Privacy and Informed Consent for Research in the Age of Big Data

David M. Parker*, Steven G. Pine**, and Zachary W. Ernst***

ABSTRACT

Big Data collections derived from medical records present regulatory and privacy challenges while holding significant promise for advancements in biomedical research. The growth of Big Data has been spurred by technological advances and the increasing use of electronic medical records. In this article, we explore how the concept of a rights to privacy and confidentiality for research subjects has developed, through both HIPAA and the Common Rule, as well as in the European Community's General Data Protection Regulation (GDPR). We analyze how developments in regulations governing human subjects research reflect both a heightened societal concern for individual privacy and confidentiality and a recognition that research may be of sufficient importance to society to outweigh those individual concerns. We review how new efforts to improve informed consent procedures in the Common Rule fall short and propose a more frank and straightforward approach.

*David Parker is Of Counsel within the Health Care and Higher Education practices of K&L Gates LLP, Research Triangle Park, NC.
**Steven Pine is an Associate within the Health Care practice of K&L Gates LLP, Research Triangle Park, NC.
***Zachary Ernst is an Associate within the Health Care practice of K&L Gates LLP, Research Triangle Park, NC.

Table of Contents

I. INTRODUCTION

Technical capacities to retrieve, transmit, store, and manipulate data and a concurrent push toward adoption of electronic medical record systems have fueled development and availability of Big Data collections of biomedical data. Such collections have already helped researchers achieve breakthroughs in understanding and treating disease, but their misuse can create significant privacy issues. Big Data presents enormous opportunities for the advancement of biomedical research, but unhelpful regulation creates challenges for researchers and research subjects alike. While regulation and enforcement trends appear to reflect increased vigilance about protecting the privacy and security of personal data, in some ways they also result in patients assuming that they have control over personal data that in fact they do not possess. Citizens are generally willing to compromise their privacy interests for the greater good, but that exchange should be based upon honest and informed transactions where research subjects knowingly barter for societal advances that may (or may not) also mean improvements in their own health and well-being. Evolving societal tolerance for collectivization of personal data and favorable attitudes about using biomedical data in human subjects research may

indicate that more relaxed and straightforward regulatory approaches would be acceptable, at least in the research context.

In the remainder of Part I, we define Big Data and how its use in research involving human subjects has been fueled by technological advances and the rise of electronic medical records. Part II discusses the history of human subjects research regulation in the United States and current regulations addressing privacy and consent, both in the United States and the European Union. Part III will identify challenges with using Big Data under the current regulatory system. Finally, Part IV proposes a middle ground approach that balances societal benefits with individual autonomy by making changes to the broad consent standard.

A. Defining Big Data and How it is Used in Research

"Big Data" is a label placed with some degree of imprecision on the results of increasingly sophisticated capacities to obtain, store, and manipulate data in volumes and at speeds that were previously not achievable. More than simply a huge collection of information, Big Data can be conceived as having at least two dimensions in addition to volume: velocity, or the speed of processing data; and variety, meaning that different forms of data from differing types of sources can be assimilated into one aggregation.[1] Thus, Big Data refers both to large, often heterogeneous databases as well as to the process of analyzing very large datasets.[2] Such datasets can include primary medical data harvested from electronic medical records, research data obtained from human subjects outside the scope of routine patient care, and data from nontraditional sources such as fitness trackers and grocery store receipts that provide information on lifestyle choices.[3]

Big Data allows companies and researchers, through the application of algorithms and artificial intelligence, to identify "patterns, links, behaviors, trends, identities, and practical knolwedge."[4] Such an approach presents complicated ethical challenges, and to some, Big Data can seem like Big Brother.[5] Unlike traditional research projects where one person

1. *See* Brent Daniel Mittelstadt & Luciano Floridi, *The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts*, 22 SCI. & ENGINEERING ETHICS 303, 309 (2016).

2. *Id.*

3. *See* Barbara J. Evans, *Power to the People: Data Citizens in the Age of Precision Medicine*, 19 VAND. J. ENT. & TECH. L. 243, 244 (2017) [hereinafter Evans, *Power to the People*].

4. *See* Anita L. Allen, *Protecting One's Privacy in a Big Data Economy*, 130 HARV. L. REV. F. 71, 71 (2016).

