Advancing the Use of HIE Data for Research

Michael J. Saks,[*] Adela Grando,[**] Chase Millea,[***] & Anita Murcko[****]

ABSTRACT

Health Information Exchanges (HIEs) are centralized repositories of patients' health records. The records come from most or all of the providers and health-care organizations within a given region or locale. Used mainly for clinical care, patients' records can generally be accessed by any of the patients' health-care providers, care coordinators, or payors, enabling them to see comprehensive and up-to-date health information pertaining to the patient.

Patients' records and HIEs are heavily regulated by federal and state law both to achieve effective flow of information and ensure the privacy and security of the data.

That same data could be a remarkable resource for medical researchers to use to improve health and health care. But the data sit unused by researchers. Why? One of the principal barriers to research use of those data has been the law. This article reviews what those policy problems are and discusses a range of solutions. In addition, innovations in information technology have been proposed that would ameliorate the same problems that the law confronts.

Legal reforms could remove the legal barriers as well as facilitate the technological advances needed to make HIE databases more available to medical researchers, while protecting the privacy and security of the data.

---

[*] Sandra Day O'Connor College of Law, ASU
[**] College of Health Solutions, Biomedical Informatics, ASU
[***] Health Current, Phoenix, AZ
[****] College of Health Solutions, Biomedical Informatics, ASU

INTRODUCTION

> The availability of electronic health records is "not just about patients having access to their data and making it easier for them and their doctors, but it could really push the new generation of research and coming out with more cures and more personalized treatment."
>
> *Seema Verma*[1]

An exceptional resource for medical research sits almost entirely untouched by researchers. That resource consists of huge databases of health information in the possession of health information exchanges (HIEs) around the nation.[2] Though HIEs were created to facilitate patient care, the data they hold offer an unprecedented opportunity for medical (and other) research.[3] The maxim "if you build it they will come"[4] might apply to many creations but, so far, not to this invaluable resource.[5] There are several barriers to using those data, but one of the greatest challenges is the law.[6] For the most part, those legal barriers are inadvertent.[7] But some—such as an existing Arizona statute[8]—are so impassable that research is all but prohibited.[9]

The potential usefulness of HIEs as sources on which researchers of various kinds might draw cannot be overstated. Putting those data to use for research would greatly increase the return on the public investment in

---

1. *Seema Verma's the Point Person on Federal Health Care*, AARP (June 4, 2019), https://www.aarp.org/politics-society/government-elections/info-2019/seema-verma-administrator-cms.html [https://perma.cc/QJ9G-CYWF] (quoting Seema Verma, Administrator, Centers for Medicare and Medicaid Services).

2. *What is HIE?*, HEALTHIT.GOV (May 1, 2019), https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/what-hie [https://perma.cc/MPM2-ZYNU]. These are also known as health information exchange organizations (HIOs). *Health Information Exchange (HIE),* ARIZ. HEALTH CARE COST CONTAINMENT SYS., https://www.azahcccs.gov/AHCCCS/Initiatives/HIT/HIE.html [https://perma.cc/HT3T-D568].

3. *See* Lucila Ohno-Machado, *Electronic Health Records and Health Information Exchange*, 25 J. AM. MED. INFORMATICS ASS'N 617 (2018).

4. FIELD OF DREAMS (Gordon Company 1989).

5. Dori A. Cross, Jeffrey S. McCullough, & Julia Adler-Milstein, *Drivers of Health Information Exchange Use During Postacute Care Transitions*, 25 AM. J. MANAGED CARE e7, e11–12 (2019).

6. Cason D Schmit et al., *Falling Short: How State Laws Can Address Health Information Exchange Barriers and Enablers*, 25 J. AM. MED. INFORMATICS ASS'N 635, 636 (2018).

7. *See id.* at 638–41.

8. *See* ARIZ. REV. STAT. ANN. §§ 36-3801 to -3809 (2020).

9. Schmit et al., *supra* note 6.

creating HIEs.[10] Data collection for research is costly, effortful, and time-consuming.[11] Access to HIE data would reduce most of those data collection burdens, and facilitate the knowledge-building that science exists to do and on which progress in health and health care depend.[12]

For example, researchers at Case Western Reserve University recently found Parkinson's disease to be three times more likely to develop in people who had had appendectomies.[13] The study involved more than 62.2 million patients.[14] It found that nearly one percent of those who had had appendectomies developed Parkinson's compared to fewer than a third of a percent of those who had not had appendectomies.[15] The next step would be to determine what about appendectomies or the conditions requiring them contributed to Parkinson's so that a means of prevention from this particular cause can be found. Smaller studies are less able to detect low-frequency relationships.[16] Conduct similar studies to discover other apparent causes and their preventions and the incidence of Parkinson's might be reduced appreciably. Conduct similar studies to discover causes and preventions of other diseases, and the public's health could improve markedly.[17] The large databases of HIEs offer a remarkable new opportunity to discover clues to the elusive causes of some illnesses, such as autism.[18]

In addition to research to improve population health through identification of causes or risk factors associated with disease, other benefits of tapping HIEs for research purposes could include improvement in diagnostic tools and techniques, assessment of the efficacy (and side effects) of treatments, real-time monitoring of patient health, longitudinal studies of health and

---

10. Valerie A. Yeager et al., *Challenges to Conducting Health Information Exchange Research and Evaluation: Reflections and Recommendations for Examining the Value of HIE*, EGEMS 7–9 (Sept. 4, 2017), https://egems.academyhealth.org/articles/10.5334/egems.217/galley/194/download/ [https://perma.cc/9ZZ3-AGFP].

11. *Id.* at 9.

12. *Id.*

13. MOHAMMED Z. SHERIFF ET AL., PARKINSON'S DISEASE IS MORE PREVALENT IN PATIENTS WITH APPENDECTOMIES: A NATIONAL POPULATION-BASED STUDY (2019), http://meetings.ssat.com/abstracts/2019/739.cgi [https://perma.cc/A2PV-ZC5Y].

14. *Id.*

15. *Id.*

16. A. Hackshaw, *Small Studies: Strengths and Limitations*, 32 EUR. RESPIRATORY J. 1141 (2008), https://erj.ersjournals.com/content/erj/32/5/1141.full.pdf [https://perma.cc/HV25-C3A6].

17. *See* MARY DEVEREAUX, THE USE OF PATIENT RECORDS (EHR) FOR RESEARCH (2013), https://medschool.ucsd.edu/som/dbmi/education/seminars/Documents/11-8-2013-EHR%20for%20research.pdf [https://perma.cc/ZK28-TSA8].

18. *See, e.g.*, Jessica Wright, *Electronic Medical Records May Reveal Subgroups of Autism*, SPECTRUM (Jan. 16, 2014), https://www.spectrumnews.org/news/electronic-medical-records-may-reveal-subgroups-of-autism/ [https://perma.cc/P54M-DASS].

health-related behavior, improvement in patient safety, pharmacovigilance (post-marketing surveillance of drugs and devices), differential care as a function of social circumstances (race/ethnicity, sex/gender, socioeconomic status), ways to improve cost-effectiveness of health care, public health initiatives, and more.[19]

Research less directly connected to health care might also benefit from employing HIE databases, ranging from discoveries in basic science to epidemiologic research relevant to mass disease and injury (e.g., toxic substances, occupational illnesses), life span development, and other topics.

Despite the potential value-added of HIE data, very little research has thus far made use of HIEs. The purpose of this article is to explore the causes and to contemplate potential solutions to the problem of non-use of this invaluable data resource. Part I describes HIEs—what they are and what purposes they serve and potentially could serve. Part II examines barriers to the wider use of HIEs for advancing biomedical and other research. Part III discusses the need to balance, on the one hand, privacy and autonomy and, on the other hand, the desirability of making health data available for research purposes in order to grow more knowledge more efficiently. Various options exist in current law, or have been proposed, for balancing those competing goods. Part IV explicates those options.

I. HEALTH INFORMATION EXCHANGES

HIEs—organizations serving as a neutral "hub" of information for health-care stakeholders—have become integral components of health-care communities throughout the country.[20] By providing impartial forums in which competing entities can share critical health information, they process millions of health-care transactions daily[21] and continue to grow.

As a patient visits various health-care providers within a given area, each provider can access the patient's health record as it develops from visits to providers throughout the community (so long as all are participants in the HIE).[22] The available health data include patients' clinical, pharmacy, laboratory, radiology, and other records.[23] The HIE facilitates coordination of

19. *See generally* Jason S. Shapiro et al., *Using Health Information Exchange To Improve Public Health*, 101 AM. J. PUB. HEALTH 616 (2011) (discussing possible public health benefits of HIEs).

20. HEALTHIT.GOV, *supra* note 2.

21. In Arizona alone, there are more than 2.5 million transactions weekly. *Health Current Statistics*, HEALTH CURRENT, https://healthcurrent.org/hie/network-by-the-numbers/ [https://perma.cc/2NPC-3SD7] (reflecting most recent twelve months).

22. HEALTHIT.GOV, *supra* note 2.

23. *Id.*

care among the patient's multiple caregivers and payors.[24] Redundancy of testing (and associated cost) is reduced.[25] A caregiver can receive alerts about a patient whose condition needs to be monitored, such as clinical test results that have become available, or changes in a patient's condition reflected in emergency department visits or hospital admissions.[26] To give a dramatic example, emergency department staff can gain instant access to vital health information about a newly arrived patient urgently in need of care.[27]

These growing databases of clinical transactions afford another opportunity to improve patient health: vast quantities of information for research.

A. Contents of HIE Databases

The inventory of data contained within HIEs include most of what is part of patients' electronic health records (EHR) maintained by participating providers and health-care organizations.[28] As those originating records expand over time in the range of their content, so too will the contents of HIEs. In the foreseeable future, we should not be surprised to see HIE databases include information from personal health devices that perform biometric monitoring, data from mobile health applications, and genomic data (necessary for delivering personalized medical care).[29] Also included are likely to be social and behavioral data (sex/gender, race/ethnicity, education, occupation, exercise, diet, socioeconomic data, living environment, etc.), which might constitute risk factors for future health problems, or moderating variables (interactions) for drugs and other treatments—collectively referred to as social determinants of health. Anything else that is potentially relevant to health could, in theory, be harvested from the internet: how and where we

---

24. *Id.*

25. *Id.*

26. *Id.*

27. Kyle Murphy, *Using Health Information Exchange To Reduce Strain on EDs*, EHR INTELLIGENCE (Aug. 25, 2016), https://ehrintelligence.com/news/using-health-information-exchange-to-reduce-strain-on-eds [https://perma.cc/C64D-NBXK].

28. *FAQ: Health Information Exchange (HIE)*, HEALTHCARE INFO. & MGMT. SYS. SOC'Y (Oct. 29, 2014), https://www.himss.org/library/health-information-exchange/FAQ [https://perma.cc/L2C5-C5WF].

29. *See, e.g.*, Alessandro Blasimme et al., *Data Sharing for Precision Medicine: Policy Lessons and Future Directions*, 37 HEALTH AFF. 702 (2018) ("Data sharing is a precondition of precision medicine. . . . [W]e suggest leveraging emerging technologies to streamline robust informed consent procedures and privacy-preserving data processing, and we propose the introduction of reciprocity-based data-access models to promote data quality."); *see also* Jessica D. Tenenbaum et al., *An Informatics Research Agenda To Support Precision Medicine: Seven Key Areas*, 23 J. AM. MED. INFORMATICS ASS'N 791 (2016).

drive, our interactions with the ever-expanding internet of things, purchase and search data, always-on wearable technologies with voice and video interfaces, our social networks, and our use of social media.[30] Anything that might reveal illness, precursors of illness, or predictors of illness would be of potential value. The utility of much of the data will depend, of course, on whether researchers are able to test the relationships among these variables.

Masses of digital information that describe who we are and predict what we will do already are being collected and sold by businesses such as Facebook and Google to advertisers, data brokers, and to anyone else who believes they can use those data for *their* benefit—for anything from turning profits to winning elections.[31] Why not also allow one's data to be used for *our* benefit by researchers and health-care providers to try to keep us healthier longer? But, as we shall see, the barriers to analyzing our health data to improve our health and protect us from illness are subject to greater restriction than are the data we already give away to businesses, tech companies, and commercial data brokers.

B. Methodological Advantages and Limitations

In addition to cost savings from making repeated collection of the same data by different researchers unnecessary, the colossal size of the HIE database would increase the quality and utility of the research in various ways. Among them are: sample sizes in every analysis would be far larger than that afforded by conventional data sources, the search for small subpopulations would be facilitated, the study of relatively rare diseases' subtle relationships would be facilitated, and the capacity to control statistically for confounding variables would be greatly enhanced.