5. *Id.* at 72–74.

consents to participate in one project,[6] Big Data by its nature can affect a wider group or classification of people, and the ways in which it can be connected, repurposed, updated and used in a context entirely different from that in which it was originally obtained continue to grow as technology advances.[7] Set loose from traditional technical and financial constraints and limitations on use, re-use and repurposing, Big Data stresses preexisting ethical concepts of research and informed consent.[8] The horse not only cannot be put back in the barn, it may no longer be a horse at all.

Big Data in health care has many sources. Health care providers and major health systems have access to a robust array of clinical and payor data, both from enhanced processes for collecting and storing internal data, as well as through access to external clinical records, payor claims data, and lab, pharmacy, and provider data. In addition, health systems are also increasingly collaborative and may participate in an accountable care organization,[9] clinically integrated network,[10] or health information exchange[11] where big data is bi-directionally created and shared.[12] This bi-

6. *See* Jacob Metcalf & Kate Crawford, *Where are human subjects in Big Data research? The emerging ethics divide*, BIG DATA & SOC'Y, Jan.–June 2016, at 1, 2. This traditional concept of how people participate in research is somewhat undercut by a longstanding regulatory scheme whereby human subjects research is exempt from informed consent requirements involving existing, deidentified data because such data, once deidentified, are not deemed to involve human subjects. *Id.* at 1; *see also* 45 C.F.R. § 46.102(e)(1) (2018) (defining terms used within the policy). Moreover, in some circumstances existing identifiable data may be used for secondary research without informed consent for that particular study. *See* 45 C.F.R. § 46.104 (2018).

7. *See* Metcalf & Crawford, *supra* note 6, at 2.

8. *Id.*

9. Broadly speaking, accountable care organizations are "groups of doctors, hospitals, and other health care providers, who come together voluntarily to give coordinated high-quality care to their Medicare patients." *Accountable Care Organizations (ACOs)*, CTRS. FOR MEDICARE & MEDICAID SERVS., https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ACO/ (last updated Mar. 8, 2019).

10. A clinically integrated network is a similar concept to an ACO, as it likewise generally refers to a separate legal entity that manages an affiliation of health systems and/or physician practices that have agreed to come together to coordinate the provision of health care services. *See Accountable Care Organizations*, AM. ACAD. OF FAMILY PHYSICIANS, https://bit.ly/2GcqKcU (last visited Aug. 1, 2019). Generally, a CIN refers to a legal entity established for an affiliation with respect to individual commercial plans, versus the Medicare-aim of an ACO. *Id.*

11. The Office of the National Coordinator for Health Information Technology (ONC) refers to an "[e]lectronic health information exchange" as a mechanism to allow "doctors, nurses, pharmacists, other health care providers and patients to appropriately access and securely share a patient's vital medical information electronically—improving the speed, quality, safety and cost of patient care." *What is HIE?*, OFFICE OF THE NAT'L COORD. FOR HEALTH INFO. TECH., https://bit.ly/2W5t5iH (last updated May 1, 2019).

12. *Interoperability and Bi-Directional Health Information Exchange*, HALFPENNY TECHS (Apr. 19, 2016), https://bit.ly/2GaDFej.

directional data sharing leads to health systems often wearing multiple hats when using big data and a health system might, for example, create data, own data, receive and transfer data, manage data, and conduct research with data.

Health care providers and systems use Big Data to improve patient care through the monitoring of diseases and assisting in clinical decision-making.[13] Further, the rise of Big Data has led to a wide variety of secondary uses by groups both within and outside of the health system, including research, where it has demonstrated promise.[14] But Big Data is also capable of being used to determine whether certain groups of people are more or less likely to develop specific health conditions. This use can lead to individual or class discrimination.[15]

Secondary research uses of health care Big Data can trigger a variety of regulatory considerations.[16] A health system's use or transfer of data to other parties, including researchers, will generally require compliance with a number of state and federal laws, including the Health Insurance Portability and Accountability Act of 1996[17] (HIPAA) and the Federal Policy for the Protection of Human Subjects[18] (the Common Rule).[19]

13. *See* Anne S.Y. Cheung, *Moving Beyond Consent for Citizen Science in Big Data Health and Medical Research*, 16 NW. J. TECH. & INTELL. PROP. 15, 33 (2018).

14. "In 2016, noteworthy papers discussing secondary use of patient data focused on studying and improving the quality of clinical data, issues in sharing data, and predicting health outcomes using clinical data." D.R. Schlegel & G. Ficheur, *Secondary Use of Patient Data: Review of the Literature Published in 2016*, 26 Y.B. MED. INFORMATICS 68, 69 (2017).