---

30. Wullianallur Raghupathi & Viju Raghupathi, *Big Data Analytics in Healthcare: Promise and Potential*, 2 HEALTH INFO. SCI. & SYS. 1, 1 (2014) ("The totality of data related to patient health care and well-being" includes traditional clinical and test results, as well as individual patient "social media posts, including Twitter feeds (so-called tweets), blogs, status updates on Facebook and other platforms, and web pages . . . ."); *see also* Michael N. Cantor & Lorna Thorpe, *Integrating Data on Social Determinants of Health into Electronic Health Records*, 37 HEALTH AFF. 585 (2018).

31. Peter Groves et al., *The 'Big Data' Revolution in Healthcare*, CTR. FOR U.S. HEALTH SYS. REFORM BUS. TECH. OFF. 1, 3 (2013), https://www.academia.edu/9736258/The_big_data_revolution_in_healthcare [https://perma.cc/R7WL-5VZN] ("[P]ayors can learn about patients' finances, buying preferences, and other characteristics through companies that aggregate and sell consumer information, such as Acxiom and Accurint."). *See generally*, Nicholas Confessore, *The Unlikely Activists Who Took on Silicon Valley—and Won*, N.Y. TIMES MAG. (Aug. 14, 2018), https://www.nytimes.com/2018/08/14/magazine/facebook-google-privacy-data.html [https://perma.cc/6S9D-9T24].

On the other hand, the form and quality of data in the EHR are not entirely adequate for research purposes. As noted in a report prepared by the Leveraging EHR for Clinical Research Now! Think Tank, "EHR data are not created for the purposes of clinical research, so numerous challenges arise from the structure, content, form, and completeness of these data."[32] Hoffman and Podgurski have catalogued a variety of problems with the content and data structure of large health-record databases, including: selection bias,[33] measurement bias, data entry errors, incomplete or fragmented data, data coding errors, lack of standardization, and extraction errors due to software failures.[34] Studies have found that EHR data are of better quality and more complete for some kinds of patients (e.g., those with continuous insurance coverage) and for some kinds of services (those covered by insurance, those not requiring referrals to other providers).[35] The collaboration of researchers with clinicians and EHR specialists is likely to improve data quality in ways that not only will make research more informative, but will also benefit patients and providers.

The databases we have been describing have inherent limitations for use in health-care research and other research.[36] Importantly, the data are observational (correlational, non-randomized) as opposed to interventional (experimental, randomized, controlled), and therefore cannot rigorously facilitate research seeking to discover cause-effect relationships.[37] Researchers can attempt to remove the effects of confounding variables statistically, but doubts will unavoidably remain.[38] Matters can be made worse by unsophisticated analysis, which can misinterpret mere associations

---

32. Sudha R. Raman et al., *Leveraging Electronic Health Records for Clinical Research*, 202 AM. HEART J. 13, 17 (2018).

33. Affected, for example, by which portions of the population lack health-care coverage and are missing from the database, or whether patients opt into or out of inclusion in HIEs. *See* Sharona Hoffman & Andy Podgurski, *Balancing Privacy, Autonomy, and Scientific Needs in Electronic Health Records Research*, 65 SMU L. REV. 85, 114 (2012).

34. *Id.* at 85.

35. *E.g.*, J. Heintzman et al., *Agreement of Medicaid Claims and Electronic Health Records for Assessing Preventive Care Quality Among Adults*, 21 J. AM. MED. INFORMATICS ASS'N 720 (2014); Jennifer E. DeVoe et al., *Electronic Health Records vs Medicaid Claims: Completeness of Diabetes Preventive Care Data in Community Health Centers*, 9 ANNALS FAM. MED. 351 (2011).

36. Mayer-Schönberger and Cukier identify three elements of the big data environment: access to huge amounts of data; use of messy rather than clean data; and the use of large, messy data sets to discover correlations and make predictions. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK (2014).

37. *Id.*

38. *See* Andrea C. Skelly, Joseph R. Dettori & Erika D. Brodt, *Assessing Bias: The Importance of Considering Confounding*, 3 EVIDENCE-BASED SPINE CARE J. 9 (2012).

for causation, and other erroneous inferences.[39] Flawed data analysis conclusions can, in turn, affect policy decisions and population health, as well as individual health care.[40]

On the other hand, variables contained in HIE databases can be linked to interventional research and thereby enrich the findings of randomized studies.[41] The HIE data of patients who volunteer to serve in randomized experiments could be accessed more efficiently and inexpensively than contacting each source of the original data. Moreover, research, unlike clinical care, is less vulnerable to the potential harms of certain kinds of errors in patients' records, namely, random errors.[42]

Biases that are non-random, however, are likely to result in misleading research conclusions. One source of bias is central to the focus of this article: the degree of patient control over researcher access to the patient's data, and how that control is exercised. The more control we as individuals have over the uses to which our personal health information is put, the more we are able to pick and choose which data to withhold from analysis, or for use by researchers funded by some sources and not others, with the result that the more spurious correlations will seep into the data and the results will be less trustworthy.

II. BARRIERS TO WIDER RESEARCH USE

Studies making use of HIE repositories are remarkably few, especially considering their tremendous potential to grow new knowledge. Why? And what will need to be done to overcome whatever the barriers might be?

Addressing a rather different question—namely, why HIEs are expanding more slowly than expected—Michelle Mello and colleagues suggest that

---

39. *See id.*

40. Examples of such errors include hormone replacement therapy, see JERRY AVORN, POWERFUL MEDICINES (2004), and the relationship between psychiatric disorders and abortions, see Julia R. Steinberg et al., *Fatal Flaws in a Recent Meta-Analysis on Abortion and Mental Health*, 86 CONTRACEPTION 430 (2012), https://www.sciencedirect.com/science/article/abs/pii/S0010782412001503 [https://perma.cc/5QY9-V6RN].

41. This is done by adding important additional dependent variables. *See* Hoffman & Podgurski, *supra* note 33, at 90.

42. Random errors will inject noise into the data distributions, diluting correlations and enlarging error terms in significance tests. But they will not lead relationships to point away from the correct direction and toward an incorrect direction. What they can do is increase the risk of Type II error (producing non-significant statistical conclusions when a true relationship exists). In that sense, the results will be more conservative. On the other hand, the enormous sample sizes that will be involved reduce the risk of Type II error considerably, so the studies will be robust against random error.

barriers of the recent past were technological, financial, due to federal and state privacy laws, and due to the reluctance of providers and health-care organizations to share patient data in a competitive environment.[43] Mello et al. explain that most of those barriers have been removed, certainly at the federal level; and some did not actually exist in the first place, but were only imagined (through ignorance, misunderstanding, or offered as a subterfuge to conceal a desire to monopolize patients).[44]

Technological challenges have been met by government and market initiatives.[45] Notably, federal law provided financial assistance and incentives to encourage adoption of EHRs and to promote the networking of those records (particularly through HIEs).[46]

The flow of patients' health information is facilitated by a number of provisions in federal law and agency initiatives. The Health Insurance Portability and Accountability Act (HIPAA) and its implementing regulations empower providers to disclose patients' health records without patient consent or authorization for purposes of treatment, payment, and health-care organization operations.[47] Several federal initiatives offered clarification and guidance to health-care providers and organizations about those regulations, and pointed out that disclosing providers are not liable for what lawful recipients of data might later improperly do with the information.[48] Furthermore, revised rules on the confidentiality of substance abuse records are making it easier for patients to authorize wider disclosure when they wish to.[49]

Additionally, recent changes in law aim to prohibit data blocking (so providers may not withhold patient information), make improvements in record-matching capability (so all of the records of a single patient can be more accurately linked together), and promote efforts to facilitate greater interoperability (among different EHR systems).[50] The Office of the National Coordinator for Health Information Technology (ONC)—an agency within

---

43. Michelle M. Mello et al., *Legal Barriers to the Growth of Health Information Exchange—Boulders or Pebbles?*, 96 MILBANK Q. 110, 110 (2018).

44. *Id.* at 132.

45. By 2016, 96% of hospitals had switched from paper records to EHRs, as had 78% of physicians. Dr. Karen B. DeSalvo & Dr. Vindell Washington, *By the Numbers: Our Progress in Digitizing Health Care*, OFF. OF NAT'L COORDINATOR FOR HEALTH INFO. TECH.: HEALTH IT BUZZ BLOG (Sept. 29, 2016), https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/numbers-progress-digitizing-health-care [https://perma.cc/DA3K-4LT5].

46. The Health Information Technology for Economic and Clinical Health Act, 42 U.S.C. § 17901 (2018).

47. 45 C.F.R. § 164.506 (2019).

48. 45 U.S.C. § 300jj–52(b)(3)(B) (2018).

49. Confidentiality of Substance Use Disorder Patient Records, 42 C.F.R. § 2 (2019).

50. 45 U.S.C. § 300jj–52 (2018).

the Department of Health and Human Services (HHS)—and the Centers for Medicare and Medicaid Services (CMS) are furthering this legislation with proposed rules that encourage more widespread sharing of data (and patient control of that sharing), and penalize actors who engage in "information blocking."[51]

Mello et al. conclude that the inconsistent complex of state health-care privacy laws is one of the chief obstacles to greater participation with and expansion of HIEs.[52]

Our question begins where Mello et al.'s question ends. That is, why aren't the HIEs that already exist being employed for medical (or other) research? Carol Parker and colleagues have been studying the researchers who ought to be using HIEs in their research.[53] In a recent search of the research literature, they found only eighteen studies that employed data from HIEs, and sixteen of those were studying the HIEs themselves.[54]

In another study, Parker et al. examined data from a recent installment of an ongoing national survey of HIEs.[55] Only 23% of HIEs reported that they supported research, 47% were planning to do so at some time in the future, and 30% did not, and did not plan to, work to facilitate research uses of their data.[56] Comparing the characteristics of HIEs that did support versus those that were planning to support research activity in the future, Parker et al. found that those that supported research, compared to those that did not yet, were more likely to be able to create de-identified datasets, had prepared data-use agreements, required research to be approved by an Institutional Review Board (IRB), required approval of data sharing research proposals by an

---

51. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program, 84 Fed. Reg. 7424 (proposed Mar. 4, 2019) (to be codified at 45 C.F.R. pts. 170–71); Medicare and Medicaid Programs; Patient Protection and Affordable Care Act; Interoperability and Patient Access for Medicare Advantage Organization and Medicaid Managed Care Plans, State Medicaid Agencies, CHIP Agencies and CHIP Managed Care Entities, Issuers of Qualified Health Plans in the Federally-Facilitated Exchanges and Health Care Providers, 84 Fed. Reg. 7610 (proposed Mar. 4, 2019) (to be codified at 45 C.F.R. pts. 406–07, 422–23, 431, 438, 457, 482, 485, 156).

52. *See* Mello et al., *supra* note 43, at 121–22.

53. Carol Parker, Mathew Reeves & Michael Weiner, *Health Information Exchanges—Unfulfilled Promise as a Data Source for Clinical Research*, INT'L J MED. INFORMATICS, Mar. 2016 at 1, 1.

54. *Id.*

55. Carol Parker, Julia Adler-Milstein, Mathew Reeves, & Michael Weiner, *Health Information Exchange Organizations and Their Support for Research: Current State and Future Outlook*, 54 INQUIRY: J. HEALTH CARE ORG, PROVISION, & FINANCING 1 (2017).

56. *Id.* at 3. The word "support" should be understood to mean willing and able to work with researchers to share data in possession of the HIE. It does not mean providing financial support to the researchers.

oversight body, and had other policies and procedures in place to facilitate data sharing.[57]

A fundamental problem might be that researchers are unaware of the data resource that is the HIE. Consequently, few if any researchers are knocking on the doors of HIEs. In turn, few HIEs are prepared to receive them. But if researchers were to knock, fewer than a quarter of HIEs would be ready to work with them, so why bother knocking? This is something of a reverse chicken-and-egg problem. How do you get one side of the relationship moving when the other side isn't ready? Whichever side of the partnership makes the first move will be wasting its time because there is no partnership. HIEs and researchers will have to evolve their relationship and do so together. HIEs and research communities must work together to understand opportunities for cooperation (i.e., understanding what data are available and for what purposes the data might be useful). Perhaps what is required will be government initiatives to jump-start those collaborations by incentivizing researchers to use HIEs and helping HIEs to prepare themselves to work with researchers. Such a program could invite proposals from teams consisting of HIEs partnering with teams of researchers who have jointly generated a plan for working together. As HIEs mature further, cross-sector communication with the research community will open the door to opportunities that can last well into the future.

As we discuss at greater length below, federal law has removed (or never imposed) a variety of potential barriers to disclosing HIE data for research purposes.[58] A number of different paths exist to enable disclosure of patient data to researchers, and at the same time ensure privacy and data security. These are described in Part IV.

Finally, there is one persistent barrier that overlaps with what Mello et al. found on the clinical care side: state health record privacy laws. HIPAA creates only a floor of privacy protection for patients' health records; where state laws more stringently protect patient privacy, HIPAA defers to those state laws.[59]

That creates a tangle of differing state laws that make the sharing of data across state lines for research purposes difficult if not impossible. The data in each state's HIE will reflect inconsistencies in state-level rules such as in regard to what variables to include, how long records must be retained, who may access for what research purposes, lack of interoperability, and other

57. *Id.* at 4.
58. *See infra* text accompanying notes 163–171.
59. 45 C.F.R. § 160.203(b) (2019).

variations.[60] Some state laws make even intrastate research using a single statewide HIE so nearly impossible that it is tantamount to banning the research. Arizona provides an example of such a law, discussed below in Part IV.A.

Two considerations are paramount for increasing the utility of HIEs for research. First, in an era of fierce competition and regulatory fluidity, building patients' and communities' trust is both difficult and necessary for research uses of HIE data as well as expansion of HIEs for clinical purposes. Second, concerns about data governance and privacy controls (e.g., protections against re-identification of de-identified data) must be addressed to limit adverse regulation and encourage stakeholder participation.

III. BALANCING THE PUBLIC GOOD AND INDIVIDUAL PRIVACY-AUTONOMY

The essential tension that must be resolved is between the benefits to be gained by putting large amounts of patient data to work to discover ways to keep us healthier longer, on the one hand and, on the other, the benefits that patients derive from exercising individual control over their personal health information. The conflict is between two goods, two different kinds of well-being.[61]

The benefits of research to all of us as individuals and to the population's health should be obvious at this point in history. A report by ONC summarizes:

---

60. John W. Hill et al., *A Proposed National Health Information Network Architecture and Complementary Federal Preemption of State Health Information Privacy Laws*, 48 AM. BUS. L. J. 503, 531 (2011). John W. Hill et al. further explain that "[t]he current patchwork system of privacy laws is wastefully inefficient in requiring HCPs [health care providers] in different states to maintain knowledge of multiple legal regimes. When PHI is transmitted electronically across state lines, HCPs must be familiar with three sets of laws: federal law, laws of the state where the transmission originates, and laws of the destination state. Even when electronic transmissions are intrastate, whether federal or state law governs depends upon which set of laws is more stringent on the relevant point. State law is more stringent if, for instance, it either prohibits or restricts a use or a disclosure when HIPAA would allow the use or disclosure, or it provides patients with greater rights of access to, or amendment of, their records than HIPAA affords." *Id.* (citations omitted).

61. *See, e.g.*, DEP'T. OF HEALTH AND HUM. SERV., FED. HEALTH IT: STRATEGIC PLAN 2015–2020 (2015) (specifying four strategic goals: "advance person-centered and self-managed health; transform health care delivery and community health; foster research, scientific knowledge, and innovation; enhance nation's health IT infrastructure"); PRESIDENT'S COUNCIL OF ADVISERS ON SCIENCE & TECHNOLOGY, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (2014); PRESIDENT'S COUNCIL OF ADVISERS ON SCIENCE & TECHNOLOGY, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE (2014).

> Research drives innovation and scientific discovery. As health IT use increases among individuals, providers, and public health entities, it creates a rich source of data. This affords exciting new opportunities to use health IT and new data to enhance clinical decision support, deepen quality improvement, reduce health disparities, improve post-marketing surveillance of the effects of medical drugs and medical devices, enhance care transitions, and enable research on prevention, diagnosis, and treatment of disease and disability.[62]

In short, without research there can be little growth of knowledge, and without data there can be no research.

From the perspective of individual patients, however, data sharing means actual or potential loss of privacy and control over personal health information.[63] The most concrete harms could occur if an employer, insurer, lender, judicial decision-maker, or other person or entity in a position to make a decision that affects the person learned of a health problem that led to making an adverse decision concerning the person.

Employers might not want to hire, promote, or invest in those with certain health risks. Personal control over disclosure, therefore, increases the ability of individuals to reduce the risks of such bias, or to weigh for themselves whether the benefits of disclosure are worth any accompanying risks. Much as the right of private property secures for the property owner the power to exclude others,[64] the concept of informational privacy would empower the owners of the information to disclose or withhold disclosure to others as they wish.[65]

Additional types of harm might occur. These include embarrassment if others learned of a condition that is viewed as discreditable or evokes pity in the patient's social world. Some patients might not want their data to contribute to a study the results of which could potentially lead to health-care

---

62. DEP'T. OF HEALTH AND HUM. SERV., *supra* note 61, at 21.

63. *See* Maria Adela Grando et al., *A Study To Elicit Behavioral Health Patients' and Providers' Opinions on Health Records Consent*, 45 J. OF L., MED. AND ETHICS 238 (2016); Hiral Soni et al., *Perceptions and Preferences about Granular Data Sharing and Privacy of Behavioral Health Patients*, MEDINFO 2019: HEALTH AND WELLBEING E-NETWORKS FOR ALL 1361 (2019).

64. Abraham Bell & Gideon Parchomovsky, *Property as the Right To Be Left Alone*, U. PA. L. REV. (forthcoming) (arguing for restoring "the symbiotic relationship between privacy and property" law).

65. The situation invites free-riding that benefits some individuals in the short term while harming everyone in the long term. Everyone wishes to avail themselves of advances in health knowledge. Some or many individuals will see benefit in withholding their personal data to researchers. But the more that people do the latter, the less there will be of the former.

practices or policies that would be in conflict with the data donor's values.[66] Even though such findings would emerge even without the use of that particular individual's data, some patients might nevertheless be troubled to know that *their* data contributed to the findings. Groups—racial, tribal, ethnic, national-origin—might be troubled by findings they feel stigmatize them or tend to support stereotypes about their group. Accordingly, they might not want members of their group to allow their personal data to be used in a study that could lead to such findings, regardless of the preferences of the individuals.[67]

A thoughtful set of general principles for balancing the beneficial use of health data and the interests of patients has been proposed by a committee of experts from the leading medical informatics organization.[68] These principles are listed in Appendix 1 and will echo through some of the discussion below.

Paradoxically, at the same time that law and biomedical ethics approach these issues with the utmost care and concern about risks (mostly economic) to patients and patients' sensibilities, those same individuals are exposed to quite a variety of personal information disclosure risks that go largely uncontrolled.[69] Indeed, they themselves often give away their most personal health and other data with little apparent concern.[70] Accidental leakage of data from government regulated health-care databases, or even hacking, are small sporadic rivulets when viewed against a background of the river of information that flows constantly from planned commercialization of personal data.[71]

---

66. For example, a patient might not want to contribute to findings that could lead to an increase/decrease in the availability of abortion.

67. In this situation, it probably would be easier for the group to use its influence to prevent such a study from being conducted or to suppress the findings than it would be to police the choices of its many individual members. Moreover, the group might consider that it is better to learn of a genetic trait or health problem that occurs disproportionately among members of the group so that actions can be taken to ameliorate the problem, rather than to keep the group defenseless to the biological condition.

68. George Hripcsak et al., *Health Data Use, Stewardship, and Governance: Ongoing Gaps and Challenges: A Report from AMIA's 2012 Health Policy Meeting*, 21 J. AM. MED. INFORMATICS ASS'N 204 (2014).

69. *See, e.g.*, Rob Copeland, *Google's 'Project Nightingale' Gathers Personal Health Data on Millions of Americans*, WALL STREET J. (Nov. 11, 2019), https://www.wsj.com/articles/google-s-secret-project-nightingale-gathers-personal-health-data-on-millions-of-americans-11573496790 [https://perma.cc/FHG3-ACBQ].

70. *See, e.g.*, Lisa Bari & Daniel P. O'Neill, *Rethinking Patient Data Privacy in the Era of Digital Health*, Health Affairs Blog (Dec. 12, 2019), https://www.healthaffairs.org/do/10.1377/hblog20191210.216658/full/ [https://perma.cc/ASW2-HQPG].

71. *See, e.g.*, Copeland, *supra* note 69.

Critics warn that expanded commercialization threatens to take patients' data beyond traditional health-care settings and entities and lead to excessive monitoring and advice-giving, surveillance prompts, and marketing by third parties.[72] While sometimes these alerts will be beneficial, overall they expose patients, and consumers more generally, to new vistas of annoyance. Imagine something like robo calls, but aimed at each of us, individually, telling us of some bodily imperfection or remote threat that has been detected and urging us to make an appointment for care or to buy a product to address the condition. Imagine lists of people suffering particular medical and social ills being sold on the data market.[73]

Those intrusions need not result from disclosures of patients' health records. They can result from internet searches about health problems, which search engines collect and sell to businesses.[74] Or companies can monitor the purchase patterns of their customers and infer health issues. In one infamous example, based on product purchase patterns, Target Corporation correctly determined that a high-school-aged girl was pregnant and began mailing coupons for baby products to her home.[75] The mailings were noticed by her father who, until then, had not been aware of her pregnancy.[76]

When invited to do so outside the formal health-care system, many people are willing to give away their personal information—for what they think could benefit others, for a trifling amount of money, or for no reason at all—unconcerned about the risks they might be creating for themselves.[77] For example, people who join PatientsLikeMe contribute large amounts of valuable personal data in exchange for a t-shirt, while typically overlooking the member (user) agreement, which warns:

---

72. Janine S. Hiller, *Healthy Predictions? Questions for Data Analytics in Health Care*, 53 AM. BUS. L. J. 251, 265–67 (2016).

73. The World Privacy Forum found that lists of rape victims, AIDS sufferers, and individuals with dementia were all being offered for sale. *See* Melanie Hicken, *Data Brokers Selling Lists of Rape Victims, AIDS Patients*, CNN (Dec. 19, 2013, 12:38 PM), http://money.cnn.com/2013/12/18/pf/data-broker-lists/ [https://perma.cc/JTJ4-T56H].

74. Heather Kelly & Scott McLean, *Your Browser History Is for Sale, Here's What You Need to Know*, CNN (Apr. 6, 2017), https://money.cnn.com/2017/04/05/technology/online-privacy-faq/ [https://perma.cc/5UYT-WSVS].

75. Kashmir Hill, *How Target Figured out a Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012, 11:02 AM), https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#361fc05a6668.

76. *Id.*

77. *See, e.g.*, Edward C. Baig, *DNA Testing Can Share All Your Family Secrets. Are You Ready For That?*, USA TODAY (July 4, 2019), https://www.usatoday.com/story/tech/2019/07/04/is-23-andme-ancestry-dna-testing-worth-it/1561984001/ [https://perma.cc/NGA3-76SD].

> It is possible that a Member could be identified using information shared on PatientsLikeMe (and/or in conjunction with other data sources). A Member could be discriminated against or experience repercussions as a result of the information shared. For example, it is possible that employers, insurance companies, or others may discriminate based on health information.[78]

A study of how PatientsLikeMe members move through the registration process found that they did so much as people sign up for other internet services: none of them opened the "Terms and Conditions" or the "Privacy Policy"—they just clicked through and agreed to whatever they were invited to agree to.[79]

By now we all know that on social media platforms *we* (that is, our data) are the product. Platforms like Facebook, Snapchat, Instagram, and Twitter let us post and view information with our family and friends; in return, we allow them to collect data about us and sell it.[80]

One data business offers people a chance to be their own data brokers within the consumer space: "With the #My31 App consumers can claim a property interest on inherent human data, consent for privacy, authorize for permitted use, and elect for compensation if desired."[81] PatientSphere focuses on personal health information, offering patients "the ability to not only share" data on their own terms, "but also get paid for it."[82] Similarly, PatientTruth tells patients they can "own" and "monetize" their health data.[83]

Meanwhile, inside conventional health care, providers and organizations are expected, in the foreseeable future, to use digital devices to prompt patients to take their meds, get exercise, conduct self-examinations, and come in for tests and treatment before the patient realizes the need, especially those patients with good insurance.[84] Some of these intrusions will serve patients' interests; most of them will increase the health-care organization's revenues.