15. *See* Henry T. Greely, *Informed Consent and Other Ethical Issues in Human Population Genetics*, 35 ANN. REV. GENETICS 785, 789–90 (2001).

16. The National Institutes of Health (NIH) recently issued a Strategic Plan for Data Science, observing that the wealth of data available to academic medical centers and other research enterprises through electronic health records presents both "great opportunities for advancing medical research and improving human health—particularly in the area of precision medicine—but they also pose tremendous challenges," for example, in terms of data privacy, data security, and data interoperability. *See NIH releases strategic plan for data science*, NAT'L INSTS. OF HEALTH (June 4, 2018), https://www.nih.gov/news-events/news-releases/nih-releases-strategic-plan-data-science.

17. Health Insurance Portability and Accountability Act of 1996, Pub. L. 104-191, 110 Stat 1936 (1996) (codified as amended at scattered sections of 29 and 42 U.S.C.).

18. Protection of Human Subjects, 45 C.F.R. pt. 46 (2018) ("Part 46"). The Common Rule is currently followed by twenty different federal agencies. *See Federal Policy for the Protection of Human Subjects ('Common Rule')*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2txRR8I (last updated Mar. 18, 2016) [hereinafter HHS, *Protection of Human Subjects*]. This Article focuses on the Common Rule's application to U.S. Department of Health and Human Services (HHS) requirements, contained within Part 46.

19. In addition, data that contains substance abuse records from an entity that receives some federal assistance and holds itself out as providing, or provides, substance abuse disorder diagnosis, treatment or referral for treatment, must be handled in compliance with the "Confidentiality of Substance Use Disorder Patient Records" regulations, located at 42 C.F.R. pt. 2 (2018) ("Part 2"). Part 2 records are subject to specific confidentiality provisions that are more restrictive than HIPAA regulations. *See id.* While Part 2 is beyond

Failure to comply with these laws carries significant penalties. For example, under HIPAA, improper disclosure of protected health information (PHI) or other types of noncompliance with the HIPAA Privacy Rule[20] or HIPAA Security Rule[21] can result in the U.S. Department of Health and Human Services (HHS) Office of Civil Rights (OCR) imposing corrective actions requirements[22] or civil monetary penalties.[23] Likewise, under the Common Rule, the federal government or an institutional review board (IRB) can terminate or suspend research projects for noncompliance with Common Rule requirements.[24] Finally, other local or international penalties can apply, including sanctions for violations of the General Data Protection Regulation (GDPR).[25]

HIPAA and the Common Rule each address the use of big data for research purposes. These laws define "research" as "a systematic investigation . . . designed to develop or contribute to generalizable knowledge."[26] An activity's purpose is often a key factor in determining

the scope of this paper, researchers should be sensitive to additional restrictions on this type of data. The Food and Drug Administration (FDA) also has a separate body of regulations for clinical trials under its jurisdiction. *See Clinical Trials and Human Subject Protection*, U.S. FOOD & DRUG ADMIN., https://bit.ly/2G3ukEY (last updated Apr. 22, 2019). While this paper does not encompass discussion of those regulations, it should be noted that FDA regulations have not been updated to match the Common Rule revisions and thus do not include, for example, broad consent or the HIPAA exemption. *Impact of Certain Provisions of the Revised Common Rule on FDA-Regulated Clinical Investigations*, U.S. FOOD & DRUG ADMIN., https://bit.ly/2GeIvZe (last updated Apr. 22, 2019). For a discussion of the interplay between the Common Rule and FDA regulations, the FDA released guidance regarding how recent changes may affect FDA-regulated clinical trials. *See* David M. Parker et al., *FDA Releases Guidance on How Recent Changes to the Common Rule May Affect FDA-Regulated Clinical Trials,* K&L GATES (Oct. 16, 2018), https://bit.ly/2Ku3P0K.

20. HIPAA Privacy Rule, 45 C.F.R. pts. 160, 164(A), (E) (2018); *see also The HIPAA Privacy Rule*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2DhI0Jb (last updated Apr. 16, 2015) [hereinafter HHS, *HIPAA Privacy Rule*].

21. HIPAA Security Rule, 45 C.F.R. pts. 160, 164(A), (C) (2018); *see also The Security Rule*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2yFAda4 (last updated May 12, 2017).

22. *See* 45 C.F.R. § 160.312 (2018).