---

78. Privacy Policy, PATIENTSLIKEME, https://www.patientslikeme.com/about/privacy_full [https://perma.cc/GQW6-68CK] (Jan. 12, 2020).

79. W. Rowan et al., *Exploring User Behaviours when Providing Electronic Consent on Health Social Networks: A 'Just Tick Agree' Approach*, 121 PROCEDIA COMPUT. SCI. 968, 969 (2017).

80. Lois Beckett, *Yes, Companies Are Harvesting—and Selling—Your Facebook Profile*, PROPUBLICA (Nov. 9, 2012), https://www.propublica.org/article/yes-companies-are-harvesting-and-selling-your-social-media-profiles [https://perma.cc/96KG-7X8X].

81. *For Consumers*, HU-MANITY.CO, https://hu-manity.co/home-3-3/ (last visited Dec. 18, 2019); *see also* Sarah Jeong, Opinion, *Selling Your Private Information Is a Terrible Idea*, N.Y. TIMES (July 5, 2019), https://www.nytimes.com/2019/07/05/opinion/health-data-property-privacy.html?searchResultPosition=7 [https://perma.cc/RD58-X5M4].

82. Jeong, *supra* note 81.

83. *Id.*

84. Secure communications with one's patients are permissible.

How do patients themselves feel about the dilemma of sharing data for research versus maintaining tight control for protection of privacy and autonomy? Surveys generally find substantial proportions of the public expressing the desire for control of what uses their personal health information is to be put.[85] But the examples above seem to suggest that what they say about information privacy and what they do about it are at considerable odds with each other. In some (or even the same) surveys where patients express a desire to control whether health information may be shared with researchers, two-thirds also say that they believe that "[r]esearch that could be beneficial to people's health is more important than protecting people's privacy."[86]

A major problem with such surveys seems to be that context and nuance matter in ways that cause results to be volatile. The particular way a question is asked, and what information is supplied along with the question, could be pushing responses in contradictory directions, and could explain why privacy appears paramount in some surveys and support for research is in others.[87] At the same time, one survey that carefully asked some respondents their views about the disclosure of their health record *with* their identifying information included versus *without* any identifying information did not produce much of a difference in patients' attitudes about the shareability of the data.[88] The one factor that ethicists, lawmakers, and other professionals assume to be *the* most critical factor in regard to data sharing for research (i.e., identifiability) seemed not to matter much to the survey respondents.[89] Until more clarifying

85. COMM. ON HEALTH RESEARCH & THE PRIVACY OF HEALTH INFO.: THE HIPAA PRIVACY RULE, IOM, BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 247, 268 (Sharyl J. Nass et al. eds., 2009).

86. Donald J. Willison et al., *Alternatives to Project-Specific Consent for Access to Personal Information for Health Research: What Is the Opinion of the Canadian Public?*, 14 J. AM. MED. INFORMATICS ASS'N 706, 708 (2007) (in the same study, sixty percent felt their permission should be obtained before accessing data for research from their medical records).

87. Mhairi Aitken et al., *Public Responses to the Sharing and Linkage of Health Data for Research Purposes: A Systematic Review and Thematic Synthesis of Qualitative Studies*, 17 BMC MED. ETHICS 73 (2016) (reviewing relevant qualitative studies conducted in the UK and North America between 1999–2013, and concluding: "Key themes identified across the corpus of studies related to the conditions necessary for public support/acceptability, areas of public concern and implications for future research. The results identify a growing body of evidence pointing towards widespread general—though conditional—support for data linkage and data sharing for research purposes. Whilst a variety of concerns were raised (e.g. relating to confidentiality, individuals' control over their data, uses and abuses of data and potential harms arising) in cases where participants perceived there to be actual or potential public benefits from research and had trust in the individuals or organisations conducting and/or overseeing data linkage/sharing, they were generally supportive. The studies also find current low levels of awareness about existing practices and uses of data.").

88. *Id.*

89. *Id.*

research develops, current survey findings on these issues do not seem to be a dependable guide to what patients and consumers really want.[90] None of which is to say that the preferences of the public are not of central importance. Ultimately, in a democracy, the interests of the public should dictate policy. But discovering those preferences is not a simple matter.

The policy situation is not fixed in stone, of course—with data inside the health-care system being highly regulated and much personal data in the world of commerce enjoying nearly unlimited wild-west style freedom. As commercial interests extend their reach farther into our lives, and use our data in troubling ways, regulation becomes increasingly likely. Thus, the Federal Trade Commission's *Data Broker Report* recommended legislation to protect people from data disclosures and abuses that create risks for consumers.[91] Industry best practices also were recommended, as was legislation to regulate three types of data products: marketing, risk mitigation, and people search.[92] Recently, the FTC voted to fine Facebook $5 billion for mishandling personal data.[93]

But against that background, we might wonder whether data in the health-care world, particularly for use by researchers to advance knowledge aimed at making us healthier, might be somewhat over-regulated. The following Part examines the range of options that exist in current law, are under consideration, or have been proposed regarding whether and how health-care information should be regulated in relation to data-sharing for research.

---

90. A more helpful survey, though a more time-consuming and expensive one, might more concretely show patients what the researchers would see: a massive matrix of de-identified numbers or a screen in which researchers can input data analysis queries but view no raw data at all. Then, the survey respondents could be asked if they would object to their health data being made routinely available for records-review research without being asked for permission or if they wanted the right to approve or disapprove its use in general or for particular studies. If they do want greater control, the nature of their concerns could be probed: what data would they wish to hold back and from whom for what uses? A random half of the sample in such a survey could be asked about the sharing of all patients' data, or other patients' data (in which the focus is not on them individually).

91. FED. TRADE COMM'N, DATA BROKERS, A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (May 2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [https://perma.cc/6NLZ-KNER].

92. Hiller, s*upra* note 72.

93. Cecilia Kang, *F.T.C. Approves Facebook Fine of About $5 Billion*, N.Y. TIMES (July 12, 2019), https://www.nytimes.com/2019/07/12/technology/facebook-ftc-fine.html?campaign_id=60&instance_id=0&segment_id=15174&user_id=50f8bb062fa6d2fae19b80f552e1ae62&regi_id=48668335ing-news [https://perma.cc/RL7M-NU3J].

IV. RANGE OF POSSIBILITIES

A range of proposals have been advanced in efforts to balance the use of HIE data repositories for research against the need to provide adequate safeguards for the privacy of patients' records. The ideal arrangement would find a way to make as available as possible the wealth of data for growing health-care knowledge while fully addressing the caution that "[h]ealth-care providers and policy makers should ask hard questions about how harms to personal privacy can be avoided, stigmas prevented, and threats of unbridled commercialization ameliorated."[94] In the alternatives described below, we generally proceed from most restrictive regulations to most facilitative.

A. Specific Authorization (Informed Consent)

At one extreme, access to data would require that consent/authorization be obtained from each patient for each use of their data for each study being conducted. Such a requirement would have the ethical advantage that permission to use each research participant's data has been specifically granted by the individuals whose data are being analyzed. That such a rule would be burdensome—to patients and researchers alike—is obvious, especially when considering that meaningful informed consent requires explaining to patients the proposed data use (or presenting them with a lengthy informed-consent document) and offering to answer their questions. The individual, study-by-study, consent approach likely would produce the maximum non-participation as many potential research participants opt out, perhaps for no reason other than that they have grown tired of being contacted again and again for successive studies and are asked to devote their time to the informed consent process. Consequently, this approach would likely produce the most unrepresentative of samples and would almost certainly discourage many studies from being undertaken at all.[95]

Arizona's health information organization (HIO) statute contains a provision that exemplifies this extreme restriction on using data for research.[96] That Arizona law prohibits the transfer of data for any *research*

94. Hiller, *supra* note 72, at 252.

95. HIPAA requires individual patient authorization for disclosure of identifiable data for research purposes, but it has numerous exceptions to that general rule. *See, e.g.*, 45 C.F.R. § 46.104. HIPAA's Privacy Rule defines research as, "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge." 45 C.F.R. § 164.501.

96. A.R.S. § 36-3805(C):

purpose, no matter how well scrubbed the data might be of individual identifiers (i.e., including only data deemed "de-identified" under federal law), without the *health-care provider* (in whose clinic or office the data originated) obtaining the patient's informed consent, separately for each research project for which the patient's data are being sought.[97] This creates a triple burden: researchers have to recruit dispersed providers who would then have to contact their patients (again and again) in order to manage the consent process on behalf of the research project. That is a lot for researchers, providers, and patients to do to get to analyze data that could have been far more efficiently shared by the HIE (for many patients of many providers) after having been rendered anonymous by the HIE.

The target of that provision of Arizona law is not research so much as it is research employing HIE data. If the concern was that patients should not become part of scientific research in which the patient's identity becomes known to a researcher without first having the opportunity to authorize or refuse, HIPAA already requires patients' authorization under such circumstances.[98] If the concern were that privacy breaches might lead to harm to a patient, HIPAA, the Health Information Technology for Economic and Clinical Health Act (HITECH), and regulations governing human subjects research (generally referred to as the Common Rule) already require careful protection of the data from leaks, breaches, or intentional further disclosure by the researchers.[99] Researchers rarely are interested in the identity of the individuals whose data they are analyzing; they are interested in finding relationships among variables. They are almost always happy to work with a version of the EHR from which identifying information has been completely removed by the HIE. If the concern were that even such de-identified data

---

> A health information organization may not transfer individually identifiable health information or de-identified health information . . . to any person or entity for the purpose of research or using the information as part of a set of data for an application for grant or other research funding, unless the health care provider obtains consent from the individual for the transfer. A health care provider must document that it has provided a notice of transfer to the individual and that the individual has received and read and understands the notice. Documentation must be in the form of a signature by the individual indicating the individual has received and read and understands the notice and that the individual gives consent to the transfer of information. For the purposes of this subsection, "consent" means that a health care provider participating in a health information organization has provided a notice to the individual that is in at least twelve-point type and that describes the purposes of the transfer.

97. *Id.*
98. 45 C.F.R. §164.508(a)(1) (2019).
99. *See* 45 C.F.R. Part 46 (2019).

might be used to re-identify who was its source, Arizona HIO law already prohibits re-identification of individuals.[100]

The burden of repeatedly seeking and granting (or denying) consent could be reduced by expanding consent to authorize whole categories of studies. Information technologies are being developed to provide patients with the ability to consent to disclosure of their health data in a more granular fashion—specifying which categories of the information in their EHRs may be disclosed to which members of their care team (e.g., providers, care coordinators, etc.).[101] The same technology should be able to offer patients a similar opportunity to indicate which of their data they authorize to be disclosed to which kinds of researchers or research, being conducted for what kinds of purposes, and funded by what types of sources.[102] Even with consent granted (or withheld) in granular fashion, preferences likely will change over time for some patients. Provisions would need to be made for periodically tweaking one's granular preferences for consent to disclosure of data for research uses.

Complicating the situation further, there is increasing interest in and use of samples from biobanks for reference purposes (with the need to link those to patient records), and eventually genomic data will be a routine part of a patient's health record.[103] Moreover, some or many research efforts are a step on a path of research and development that will lead to the creation of products or services that will eventually be commercialized.[104] Where the potential is real for the passage of data from research to commercial applications, such information should be provided to the data subject for approval or disapproval of sharing. Granular consent procedures could accommodate all of these possibilities.

While the law certainly can secure for us (patients) the absolute power to grant or withhold our consent, and granular consent can facilitate the more

---

100. A.R.S. § 36-3804(F) provides that: "[a] person who receives de-identified information from the health information organization may not use such de-identified information, either alone or in combination with other information, to identify an individual."

101. *See* CONSENT2SHARE, https://bhits.github.io/consent2share/ [https://perma.cc/4FNW-MYQR].

102. *See* Adela Grando & Richard Schwab, *Building and Evaluating an Ontology-Based Tool for Reasoning About Consent Permission*, 2013 AMIA Ann. Symp. Proc. 514 (2013), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900195/ [https://perma.cc/9FC5-ZAJU]; Hyeoneui Kim et al., *iCONCUR: Informed Consent for Clinical Data and Bio-Sample Use for Research*, 24 J. AM. MED. INFORMATICS ASS'N 380, 380 (2017).

103. *See, e.g.*, Denise Roland, *How Drug Companies Are Using Your DNA To Make New Medicine*, WALL ST. J. (July 22, 2019), https://www.wsj.com/articles/23andme-glaxo-mine-dna-data-in-hunt-for-new-drugs-11563879881 [https://perma.cc/2LED-MDMV].

104. *See, e.g.*, Michael Szycher, COMMERCIALIZATION SECRETS FOR SCIENTISTS AND ENGINEERS (2016).

precise exercise of that power, whether we can make such choices meaningfully has been questioned.

> [R]equiring individual notice and consent for collection and use of data in the big data environment "is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data. It is simply too complicated for the individual to make fine-grained choices for every new situation."[105]

As an alternative to the burdens of individual, use-by-use, study-by-study consent—for both researchers and patients—less demanding procedures exist, which still are attentive to confidentiality and protection of health records. We discuss these next.[106]

B. Waiver of Authorization

Unlike in clinical health care, HIPAA does not permit routine disclosure of patients' identifiable health information for research without the patient's authorization. To use information for research purposes, HIPAA generally requires specific authorization by each patient for use of their identifiable health records.[107]

However, HIPAA offers other pathways to disclosure in addition to specific authorization by patients for research use. One of those is a "waiver of authorization" for disclosures under limited specified circumstances.[108] An IRB or a Privacy Board may approve a waiver or alteration of authorization under the Privacy Rule (a component of HIPAA) if it finds that the research could not practicably be conducted without access to and use of the protected health information (PHI)[109] Furthermore, the reviewing body must find that the proposed disclosure of PHI involves no more than a minimal risk to the privacy of individuals.[110] That minimal risk must be backed by *at least* three things: an adequate plan to protect the identifiers from improper use and disclosure, an adequate plan to destroy the identifiers at the earliest opportunity consistent with conduct of the research (absent a health or

---

105. Hiller, supra note 72, at 267 (quoting PRESIDENT'S COUNCIL OF ADVISORS ON SCI. AND TECH., EXEC. OFFICE OF THE PRESIDENT, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE 38 (2014), https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf [https://perma.cc/L8W2-D7D3]).

106. In all variations of these proposals, uses beyond the one or ones authorized, or further disclosure by researchers to additional parties, would be prohibited.

107. 45 C.F.R. § 164.508(a)(1).

108. *See* 45 C.F.R. § 164.512(i)(1)(i).

109. § 164.512(i)(2).

110. § 164.512(i)(2)(ii)(A).

research justification or legal requirement for retaining the identifiers longer), and adequate written assurances that the PHI will not be reused or disclosed to any other person or entity (except as required by law).[111] The entity disclosing the data must then document when and which IRB or Privacy Board approved the waiver or alteration, the IRB or Privacy Board's signed finding that the three minimal criteria have been met, and a description of the PHI for which use or access has been determined to be necessary by the IRB or Privacy Board.[112] Thus, under limited circumstances, HIPAA makes an exception allowing disclosure of individually identifiable patient information.

C. De-identification

If I am concerned about something in my health record that, if traced back to me, could embarrass or cost me, my worry ought to be dispelled if the data cannot be traced back to me.[113] Reflecting that simple insight, the major response to the risks of sharing patients' health data for research has been to decouple the health record data from any identifying information in the record. HIPAA has adopted this view. HIPAA covered entities or their business associates[114] are permitted to disclose patients' health information for research purposes if the information has been rendered anonymous (i.e. de-identified). Put differently, HIPAA classifies only individually identifiable health information as PHI. Information that is not PHI is not subject to HIPAA regulation—except that HIPAA sets the standards under which information is deemed to be de-identified and thus does not fall into the category of PHI.[115]

HIPAA specifies the de-identification process, as well as the requirements for re-identification if the data are assigned a code or other means of record identification—in that instance, disclosure of that code would constitute disclosure of PHI.[116] The Rule allows for two methods to demonstrate that data are de-identified: (1) a qualified statistician determines that the risk is very small that the information could be used alone or in combination with other available information to identify the patient, or (2) removal of eighteen

111. § 164.512(i)(2)(ii).

112. 45 C.F.R. § 164.512(i)(2).

113. The one major exception is group identity. But, as mentioned earlier, that can be handled more effectively and efficiently in ways other than by giving patients belonging to that group control over disclosure and then trying to convince or compel them to withhold disclosure.

114. HIEs (aka HIOs) fall within the definition of a business associate. *See* 45 C.F.R. § 160.103.

115. § 164.514(a).

116. § 164.502(d); § 164.514(a)–(c).

common identifiers of the individual or of any relative, employer, or household member of the individual (commonly referred to as the "Safe Harbor").[117] HIPAA specifies the de-identification process, as well as the requirements for re-identification if the data is assigned a code or other means of record identification—in that instance, disclosure of that code would constitute disclosure of PHI.[118] In essence, it consists of making sure that eighteen identifiers of the patient and of the patient's relatives, employers, or household members have been removed from the health record. Those eighteen identifiers are listed in Appendix 2.

Similarly, the European Union, most recently through its General Data Protection Regulation (GDPR), which came into effect in May 2018, carefully regulates disclosure only of personally identifiable information. But, whereas HIPAA is limited to health records, the GDPR extends to *all* "Personal Data," defined as "any information relating to an identified or identifiable natural person."[119] Like de-identification in HIPAA, anonymized data are not subject to GDPR regulation.[120] Worth noting is that the GDPR is generally more rigorous and more protective of Europeans than HIPAA and HITECH are of Americans.[121] And yet it, too, finds removal of identifying information to be a powerful privacy protection.

---

117. § 164.514(a)–(c).

118. § 164.514(c)(2).

119. Commission Regulation 2016/679 of 27 April 2016, General Data Protection Regulation, 2016 O.J. (L 119) ¶ 1.

120. *Id.* at ¶ 26. De-identification and anonymization do not map onto each other precisely. But for purposes of our discussion, they may be regarded as synonymous.

121. Among other provisions: The GDPR asserts its jurisdiction to any organization, anywhere in the world, that is processing the personal data of EU residents. *Id.* at art. 3 ¶ 1. Organizations may no longer use complicated, lengthy, incomprehensible agreements requesting consent; clear, intelligible, easily accessible terms of service must include a statement of the purposes for which data are sought. *See id.* at art. 1 ¶ 39. The request for and grant of consent must be clear and distinguishable from other matters, and withdrawal of consent must be easy. *See id.* at art. 1 ¶ 65. Except under specified conditions, "[p]rocessing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited." *Id.* at art. 9 ¶ 1. Data subjects have a right to be informed by any entity using their data of what personal data are being processed, where, for what purposes, to obtain a copy of the personal data at no charge, and to request that their data be erased (also referred to as the right to be forgotten). *Id.* at art. 1 ¶ 65; *But see id.* at art. 1 ¶ 10 (such requests for data erasure must be weighed against the public interest in the availability of the data). The GDPR incorporates into law the concept of "privacy by design," requiring data protection to be planned for as part of the architecture of a data gathering and processing system at the outset of system design. *Id.* at art. 25 ¶ 1. Data controllers are permitted to retain and process only those data which are absolutely necessary for the completion of the purposes for which they were

Unlike HIPAA, the GDPR does not specify particular items of information the removal of which produce a de-identified or anonymized record. The EU's approach is, instead, a set of guidelines for determining which data are *identifiers* and therefore must be removed to render the data anonymous.[122] An *identifiable person* is one who can be identified, directly or indirectly.[123] "[P]articular pieces of information . . . which hold a particularly privileged and close relationship with the particular individual" are termed *identifiers*.[124] Identifiers can be separated into those with potential to lead to identification of an individual and those which afford only indirect clues to a person's identity.[125] *Anonymous data* means "[a]ny information relating to a natural person where the person cannot be identified, whether by the data controller or by any other person, taking account of all the means likely reasonably to be used either by the controller or by any other person to identify that individual."[126] Related concepts for anonymous and anonymized data reflect the varying nature of the identifiers that have been removed or neutralized: *de-identified, non-identifiable, irretrievably unlinked, irreversibly de-identified, unlinked-anonymized, or irreversibly anonymized*.[127]

Relating these concepts back to the use of HIE data for research, for studies in which research participants must be recontacted, or where data outside what is held by the HIE need to be linked with the HIE data (such as from biobanks), *pseudonymization* can be used. That is, identification data

---

collected (known as "data minimization"), and to limit access to personal data to those who are needed to process it. *Id.* at art. 89 ¶ 1. To help ensure compliance with the provisions of the GDPR, every organization whose core activities include large scale data gathering of EU persons must appoint a professionally qualified Data Protection Officer. *Id.* at art. 37 ¶ 1. Violations of the GDPR can result in fines as high as four percent of an organization's annual global revenue, or 10 million euros, whichever is greater. *Id.* at art. 83 ¶ 4.

122. Article 29 Data Protection Working Party, *Working Document on the Processing of Personal Data Relating to Health in Electronic Health Records* 2 (Article 29 Working Party, Working Paper No. 131, 2007), https://www.dataprotection.ro/servlet/ViewDocument?id=228 [https://perma.cc/89F8-S7ZA].

123. *Id.*

124. *Id.*

125. *Id.*

126. Article 29 Data Protection Working Party, *Opinion 4/2007 on the Concept of Personal Data*, at 21, 01248/07/EN WP 136 (June 20, 2007), https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf [https://perma.cc/6DNB-2JKY].

127. "The transformation rules applied to data during deidentification fall into four categories: conversion (transformation of the data value into another one), truncation (only a part of the data is transmitted), access restriction (only authorized users can query the information) and removal (the information is removed from the result)." Bernice S. Elger et al., *Strategies For Health Data Exchange For Secondary, Cross-Institutional Clinical Research*, 99 COMPUTER METHODS & PROGRAMS IN BIOMEDICINE 230, 239 (2010). For example, turning birthdate into age in whole years as of the extraction date of the data.

are replaced with codes, the researchers receive only the coded data, and the key linking the two is stored by others, such as the custodian of the data, the HIE.[128] The researchers would not have access to the codes. After the data required by the study are collected, the file containing the codes would be destroyed in accord with the research protocol and data-use agreement (likely also overseen by an IRB or Privacy Board).

Increasingly, however, concerns have arisen that de-identification can be achieved in less than an absolute fashion. That is because a handful of variables—e.g., genetic, geographic, birthdate—might allow members of a large group of patients to be reduced in statistical space to single individuals.[129] Finding a unique combination of variables in a database does not in itself constitute identifying (or re-identifying) the individual described by those data.[130] The leap still must be made to an actual person, to a named individual.[131] Perhaps the first protection against re-identification is that it is hard to imagine scenarios in which researchers would want to know the names of the people whose data are in the repository. Researchers are interested in the content of variables, not the names of the individuals being measured by those variables. A concern heard more often is the risk of a data breach, such as hacking of the de-identified data by others who would want to re-identify the individuals.[132] Obviously, those hackers are taking on a greater challenge than they would be if they attacked the HIE directly, or the EHRs in the health-care organizations where the patients' data originated. Those latter sources would, if hacked, deliver up patients' identifying information without the need to play detective in the effort to re-identify.

In any event, more rigorous protections against re-identification of de-identified databases exist than the inherent protections just described, and

---

128. The complexity of pseudonymization is illustrated by a discussion in Bernice Elger et al., *supra* note 127, at 232–33.

129. That individual likely would not have a name attached, but the potential for linking a name to the statistically isolated individual might itself be concerning. A recent study reports that data scientists were able to individualize 99.98% of Americans from most datasets with as few as fifteen attributes. Luc Rocher et al., *Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models*, NATURE COMM. (July 23, 2019), https://doi.org/10.1038/s41467-019-10933-3 [https://perma.cc/5EL3-XMXU]. Though names are not included, a company, such as an insurer, might be able to locate its one customer who shares the attributes of an individual in the dataset.

130. *See generally* U.S. DEP'T OF HEALTH & HUM. SERVICES, GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE (2012). *See generally supra* note 119.

131. *Id.*

132. *See generally* Mats G. Hansson et al., *The Risk of Re-identification Versus the Need to Identify Individuals in Rare Disease Research*, 24 EUR. J. HUM. GENETICS 1553 (2016).

these inherent protections could be strengthened further.[133] Because research data, especially in the health-care context, are tightly regulated—not only by HIPAA and the HITECH Act, but also by the Common Rule applied by federal agencies, IRBs and Privacy Boards, in data-use agreements, and by other enforceable rules—researchers can be, and are, prohibited from attempting to re-identify anyone whose data are in the database they are analyzing.[134] Any researcher who, without authorization, re-identified anyone would immediately be in violation of HIPAA's Privacy Rule because it would then constitute "individually identifiable health information" and it would be in unauthorized hands.[135] An IRB-approved research protocol would specify whether the research was to employ identifiable data or only de-identified data; to re-identify the latter would violate the IRB approval and invite sanctions.[136] In addition to federal penalties enforced by the Office of Civil Rights (OCR) within HHS, some state laws, such as Arizona's, as noted earlier, prohibit re-identification of de-identified data.[137] The penalties for a violation can include debarring a researcher from eligibility for federal research grants, which would likely be a career-ending consequence.[138]

D. Regulation that Encourages Voluntary Practices

Other than health records and some financial data, Americans' personal data have generally been protected, if at all, largely by the self-regulation of the businesses and industries that come into possession of the data. In these realms, the federal government has been more willing to offer advice than regulation.