23. *See* 45 C.F.R. §§ 160.400–.426 (2018). Depending on the severity of the violation, Office of Civil Rights (OCR) can impose civil monetary penalties of up to $50,000 for each violation, not to exceed $1,500,000 for identical violations during a calendar year. *See* 45 C.F.R. § 160.404(b).

24. *See* 45 C.F.R. § 46.113 (2018).

25. Sanctions for violations of the E.U. General Data Protection Regulation (GDPR) can be up to 4% of a company's global revenue or €20 Million (whichever is greater). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), art. 83, 2016 O.J. (L 119) 82–83, https://eur-lex.europa.eu/eli/reg/2016/679/oj [hereinafter GDPR].

26. 45 C.F.R. §§ 46.102(*l*) (Common Rule definition), 164.501 (HIPAA definition) (2018).

whether an activity is designed to develop or contribute to generalizable knowledge. That said, the HHS Office of Human Research Protections (OHRP), Secretary's Advisory Committee on Human Research Protections (SACHRP) has advised that activities may cross the line and become research if the design, purpose, or resultant information of the activity contributes to generalized knowledge.[27]

HIPAA applies to "covered entities"—a classification that includes health plans, health care clearinghouses, and health care providers that transmit health information electronically in connection with certain defined HIPAA transactions—as well as to business associates of covered entities.[28] HIPAA then regulates a covered entity's or business associate's use or disclosure of PHI, including establishing whether patient consent is required for the use and disclosure.[29] HIPAA applies to both research and non-research uses and disclosures of data by covered entities.

The Common Rule is designed to protect human subjects in connection with research conducted or supported by one of twenty different federal departments and agencies, including HHS.[30] The HHS Common Rule requirements are codified at 45 C.F.R. Part 46. The Common Rule applies to "human subject research"[31] conducted or supported by the federal government or otherwise covered by an HHS OHRP approved federal-wide assurance.[32]

For health systems with an affiliated research enterprise (e.g., academic medical centers), research uses of big data will often qualify as human subjects research under the Common Rule.[33] Further, while the Common Rule may not apply to privately funded research, many health systems may have developed a federal-wide assurance that will apply across all institution activities, often known as a "check-the-box" assurance.[34]

27. *See Attachment C – SACHRP Recommendations on Benchmarking*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2uTmc4T (last updated Oct. 28, 2016).

28. *See* 45 C.F.R. § 164.104(a) (2018). An ACO created as part of the Medicare Shared Savings Program, or a similar entity created to organize a CIN, will generally not meet the definition of a covered entity as these entities do not fit within the definitions of health care provider, health plan, or health care clearinghouse found at 45 C.F.R. § 160.103.

29. *See* 45 C.F.R. § 160.103 (2018) (defining "Protected health information").

30. *See* 45 C.F.R. § 46.101 (2018); *see also* HHS, *Protection of Human Subjects*, *supra* note 18.

31. Human subject research includes, *inter alia*, research using identifiable private information. *See* 45 C.F.R. § 46.102(e)(1) (2018).

32. *See* 45 C.F.R. § 46.101(a).

33. *See* 45 C.F.R. § 46.102(e)(1), (*l*); *see also* 45 C.F.R. § 46.101(a).

34. *Federalwide Assurance (FWA) for the Protection of Human Subjects*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2UrcmGG (last updated July 31, 2017).

B. Advances in Technology and the Push towards Electronic Medical Record

Computer technology developed in the 1960s and 1970s laid the foundation for the transition to an electronic medical record, but the transition was far from immediate. In 2004, a random sample of U.S. healthcare facilities found that only 13% of respondents had an electronic health record system fully implemented, while 10% did not have or did not plan to have an electronic health record system.[35]

In 2009, through the Health Information Technology for Economic and Clinical Health (HITECH) Act, the federal government funded a $27 billion incentive program for hospitals and providers to adopt electronic health records systems.[36] By March 2017, 67% of all providers reported using an electronic health record system and, by the end of 2017, approximately 90% of office-based physicians nationwide were projected to be using electronic health records.[37]

As medical records become increasingly electronic, there is a critical need for improved methods to mine the troves of data available.[38] Effective and efficient use of electronic medical records in clinical research on a large scale presents a huge opportunity for advancing research. One example of the use of the of electronic medical records in this context is eMERGE, an "NIH-funded national network organized and funded by the National Human Genome Research Institute that combines DNA biorepositories with electronic health record systems for large scale, high-throughput genetic research in support of implementing genomic medicine."[39]