---

133. *See* discussion *infra* Section 534E.

134. *See generally* U.S. DEP'T OF HEALTH & HUM. SERVICES, *supra* note 130; Commission Regulation 2016/679 of 27 April 2016, General Data Protection Regulation, 2016 O.J. (L 119) ¶ 1.

135. *See generally* U.S. DEP'T OF HEALTH & HUM. SERVICES, *supra* note 130.

136. U.S. Dep't of Health & Hum. Servs., Off. For Hum. Res. Protections, *Approval of Research with Conditions: OHRP Guidance (2010)* (Sept. 1, 2010), https://www.hhs.gov/ohrp/regulations-and-policy/guidance/guidance-on-irb-approval-of-research-wit h-conditions-2010/index.html [https://perma.cc/LY52-LDTM].

137. A.R.S. § 36-3805(C) (2019); U.S. Dep't of Health & Hum. Servs., *Summary of the HIPAA Privacy Rule* (July 26, 2003), https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html [https://perma.cc/HS6F-H6FQ].

138. *See, e.g.*, Advisory Committee on Human Radiation Experiments, *Final Report, Chapter 14: Federal Responses to Violations of Human Subjects Protections* (1995), https://bioethicsarchive.georgetown.edu/achre/final/chap14_6.html [https://perma.cc/JV5N-4B2A].

Beginning in the early 1970s, a number of federal agencies adopted Fair Information Practice Principles (FIPPS).[139] These principles are in the nature of guidelines or suggestions to assist self-regulation in various industry sectors.[140] The first of these was developed by the Advisory Committee on Automated Personal Data Systems of the then Department of Health, Education, and Welfare (HEW).[141] Known as the Code of Fair Information Practices,[142] its principles were these:

1. There must be no personal data record-keeping systems whose very existence is secret;
2. There must be a way for an individual to find out what information about [the person] is in a record and how it is used;
3. There must be a way for an individual to prevent information about [the person] that was obtained for one purpose from being used or made available for other purposes without [the person's] consent;
4. There must be a way for an individual to correct or amend a record of identifiable information about [the person];
5. Any organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take precautions to prevent misuses of the data.[143]

Building on those, the Federal Trade Commission (FTC) later developed its own more extensive Fair Information Practice Principles—reflecting concerns about how online businesses collect and use the personal information of consumers.[144] These principles are as follows.

1. *Notice/awareness* is considered the most essential principle.[145] The entity should inform consumers of the entity's information practices, including confidentiality practices (or offer the

139. *See generally* U.S. DEP'T OF HEALTH, EDUC. & WELFARE, REPORT OF THE SECRETARY'S ADVISORY COMMITTEE ON AUTOMATED PERSONAL DATA SYSTEMS, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS (1973).

140. *Id.*

141. U.S. DEP'T OF HEALTH, EDUC., & WELFARE, REPORT OF THE SECRETARY'S ADVISORY COMMITTEE ON AUTOMATED PERSONAL DATA SYSTEMS, RECORDS, COMPUTERS, AND THE RIGHTS OF CITIZENS (1973).

142. *Id.* at xx–xxi.

143. *Id.*

144. FED. TRADE COMM'N, FAIR INFORMATION PRACTICE PRINCIPLES (1998).

145. *Id.*

opportunity to learn of them), before any personal information is provided by the consumer or collected by the business from third parties about the consumer.[146]

2. *Choice/consent* means giving consumers options as to how any personal information collected from them may be used beyond those necessary to accomplish the contemplated transaction.[147] The conventional forms of presenting the opportunity to choose are either opt-in (consumers must take affirmative steps to allow collection and use of their data) or opt-out (the data will be collected unless the consumer takes affirmative steps to disapprove collection and use).[148]
3. *Access/participation* refers to the ability of individuals to examine data about themselves as well as to question and correct the accuracy and completeness of the data.[149]
4. *Integrity/security* concerns steps the business or other entity takes to assure that third-party sources are trustworthy and that the entity will exercise care (such as by cross-checking multiple sources to increase accuracy), as well as steps taken to keep the data secure (such as preventing unauthorized access or disclosure).[150]
5. *Enforcement* of the privacy practices, or *redress* if assurances are not kept, are what assure consumers that the information practices that the entity promised are real.[151] The FTC's FIPPs did not create any enforceable legal regime.[152] They did no more than to name several broad enforcement mechanisms that could potentially be adopted: industry self-regulation, legislation that creates private remedies for consumers, or regulatory regimes that are enforced by government through possible civil and criminal sanctions.[153]

The National Institute of Standards and Technology (NIST) took a different direction to privacy protection guidance with its Privacy Risk

146. *Id.*
147. *Id.*
148. *Id.*
149. *Id.*
150. *Id.*
151. *Id.*
152. *Id.*
153. *Id.*

Management Framework.[154] Hiller describes the NIST approach as "a collection of methods to support the mitigation of risks to individuals arising from the processing of their personal information within information systems."[155] In essence, the framework is a way that any entity can "first identify its goals and obligations for privacy protection, assess its systems against these governing requirements, prioritize mitigation mechanisms, and monitor for changes."[156] Hiller argues that "what is sought by technologists is a 'repeatable and measurable method . . . for identifying, prioritizing, and mitigating privacy problems.'"[157]

The NIST approach puts the onus on organizations to create ways for individuals they interact with to exercise control over health information, mobile health data, and big data. It reflects the PCAST Report's view that consumer control is not feasible, and therefore the entity collecting and using data should bear the responsibility for implementing privacy preferences. Privacy engineering objectives adopted in NIST's Privacy Risk Management Framework are predictability, manageability, and disassociability.[158] These objectives are basic building blocks for privacy-responsive systems that "bridg[e] the gap between an [entity's] . . . goals for privacy and their manifestation in information systems."[159]

The contrast between the protections currently given to personal information in the health-care context by HIPAA and the HITECH Act compared to the industry self-regulation that is favored in more commercial settings is dramatic. Abuses in the tech sector have, however, led to pressure for more aggressive government action.[160] So we can expect the future trend to be toward more regulation in the sectors overseen by the FTC and served by NIST (which is within the Department of Commerce).

E. Reducing Restrictions on Data Use

Some policy analysts have proposed that greater access to health information data repositories, such as through HIEs, should be afforded to

---

154. NAT'L INST. FOR STANDARDS & TECH., NISTIR 8062, PRIVACY RISK MANAGEMENT FOR FEDERAL INFORMATION SYSTEMS (2015), https://csrc.nist.gov/csrc/media/publications/nistir/8062/draft/documents/nistir_8062_draft.pdf [https://perma.cc/B7VD-YCZA] [hereinafter NISTIR REPORT].

155. Hiller, *supra* note 72, at 305 (citing the NISTIR REPORT, *supra* note 154, at 4).

156. *Id.*

157. *Id.*

158. NISTIR REPORT, *supra* note 154, at 18.

159. Hiller, *supra* note 72, at 306 (citing the NISTIR REPORT, *supra* note 154, at 1).

160. *See supra* notes 72–83 and accompanying text.

researchers.[161] The essential argument shared by all of those proposals is that the value to us all of advances in knowledge regarding health and well-being is enormous, and research is essential to making those advances. Undue restrictions on the ability to conduct that research harms the nation's health. The interest in advancing health knowledge substantially outweighs potential data-sharing risks associated with record-based research. They also note that analysis of records is not interventional research; it does not introduce risks of physical harm or uncertainty about the efficacy of diagnoses or treatment.[162] All that happens is that researchers and their computers run analyses on variables already contained in health records. Where commentators differ is on how freely databases should be made accessible to researchers and what kinds of privacy protections should be required.

Some point out that the risk of unwanted disclosure is far greater in the clinical care context.[163] Depending on the extent of a patient's health needs, their records are accessed frequently by numerous caregivers, insurers, and administrators. HIPAA has made it possible for those disclosures to occur routinely and without a patient's consent or authorization. The change from paper to electronic records vastly increased the risk of unwanted disclosures, and was undertaken without patient consent.[164] In the clinical EHR context, health information of many patients is disclosed often and widely, and privacy protection has been a fundamental issue.[165] The number of researchers who would handle patients' data would be far smaller, and kept within what typically is a much narrower and more secure circle of people and computers. The risk of breaches in security—leaks, hacking, accidental or intentional improper use or disclosures—is widely considered to be far

---

161. Paula Smailes, *Data-Tech Connect: The Ethics of Research Access to Electronic Medical Record Data*, CLINICAL RESEARCHER (June 1, 2017), https://acrpnet.org/2017/06/01/data-tech-connect-ethics-research-access-electronic-medical-record-data/.

162. *Supra* notes 32–42 and accompanying text.

163. Hiller, *supra* note 72, at 257.

164. NATIONAL RESEARCH COUNCIL ET AL., FOR THE RECORD: PROTECTING ELECTRONIC HEALTH INFORMATION 20 (National Academies Press 1997), https://www.nap.edu/read/5595/chapter/1 [https://perma.cc/4SLM-KPGL].

165. *See, e.g.*, Ranjit Janardhanan, *Uncle Sam Knows What's in Your Medicine Cabinet: The Security and Privacy Protection of Health Records Under the HITECH Act*, 30 JOHN MARSHALL J. INFO. TECH. & PRIVACY L. 667, 702–03 (2014) (stating that a centralized information system is necessary for sharing information but may contribute to identity theft). *See also* Leslie P. Francis, *When Patients Interact with EHRs: Problems of Privacy and Confidentiality*, 12 HOUS. J. HEALTH L. & POL'Y 171 (2012); Daniel J. Gilman & James C. Cooper, *There Is a Time To Keep Silent and a Time To Speak, the Hard Part Is Knowing Which Is Which: Striking the Balance Between Privacy Protection and the Flow of Health Care Information*, 16 MICH. TELECOMM. & TECH. L. REV. 279 (2010); Nicolas P. Terry & Leslie P. Francis, *Ensuring the Privacy and Confidentiality of Electronic Health Records*, 2007 U. ILL. L. REV. 681 (2007).

smaller in the research setting. Keep in mind, however, that the great majority of research can be conducted with de-identified data, which enormously shrinks the risk of privacy violations in the research context.

Moreover, the societal choice to sacrifice individual privacy for the public's health and safety is already reflected in laws in every state that, for example, require disclosure to government experts of the identity of patients without their consent who contract a variety of infectious diseases, become risks on the road, or appear to have been abused.

Recent changes in law are designed to promote the flow of patients' personal health information more freely and more widely than ever before. The 21st Century Cures Act (Cures Act) contains extensive provisions that aim to "advance interoperability and support the access, exchange, and use of electronic health information."[166] Among those are provisions for creation of a Trusted Exchange Framework and Common Agreement (TECFA) intended to create a network of networks and connect authorized participants (e.g., payors, vendor networks, government agencies, individuals and HIEs) so that information can flow among these entities (once they qualify) without barriers associated with differing systems, or concerns that others in the network are withholding information (information blocking).[167] The guiding principles in developing trusted exchange among qualified health information networks (QHINs) and other organizations envisioned in the TEF (e.g., participating health information networks (HINs) and participating providers) are standardization (adherence to shared standards, policies, and procedures), transparency (openness in exchanges), cooperation and non-discrimination (even among competitors), security and patient safety, and access (by other organizations and the patient).[168]

Furthermore, the Cures Act includes provisions that are aimed at facilitating research use of EHRs.[169] It proposes that researchers may remotely access PHI if security and privacy safeguards are maintained and the information is not retained.[170] Among other steps, the Cures Act directs HHS to convene a working group to study the use of PHI for research.[171]

---

166. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program, 84 Fed. Reg. 7424-01 (proposed March 4, 2019).

167. *Id.* at 7424–31.