Advances in technology have also opened the door to health data collected by mobile devices. The development and use of mobile devices and applications that collect health data operate in an interesting space in regard to health information regulation. HIPAA applies to traditional healthcare providers such as doctors and hospitals and, therefore, when a patient portal to an electronic health record ("EHR") or mobile app is established by a health care provider, HIPAA will generally apply. However, this is not typical, as most health apps are built by technology

35. R.S. Evans, *Electronic Health Records: Then, Now, and in the Future*, 25 Y.B. OF MED. INFO. (SPECIAL ISSUE) S48, S51 (2016).

36. Brian Schilling, *The Federal Government Has Put Billions Into Electronic Health Record Use: How Is It Going?*, THE COMMONWEALTH FUND, https://bit.ly/2Kdj2mZ (last visited Aug. 1, 2019).

37. *EHR adoption rates: 20 must-see stats*, PRACTICE INFUSION (March 1, 2017), https://bit.ly/2wwhlbY.

38. *See generally* NAT'L INSTS. OF HEALTH, NIH STRATEGIC PLAN FOR DATA SCIENCE (2018), https://bit.ly/2JeSvRn.

39. *Id.* at 18.

companies that are not HIPAA covered entities.[40] As a result, much of the health information collected by the devices and applications developed by technology companies fall outside the purview of HIPAA's protections.

Laws and regulations governing Big Data have failed to match the frenetic increases in complexity and sophistication of data gathering and management. In the United States, those efforts are complicated by the fragmented approach to regulating data and by the history and evolution of concepts of privacy and consent in regard to research uses of biomedical data.

II. HISTORY OF RESEARCH, PRIVACY, AND CONSENT IN THE UNITED STATES

The regulatory foundation for any human subjects research in the United States is informed consent, or some regulatory substitution for it. Researchers are able to use big data collections drawn from medical records where the patient has authorized or consented to such a use, the project is exempt from the need for such consent, or a reviewing body has waived the need for consent in a particular instance. In the United States, concerns about privacy and confidentiality were not, however, initial factors driving the development of laws requiring patient consent to engage in biomedical research. The primary focus on many of these early standards was on the physical safety of subjects and how to protect them from harms and abuses of research. However, as research principles have evolved, they have expanded beyond a focus on physical harm to incorporate ideas intertwined with issues of privacy and confidentiality: fairness, respect, and avoidance of reputational or emotional harm.[41] These concerns have come into sharper focus in recent years, with increasing awareness about data privacy, the value that health data can play in evolving research technologies, and growing concerns about the role of government intrusion in personal data.

A. The Early Development of Research Consent Standards

Early experiences of medical experimentation were tainted with many horrific abuses, where disadvantaged individuals were intentionally

40. Nicolas P. Terry, *Regulatory Disruption and Arbitrage in Health-Care Data Protection*, 17 YALE J. HEALTH POL'Y L. & ETHICS 143, 181 (2017).

41. For example, in Havasupai Tribe v. Ariz. Bd. of Regents, 204 P.3d 1063 (Ariz. Ct. App. 2008), the Arizona Court of Appeals held that an individual has a privacy interest in his or her genetic information and can state a claim for injury by alleging that their privacy interest was violated. *Id.* at 1076. The court stated that "[o]ne can think of few subject areas more personal and more likely to implicate privacy interests than of one's health or genetic make-up." *Id.* (quoting Norman-Bloodsaw v. Lawrence Berkeley Lab., 135 F.3d 1260, 1269 (9th Cir. 1998)).

subjected to harm in the name of science. Public awareness of these abuses spurred the development of standards for the appropriate conduct of research and fundamental principles of obtaining informed consent from research subjects.

Initial efforts at establishing widely accepted standards arose after World War II and the Nuremberg trials of doctors involved in the Nazi party's abusive medical experiments. In a verdict in the trial of Dr. Karl Brant, personal doctor of Adolf Hitler, the Nuremburg judges, together with expert medical advisors, issued ten principles essential to "Permissible Medical Experiments" which came to be known as the Nuremberg Code.[42] The Nuremberg Code's first principle is directed at informed consent, stating that "voluntary consent of the human subject is absolutely essential" to any type of research.[43] Tellingly, none of the Nuremberg Code principles directly addresses confidentiality.

In 1964, the World Medical Association, building on the principles established in the Nuremburg Code, issued the Declaration of Helsinki.[44] The Declaration has been updated seven times since being issued, most recently in 2013. The Declaration, while not carrying the weight of law, reflected a continued move in the direction of establishing enforceable standards.

The Declaration reflects a refinement from the Nuremburg Code on the issue of consent. The Nuremburg Code's directive that informed consent was "absolutely essential"[45] did not leave room for scenarios where the acquisition of consent from an individual was impractical or impossible. The Declaration, in contrast, directs researchers to obtain consent if at all possible and further to take "every precaution" to protect personal information and the privacy of research subjects.[46] The Declaration also prescribes multiple standards to develop the parameters of what can be considered informed consent; for example, requiring consent to be voluntary, requiring an adequate discussion of the study and its risks, and ensuring that the individual is aware of the right to refuse to participate without any reprisal.[47]

42. THE NUREMBERG CODE (1947), *available at Laws Related to the Protection of Human Subjects,* NAT'L INSTS. OF HEALTH (Feb. 2, 2005), https://bit.ly/2YH73nz (last updated June 16, 2009)*; see also Nuremberg Code,* U.S. HOLOCAUST MEM'L MUSEUM, https://bit.ly/2Iaippg (last visited Aug. 1, 2019).

43. *See* THE NUREMBERG CODE, *supra* note 42.

44. World Med. Ass'n, [WMA], *Declaration of Helsinki: Recommendations guiding doctors in clinical research* (1964), https://bit.ly/2Ta8Lsz (amended 2013).

45. *See* THE NUREMBERG CODE, *supra* note 42.

46. World Med. Ass'n, [WMA], *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects* ¶¶ 24–25 (7th ed. 2013), https://bit.ly/2rJdF3M.

47. *Id.* ¶¶ 25–32.

Notably, while not part of the initial Declaration, recent amendments specifically address medical research using identifiable data, biobanks, and repositories. The Declaration now directs researchers to seek informed consent for the collection, storage, and/or reuse of such data except in "exceptional situations" where consent is impractical and "only after consideration and approval of a research ethics committee."[48]

B. Tuskegee and the Belmont Report

The United States revolutionized its research standards in the 1970s after the public learned of the unethical Tuskegee research. This research, conducted in Macon County, Alabama on African American men with syphilis, lasted from 1932 to 1972.[49] Through these studies, research participants were withheld treatment for syphilis for years or even decades after a cure was available to allow researchers to study the long-term effects of syphilis.[50]

Following the public outcry surrounding the revelation of the Tuskegee experiments, Congress passed the National Research Act, which was signed into law on July 12, 1974.[51] This Act established the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, which on April 18, 1979, released the Belmont Report.[52] The authors of the Belmont Report identified three basic principles that were found to be particularly relevant to human subjects research: (i) Respect for Persons; (ii) Beneficence; and (iii) Justice.[53]

48. *Id.* ¶ 32.

49. Tuskegee was not the only instance of depriving research subjects of effective treatment in the United State post-World War II. *See generally* Michael A. Rodriguez & Robert Garcia, *First, Do No Harm: The US Sexually Transmitted Disease Experiments in Guatemala*, 103 AM. J. PUB. HEALTH 2122 (2013); PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, "ETHICALLY IMPOSSIBLE" STD RESEARCH IN GUATEMALA FROM 1946–1948 (2011), *available at* https://bit.ly/2IL0BBM [hereinafter PCSBI, STD RESEARCH]. From 1946 through 1948, a series of immoral and unethical research experiments were undertaken by the United States government, with the cooperation of the Guatemalan authorities, on more than 5,128 Guatemalan individuals, who were intentionally infected with bacteria that causes sexually transmitted diseases, including syphilis and gonorrhea, without their knowledge or consent. *See* Rodriguez & Garcia, *supra*; *see also* PRESIDENTIAL COMM'N, *supra*. The public was not aware of these experiments at the time of the Belmont Report, as these studies only came to light in May 2010 when the original research records were discovered and made public. *See* PCSBI, STD RESEARCH, *supra*, at 4.

50. *See, e.g.*, *U.S. Public Health Service Syphilis Study at Tuskegee*, CTRS. FOR DISEASE CONTROL & PREVENTION, https://www.cdc.gov/tuskegee/index.html (last updated Dec. 14, 2015).