168. Genevieve Morris & Elise Sweeny Anthony, *21st Century Cures Act Overview for States*, The OFFICE OF THE NAT'L COORDINATOR FOR HEALTH INFO. TECH., 1, 10 (2018), https://www.healthit.gov/sites/default/files/curesactlearningsession_1_v6_10818.pdf [https://perma.cc/MX58-85UW].

169. 21st Century Cures Act, Pub. L. No. 114-255, § 13103, 130 Stat. 1033 (2016).

170. 21st Century Cures Act, Pub. L. No. 114-255, § 2063, 130 Stat. 1033 (2016).

171. *Id.*

Against that background of benefits and risks of disclosure without consent already authorized by law, and of efforts under way to further streamline the exchange of health information, worries about sharing data with researchers seem to be out of proportion to the actual benefit-to-risk ratio of the research domain.

The most extreme proposals to reduce restriction on research use of HIE data would allow all data in an HIE (other than personal identifiers, except when those are necessary for the research—in which case additional approval would be needed) to be accessible to authorized researchers.[172] Researchers would become authorized in a manner analogous to acquiring a security clearance, that is, by meeting criteria for trustworthiness in protecting the security and confidentiality of data.[173] Such researchers would then gain access to needed data without further restriction.[174] Even then, researchers would be required to abide by legal, scientific, and ethical principles governing the analysis of health records, as well as to keep data secure, including not sharing with unauthorized persons or organizations.[175] The researchers would be subject to sanctions for violations.[176]

One proposal along these lines would require all patient records to be shared with public authorities who would combine them into aggregated databases and make those available, in de-identified form, to all researchers and possibly to the general public.[177] The aim of such proposals is to create the most comprehensive repositories possible for the most thorough analyses and most accurate results: No opting out or withholding of selected data.[178]

---

172. *See* Hoffman & Podgurski, *supra* note 33 at 94.

173. Determining what body or bodies would have the power to grant or deny these authorizations, according to what criteria, and using what procedures is obviously a complicated and potentially controversial matter of its own. *See* Hill et al., *supra* note 60; Hoffman & Podgurski, *supra* note 34, at 125 (arguing that requiring patients to share their EHRs is "ethically sound"); Deven McGraw & Alice Leiter, *A Policy and Technology Framework for Using Clinical Data to Improve Quality*, 12 HOUS. J. HEALTH L. & POL'Y 137, 156–66 (2012) (describing a framework for secondary use of clinical data to be expanded under fair information practice principles); Suzanne M. Rivera, *Privacy vs. Progress: Research Exceptionalism Is Bad Medicine*, 24 HEALTH MATRIX 49, 59 (2014) (asserting that health information should be a public resource).

174. *Id.*

175. McGraw & Leiter, *supra* note 173, at 145–46.

176. For most researchers, the most punishing sanction would be suspension or termination of access to the HIE's data. *See* AHIMA, *Sanction Guidelines for Privacy and Security Violations*, J. OF AHIMA 84 (2013).

177. Marc A. Rodwin, *Patient Data: Property, Privacy & the Public Interest*, 36 AM. J.L. & MED. 586, 615 (2010).

178. *See id.*

A similar proposal strongly favoring access would create a National Health Information Network (NHIN).[179] The NHIN would provide broad access for researchers to a unified database of national health records.[180] Allowing researchers access to patients' records would be a condition of including patients' records in the database.[181] A 2007 report of "Recommended Requirements for Enhancing Data Quality in Electronic Health Records" suggests how such an NHIN could increase information accuracy and protect against fraud.[182] These include such provisions as structured and coded data, transmission integrity, patient-identity proofing, accurate linkage to claims data, ensuring traceability of EHRs, secure messaging, and making data anonymous for use in medical research.[183] The report notes that, because HIPAA permits disclosure of PHI for purposes of treatment, payment, and health care operations without patients' explicit authorization, "special consideration must be given to scenarios involving some level of access by groups other than the primary user, such as patients themselves, visiting physicians, and payers"—and, no doubt, researchers.[184]

To work, such a national network would have to begin with federal pre-emption of state health records privacy laws.[185] The chief argument for this is simple necessity. As discussed earlier, data-sharing across state lines is hampered by the bottom-up approach, wherein each HIE's databases and practices are shaped by the conflicting laws of whatever state it is based in.[186] The creation of an NHIN highly accessible to researchers would greatly facilitate the many benefits of research sketched at the outset of this article. Minimizing the risks of privacy and dignitary harm to individuals would be achieved principally by anonymization. Researchers typically do not need,

---

179. Hill et al., *supra* note 60. Among many other issues, Hill et al. also analyze at length constitutional arguments regarding whether or not the government can create an NHIN and deploy it throughout the country. *Id.* at 555–94.

180. *Id.* at 510.

181. This could be accomplished either through explicit consent (at the time of opting-in to the database) or through the operation of law (all patient data becomes part of the database with research access allowed as part of as a database feature unless a patient opts-out of the database or out of the research component, as policy and database architecture provide).

182. RTI Int'l, *Recommended Requirements for Enhancing Data Quality in Electronic Health Record Systems*, U.S. DEP'T. OF HEALTH AND HUMAN SERVS. ES-1, 3-2 (2007), https://www.rti.org/sites/default/files/resources/enhancing_data_quality_in_ehrs.pdf [https://perma.cc/3YMJ-L7RT].

183. *Id.* at ES-5.

184. *Id.* at 4-5.

185. Other factors favoring federal pre-emption include the recognition that HIEs already cross state lines, research is not an intra-state activity, and commerce benefits from research advances.

186. *See supra* text accompanying notes 43–60.

and would not normally have, access to identifying information.[187] Today, the network of networks envisioned by the Cures Act would accomplish most, if not all, of what had been suggested for a single national network (NHIN), without imposing outright federal pre-emption.

Others argue for similar but narrower access. This approach would strengthen protection of PHI by limiting researchers' access to those particular data elements needed for each particular authorized study. That more limited access would have to be overseen by one or more vigilant gatekeepers. Each proposed analysis (or set of related analyses) of data would be a research project that had been vetted by one or more appropriate entities: IRBs, Privacy Boards, funding agencies, and/or the HIEs themselves. The sources, the data, whether they are to be de-identified or otherwise, would be specified in the approved research protocols. Increasing degrees of restriction generally require increasing cost and effort by whoever (the HIE and/or other entities) is doing the gatekeeping and creating the de-identified databases.[188] But some or all of those expenses could be charged to the research projects, which would include them in the proposal budgets they submit to funding agencies. Again, departures would be subject to sanctions.

Perhaps the most astute policy proposal is one offered by Hoffman & Podgurski. They agree entirely with the ethical conclusion that the public interest in advancing health knowledge outweighs individual interests in privacy, and justifies making databases of PHI available to researchers without requiring patients' consent.[189] They argue that only interventional research ethically requires patient consent, not the statistical analysis of existing records.[190] They also note that the law already permits the sharing of de-identified data without asking patients' consent, which they support.[191]

But for Hoffman and Podgurski that is far from the end of the issue. They note that the right to grant or withhold consent offers no additional protection from unwanted disclosure to those who *do* consent.[192] Many or most patients

187. In those studies where research required multiple data draws with the same patients, the custodians of the database should be able to accommodate the needed linkage without disclosing identifying information. In those studies where researchers needed to collect additional data (not contained in the NHIN) from patients, specific informed consent would be required and assistance in contacting those patients could be provided by the database custodians.

188. That increased cost and effort suggest why gatekeeping at the outset (deciding who may be trusted with access, with requirements that those researchers limit themselves to permissible uses) would be attractive.

189. Hoffman & Podgurski, *supra* note 33, at 126 ("The common good principle supports the imposition of certain burdens on patients, namely, depriving them of choice as to whether their EHRs are accessible to researchers.").

190. *Id.* at 124.

191. *Id.* at 95.

192. Surveys suggest this will be most patients. *Id.* at 127–28.

have an interest in the security of their data, even when it is de-identified. That is the lingering problem Hoffman and Podgurski seek to remedy.[193] They urge a high degree of sensitivity to the privacy interests of patients for both ethical (protecting the security of the patients' data) and practical (political) reasons, as well as the interest of the public in better understanding the research enterprise for which their data are sought.[194]

In place of patients' consent, Hoffman and Podgurski argue for strengthening oversight of the research process.[195] They propose, first, that an ethics board with special expertise in records-based studies and information security review proposals and approve, disapprove, or require modifications.[196] Second, researchers would be obliged to sign written assurances that they will not re-identify patients and not convey records to anyone beyond themselves.[197]

Third, the ethics board would carry out continuing review of each project, more or less intensely, depending on circumstances, and require annual reports from researchers regarding data security.[198] At any point, the board could order corrective action or withdraw approval of the research, thereby terminating the study.[199] Fourth, HHS, which is responsible for enforcing HIPAA security rules, would conduct oversight of the ethics boards, including audits, which could be unannounced.[200] Fifth, HIPAA would be amended to expand the definition of "covered entities" to ensure that researchers are subject to the law's security regulations.[201] They also suggest that consideration be given to repealing the rule that exempts de-identified databases from the HIPAA Security Rule.[202]

Sixth, Hoffman and Podgurski propose a "notice and education" regime.[203] They recommend that part of the HIPAA notice given to patients by each

---

193. Hoffman and Podgurski's proposal can be regarded as a much stronger and more extensive version of a comparable proposal from the Institute of Medicine. *See generally*, Institute of Medicine, *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*, THE NAT'L ACAD. PRESS 245 (2009), https://www.nap.edu/catalog/12458/beyond-the-hipaa-privacy-rule-enhancing-privacy-improving-health-through [https://perma.cc/MT5Q-WV34]. Most notably, the IOM proposal relies heavily on pre-approval oversight (rather than ongoing oversight) and does not include a notice-and-education component. *Id.*

194. Hoffman & Podgurski, *supra* note 33, at 133–41.

195. *Id.*

196. The ethics board could be an IRB or some other appropriate entity. *See id.* at 91, 133–37.

197. *Id.* at 112–14.

198. *Id.* at 136.

199. *Id.*

200. *Id.* at 91.

201. *Id.* at 137–38.

202. *Id.* at 97, 138.

203. *Id.* at 138–41.

provider and health-care organization should inform patients of what research uses might be made of their (de-identified) health records, explaining in general terms how and under what circumstances their data might be used.[204] It probably should also explain why, that is, the value of record-based research. The education component would consist of requiring those who conduct and who promote research to initiate wide ranging educational campaigns to help patients understand why their health records might become the subject of analysis by researchers and what benefits accrue to the population from such research.[205] Ultimately, it is the education that will empower members of the public to look out for their true interests by demanding changes to policies they object to.[206]

The proposals described thus far in this section vary primarily in regard to how much they would open up access to databases to researchers and how much they would subject researchers to oversight. A different move could put more control in the hands of patients to broaden or narrow whatever of that patient's health information is made available to researchers, which is in line with the general principles described in TEFCA as well as recent ONC and CMS proposed rules on the subject.[207] This could be accomplished by offering patients a number of choices when they first encounter a given provider who makes their data available to third parties (e.g., an HIE), and periodically inviting them to change their choices if they wish. These choices could range from broad decisions whether to allow their information to be used for research at all, whether certain categories of data must be withheld, whether the information may or may not be linked back to the patient (for possible re-contact) (presumably in coded form), and more fine-grained choices, such as what kinds of research their data may be employed in (for example, subject matter), what kinds of researchers, and what kinds of funders.[208] Allowing patients such choices obviously would frustrate efforts to create comprehensive data repositories and create problems of sample bias and possible confounding of variables;[209] therefore, it is not ideal for growing

204. *Id.* at 139.

205. *Id.* at 140–41.

206. Perhaps they will insist on more support for research and fewer restrictions on researcher access.

207. Medicare and Medicaid Programs; Patient Protection and Affordable Care Act; Interoperability and Patient Access for Medicare Advantage Organization and Medicaid Managed Care Plans, State Medicaid Agencies, CHIP Agencies and CHIP Managed Care Entities, Issuers of Qualified Health Plans in the Federally-Facilitated Exchanges and Health Care Providers, 84 Fed. Reg. 7610 (Mar. 4, 2019) (to be codified at 45 C.F.R. pts. 156, 406, 407, 422, 423, 431, 438, 457, 482, 485).

208. *See* Kim et al., *supra* note 102; Grando & Schwab, *supra* note 102.