51. National Research Act of 1974, Pub. L. No. 93-348, 88 Stat. 342 (1974).

52. OFFICE OF THE SECRETARY, U. S. DEP'T HEALTH, EDUC. & WELFARE, THE BELMONT REPORT (1979), https://bit.ly/2VKZgQq.

53. *Id*. at 4.

Respect for Persons is described as a conjunction of at least two ethical foundations, "that individuals should be treated as autonomous agents" and "that persons with diminished autonomy are entitled to protection."[54] Beneficence, the quality or state of doing good, encompasses an obligation to act in a kind and charitable manner toward people. The principle of Beneficence was intended to apply not just to individual studies, but to the "entire enterprise of research," as a caution against causing harm in the name of science.[55] The principle of Justice is concerned with how the benefits—and harms—of research should be distributed among society. The Belmont Report describes how justice has its roots in equality, what considerations justify departure from equality, and how to distribute the benefits and burdens of research.[56] The Belmont Report sought to use these principles to "provide an analytical framework" through which ethical concerns regarding research could be resolved.[57] It also introduced a concept of applying higher standards when research involves public funds.[58]

Applying these principles, the Belmont Report describes a framework of informed consent requirements for medical research. The authors of the Belmont Report acknowledged that there was an ongoing debate of the nature and possibility of informed consent; however, the Report states that respect for persons demands that individuals should have the opportunity to choose what shall or shall not happen to them.[59] Researchers should explain "the research procedure, their purposes, risks and anticipated benefits, alternative procedures (where therapy is involved), and a statement offering the subject the opportunity to ask questions and to withdraw at any time."[60] This commitment to a respect for persons, their rights, and their ability to control their involvement in research provides an introduction into the role of privacy and confidentiality into the informed consent process.

54. *Id.* at 4–5.

55. *Id*. at 5.

56. For example, justice is concerned with ensuring vulnerable or disenfranchised populations are not systematically selected "simply because of their easy availability, their compromised position, or their manipulability." *Id.* at 5–6.

57. *See id.* at 3.

58. The Belmont Report notes that with research there is not one common approach to how much information, or what sort of information an individual should be provided in connection with a research study. However, the standards in the Report provide certain frameworks, including that a researcher should never withhold information about risks, and always give truthful answers to direct questions about research; the information provided must be comprehensible and should be tailored to the audience receiving it; and a valid consent must be voluntary, and without coercion and undue influence. *Id*. at 4–9.

59. *Id.* at 6.

60. *Id*.

C. The Common Rule - Establishing Uniform Federal Agency Standards for Research

If the Belmont Report announced the government's principles on research ethics, the Common Rule created a uniform set of regulatory requirements and standards that apply to federal agencies involved with research. After the Tuskegee abuses, the Office for Protection from Research Risk (OPRR) was established, a predecessor to OHRP.[61] OHRP is tasked with protecting human subjects in biomedical and behavioral research and providing leadership to federal agencies that conduct or support such research.[62]

In 1991, directed by the leadership of OPRR, fifteen Federal departments and agencies codified, in separate regulations, a united set of human subject research protections known as the "Common Rule."[63] For the first time, these agencies took a consistent, directed approach to applying protections to human subject research.

As noted, the Common Rule applies to all human subjects research[64] conducted, supported, or otherwise subject to regulation by the federal government or otherwise covered by an OHRP approved federal-wide assurance.[65] Consent is at the heart of the Common Rule's requirements. The Common Rule imposes a series of requirements on non-exempt research, including, for example, IRB review and approval of research[66] and informed consent requirements.[67]

D. Development of Rights of Privacy and Confidentiality in Research

However, privacy and confidentiality were still not the focus of the Common Rule. Although the concept of a privacy and the right to be left alone was contemplated as a Constitutional right at least as early as William Brandeis and Samuel Warren's influential 1890 Harvard Law review article, "The Right to Privacy,"[68] enforcement of privacy rights of

61. *See OHRP History*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://www.hhs.gov/ohrp/about-ohrp/history/index.html (last updated Mar. 15, 2016).

62. *See* Notice Statement of Organization, Functions, and Delegations of Authority, 45 Fed. Reg. 37136 (June 13, 2000) (outlining the full responsibilities of the OHRP).

63. *See* HHS, *Protection of Human Subjects*, *supra* note 18.

64. Human subject research includes, *inter alia*, research using identifiable private information. *See* 45 C.F.R. § 46.102(e)(1) (2018).

65. 45 C.F.R. § 46.101(a) (2018).