209. *See supra* notes 32–42 and accompanying text.

sound new knowledge about health and health care. But note that it still permits vastly more access by researchers to EHR databases than state statutes such as Arizona's—which requires separate, new informed consent, in an all-or-nothing approach for each project.[210]

F. Technological Solutions

The policy options described thus far consist of legal rules that seek to balance the desirability of making HIE data available for research with the protection of the privacy of individual health records through various means: greater or smaller limits on what is available, more or less gatekeeping, more or less oversight, and varying sanctions for violations. As between those two goods—research versus privacy—the debate comes down to how best to strike the balance, where to place the fulcrum.

Might there be a technological solution that can deliver more of both, rather than trading off one against the other? A solution might be possible, whereby innovative technology could enable data from a HIE or a national network of HIEs[211] to be made more fully and readily available, while at the same time enhancing the privacy and security of patients' health records.

Let's begin by imagining HIEs with staff whose principal job is to carry out analyses for outside researchers.[212] One or several researchers representing an approved research project would meet with one of the HIE's statistical analysis staff, explain the analyses that need to be conducted for their project, and the HIE staff statistician would run the analyses. Later, they would meet again, go over the output, and probably see the need for additional analyses to be run. This back-and-forth process—discuss, run analyses, discuss, run more analyses—is not unlike the way senior researchers work with their graduate students. The virtue of this process would be that the only person who has any direct access to the HIE's data repository is the employee of the HIE—someone who has access anyway by virtue of being an HIE staff member. Data never leave the HIE. Researchers receive the aggregate output from the analyses, which is what they need to complete their research. Thus, the researchers have essentially unrestricted ability to carry out their research, and the patient records are as private and secure as they would be without any research taking place.

---

210. *See supra* notes 96–102 and accompanying text.

211. The Strategic Health Information Exchange Collaborative is a national network of HIEs. STRATEGIC HEALTH INFORMATION EXCHANGE COLLABORATIVE, https://strategichie.com [https://perma.cc/LDN3-53A9].

212. Fees charged to the outside research projects would pay these staffers' salaries.

Next, let's think about replacing the HIE's data analysis staffers with software, and allow the researchers to run their own analyses. This would eliminate the cost of the staffers and the inefficiency of having to explain what statistical analyses are needed to someone who is not intimately familiar with the research project.[213]

Imagine a software interface that offers a portal through which researchers can access data. Such a research portal would enable researchers to select any variables they wished to study, analyze the data, and generate aggregate results, without ever having access to *any* individual patient records. The portal would reveal only some form of "variable view" to researchers: that is the menu of variables they have to work with in setting up their analyses.[214] Researchers would not have access to, and would not need access to, "data view" or "case view." Figure 1 illustrates the two different views. Data view exposes the raw data of each case, or patient. Variable view presents only definitions and attributes of variables. This latter view is the only one researchers would be able to access. Researchers could analyze the data without risk of any patients' identities being disclosed.[215]

As to security, the system would be as secure as before, as secure as the HIE. Separate de-identified datasets would not be created and conveyed to researchers and their computers and their networks. Only HIEs which had passed muster with TEFCA (or some other agreed-to framework), would be eligible to enter into these partnerships with researchers.

What we've described is not entirely science fiction. There are existing national efforts to achieve solutions similar through the implementation of Distributed Research Networks (DRN).[216] A DRN is a computer network in which data stewards maintain data in their own environment while allowing access through Application Programming Interfaces (APIs) and controlled network functions rather than directly integrating between computer systems or exporting datasets.[217]

---

213. On the other hand, the HIE staffer would be intimately familiar with the database and the nature of its inventory.

214. "Data view" and "case view" are terms used by some statistical analysis packages to refer to a screen that displays the data individual-by-individual. *E.g.*, *SPSS Data Editor Window,* SPSS TUTORIALS, https://www.spss-tutorials.com/spss-data-editor-window/ [https://perma.cc/NX7X-SLZD]. Figure 1 illustrates the two different views.

215. Where a research design requires linking new data or following up with patients, additional solutions are needed–such as pseudonymous coding to link additional data with existing data or obtaining patients' consent to being contacted and contributing additional data.

216. *See, e.g.*, *About*, NAT'L PATIENT-CENTERED CLINICAL RES. NETWORK, https://pcornet.org/about/ [https://perma.cc/VGS7-7WMX].

217. JEFFREY BROWN ET AL., AHRQ, DESIGN SPECIFICATIONS FOR NETWORK PROTOTYPE AND COOPERATIVE TO CONDUCT POPULATION-BASED STUDIES AND SAFETY SURVEILLANCE

For example, in 2014, the Patient Centered Outcomes Research Institute (PCORI) invested more than $250 million in the development of the National Patient-Centered Clinical Research Network (PCORnet).[218] PCORnet uses a DRN and allows researchers to ask the same question of millions of people across the country all at the same time, enabling clinical research that is faster, easier, less costly and more relevant.[219] PCORnet is comprised of a coordinating center that works as a data steward and partner networks that securely collect and store data within their own institutions.[220] Partner networks include thirteen clinical data research networks, twenty people-powered research networks, and two health plan research networks.[221] PCORnet has put in place HIT techniques to ensure the security of patient data and a governance structure to monitor and ensure that the data are used appropriately.[222] The practical value of PCORnet has been already demonstrated through multiple clinical studies.[223]

Comparative effectiveness research conducted in DRNs or across HIEs is subject to different state laws and regulations as well as institution-specific policies intended to protect privacy and security of health information.[224] Efforts to develop privacy and security policy frameworks are needed if multistate research networks are used. PCORI and the Agency for Healthcare Research and Quality's Electronic Data Methods Forum have been working on the development of frameworks for addressing governance issues at state and federal level.[225]

(2009); *The Value of an API in Healthcare*, MULESOFT, https://www.mulesoft.com/resources/api/connected-healthcare [https://perma.cc/A9HS-QSCC]. *See also* Jessica M. Malenfant et al., *Cross-Network Directory Service: Infrastructure To Enable Collaborations Across Distributed Research Networks*, LEARNING HEALTH SYS. (Jan. 3, 2019), https://onlinelibrary.wiley.com/doi/epdf/10.1002/lrh2.10187 [http://perma.cc/4VPW-A75H].

218. *PCORnet 101*, PATIENT-CENTERED OUTCOMES RES. INST., https://www.pcori.org/events/2017/pcornet-101 [https://perma.cc/4WSV-QZY3].

219. BROOKINGS INST., PCORNET: BUILDING EVIDENCE THROUGH COLLABORATION AND INNOVATION (2014), https://www.brookings.edu/wp-content/uploads/2014/01/PCORnet_Discussion_Guide_final.pdf [https://perma.cc/26KR-JWRX].

220. *Id.*

221. *The Network,* NAT'L PATIENT-CENTERED CLINICAL RES. NETWORK, https://pcornet.org/clinical-research-network/ [https://perma.cc/2WTE-V567].

222. BROOKINGS INST., *supra* note 219.

223. *Impact*, NAT'L PATIENT-CENTERED CLINICAL RES. NETWORK, https://pcornet.org/past-research-studies/ [https://perma.cc/8J8W-AX2U].

224. Katherine K. Kim et al., *Development of a Privacy and Security Policy Framework for a Multistate Comparative Effectiveness Research Network*, 51 MED. CARE S66, S66 (2013).

225. Marianne Hamilton Lopez et al., *Involving Patients and Consumers in Research: New Opportunities for Meaningful Engagement in Research and Quality Improvement*, ACADEMY HEALTH-EDM F., (2012); Joe V. Selby et al., *The Patient-Centered Outcomes Research Institute*

What are currently seen as competing interests could, with the right information technology, both be maximized, and therefore served better than either is currently.

CONCLUSION

The databases of HIEs constitute a resource of exceptional potential value if they can be accessed for expansive medical and other research. That research, and the knowledge that will be generated by it, will not begin to flow until the barriers keeping researchers from using those databases are lowered or removed.

The law could be instrumental in reducing those barriers. Federal and state health record privacy laws could and should be amended to facilitate research access without undermining real interests of patients in the privacy and security of their data. A range of suggestions have been discussed for adjusting the balance between society's interest in research advances and individuals' interest in privacy. Some beneficial legal adjustments have already been put into motion at the federal level, namely the Cures Act.

On the other hand, some state laws, such as Arizona's, are particularly problematic, placing barriers to research so high as to effectively prevent HIE databases from being used for research. Those states should consider adjusting their privacy laws to better balance the interests of research and privacy. Alternatively, consideration should be given to federal pre-emption of state laws to the extent that they unnecessarily prevent research from using existing records.

Finally, technological advances under discussion have a high potential to open up access to the data while adding essentially no additional risk to privacy and security. By facilitating the development of those information technology innovations, the law could obviate much of the complicated balancing that it otherwise needs to do to make EHRs available for research while protecting privacy and security.

(PCORI) National Priorities for Research and Initial Research Agenda, 307 J. AM. MED. ASS'N 1583 (2012).

APPENDIX 1. PROPOSED PRINCIPLES OF HEALTH DATA USE FOR RESEARCH. (DEVELOPED AT AMIA 7TH ANNUAL HEALTH POLICY MEETING.)[226]

1. Access to and use of health data should be viewed as a public good. Data should be available and 'fit-for-use,' with proper security, for appropriate purposes beyond direct patient care.

2. Health data must be as consistent, comparable, timely, accurate, accessible, complete, and reliable as possible. Users must be able to track the degree to which the data have attained these attributes. Understanding the context and provenance of the data is also critical in determining their 'fitness for use.'

3. Integration and sharing of health data that currently reside in silos are necessary for the optimal use of the data.

4. The rights and responsibilities of everyone (including patients, families, providers, researchers, payers, and organizations) involved in collecting and using health data must be understood and respected.

5. Data uses must be transparent to all, including patients and their agents.

6. The potential benefits of data use must be weighed against the potential risks and costs of loss or inappropriate disclosure of personal health information.

7. Data stewards (those who collect, maintain, aggregate, analyze, and use health data) must demonstrate that they understand and are willing to assume the responsibilities of effective stewardship in order to earn and retain the support of patients and the public. Data stewards must demonstrate that they use data appropriately and in accordance with applicable laws and regulations.

8. Data use policies should not be so binding that they restrict or prevent uses of data from emerging technologies or impede as yet unknown data sources or technologies.

226. Hripcsak et al., *supra* note 68, at 206.

9. All health care system stakeholders must continue to study the benefits and risks of new data sources and uses and to refine data use principles as needed.

APPENDIX 2. DE-IDENTIFICATION IN ACCORD WITH HIPAA[227]

De-identification of protected health information (PHI) requires removal of the following 18 identifiers of the individual and of the individual's relatives, employers, or household members:

1. Names
2. All geographic subdivisions smaller than a state, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of the Census: (1) the geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and (2) the initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
3. All elements of dates (except year) directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older.
4. Telephone numbers
5. Fax numbers
6. Electronic mail addresses
7. Social Security numbers
8. Medical record numbers
9. Health plan beneficiary numbers
10. Account numbers
11. Certificate/license numbers
12. Vehicle identifiers and serial numbers, including license plate numbers
13. Device identifiers and serial numbers
14. Web universal resource locators (URLs)
15. Internet protocol (IP) address numbers
16. Biometric identifiers, including finger and voice prints
17. Full face photographic images and any comparable images
18. Any other unique identifying number, characteristic, or code

227. OFFICE OF THE CIVIL RIGHTS, GUIDANCE REGARDING METHODS FOR DE-IDENTIFICATION OF PROTECTED HEALTH INFORMATION IN ACCORDANCE WITH THE HEALTH INSURANCE PORTABILITY AND ACCOUNTABILITY ACT (HIPAA) PRIVACY RULE 7–8 (2012).

FIGURE 1. DATA VIEW CONTRASTED WITH VARIABLE VIEW IN SPSS[228]

Data

Columns are variables

Rows are cases

IDs	Rank	Gender	Athlete	Height	Weight
20183		Male	Non-athlete	5.5	192.61
20230	Freshman	Male	Athlete	6.2	
20243	Junior	Female	Non-athlete	5	128.4
20248	Freshman		Non-athlete	5.7	153.87

Variable

Columns are variables descriptions

Rows are variables

Name	Type	Label	Values
IDs	Numeric	Unique respondent identifier	None
Rank	String	Year of undergraduate education	{Freshman, Junior, Sophomore}
Gender	String	Gender description	{Female, Male}
Athlete	String	Identification as athlete or non-athlete	{Athlete, Non-athlete}
Height	Numeric	Height in feet	None
Weight	Numeric	Weight in pounds	None

228. SPSS is a computer program enabling the user to perform a wide range of statistical tests. See *supra* note 214 for a tutorial on employing the data and variable views in SPSS.