66. 45 C.F.R. § 46.109 (2018).

67. 45 C.F.R. § 46.116 (2018).

68. Frequently referred to as one of the more influential writing on the topic of privacy rights, even in the 19th century Warren and Brandeis recognized the growing role of technology played at creating an open society where "what is whispered in the closet shall

individuals was generally linked to a collection of various common law court decisions and state law requirements for many years.[69]

The introduction of federal standards related to privacy in health care matters came with the Health Insurance Portability and Accountability Act of 1996, (HIPAA), signed into law on August 21, 1996.[70] In accordance with requirements in the 1996 law, on December 28, 2000, HHS finalized its "Standards for Privacy of Individually Identifiable Health Information" (the Privacy Rule).[71]

The overarching purpose of the Privacy Rule is to develop standards and requirements related to the electronic transmission of health information, with a particular focus on the privacy of an individual's health information.[72] The Privacy Rule actually protects confidentiality, not privacy. Privacy is the right to be left alone—in this context, to choose not to share one's personal information with anyone. For medical data, privacy has never been an attainable goal. People have always been forced to divulge highly sensitive information to treating professionals, because doing so was the only way to secure adequate treatment. Confidentiality—the right to prevent further disclosure of personal information one has chosen to divulge to another—underscored the patient-physician relationship long before being enshrined in statutes and regulations.

Thus, unlike the Common Rule, HIPAA's central purpose is privacy and confidentiality, not the protection of research subjects. That said, while not the main thrust, HIPAA is also concerned with the protection of data for use in research.[73] Around the time that the Privacy Rule was being developed, the U.S. Government Accountability Office noted that

be proclaimed from the house-tops." Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 195 (1890).

69. *See generally* Daniel J. Solove, *A Brief History of Information Privacy Law*, GW LAW SCHOLARLY COMMONS (2006), https://bit.ly/2w0mfMF.

70. Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (1996).

71. Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. 82462 (Dec. 28, 2000) (codified at 45 C.F.R. pts. 160, 164). HHS subsequently published its "Standards for the Protection of Electronic Protected Health Information" (the Security Rule) in 2003, *see* Health Insurance Reform: Security Standards, 68 Fed. Reg. 8334 (Feb. 20, 2003) (codified at 45 C.F.R. pts. 160, 162, 164), and a rule implementing enforcement standards in 2006, *see* HIPAA Administrative Implication: Enforcement, 71 Fed. Reg. 8390 (Feb. 16, 2006) (codified at 45 C.F.R. pts. 160, 164). The Privacy Rule has subsequently been amended several times, most notably through significant modifications published in 2002, *see* Standards for Privacy of Individually Identifiable Health Information, 67 Fed. Reg. 53182 (Aug. 14, 2002) (codified at 45 C.F.R. pts. 160, 164), and with the final Omnibus Rule that implemented provisions from the Health Information Technology for Economic and Clinical Health (HITECH Act), *see* Modifications to the HITECH Act, 78 Fed. Reg. 5566 (Jan. 25, 2013) (codified at 45 C.F.R. pts. 160, 164).

72. *See* Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. at 82463.

73. *Id.* at 82691.

Common Rule protections were not adequately addressing issues of privacy and confidentiality in the research setting.[74] Hence, one of the tenants of HIPAA driving the development of the Privacy Rule was establishing and protecting the rights individuals have in connection with their individually identifiable health information, which includes data use in research settings.[75]

As a result, there is significant overlap in requirements between the Common Rule and HIPAA, particularly in settings where research is primarily or exclusively based on the use of health data, rather than direct involvement with a research subject. In research involving health data, concerns about privacy tend to be more salient than concerns of physical harm to an individual. As HHS has stated: "informed consent laws place limits on the ability of other persons to intrude *physically* on a person's body. Similar concerns apply to intrusions on *information* about the person."[76]

Generally, a HIPAA covered entity may not use or disclose PHI without informed consent from the patient in form of a valid written authorization[77] unless an exception applies or the use or disclosure is otherwise permitted under the Privacy Rule.[78] For example, a covered entity may use PHI for its own treatment, payment, and health care operations without obtaining patient consent.[79] In addition, HIPAA provides a number of different exemptions where PHI may be used or

74. *See* U.S. GOV'T ACCOUNTABILITY OFFICE, GAO/HEHS-99-55, MEDICAL RECORD PRIVACY: ACCESS NEEDED FOR HEALTH RESEARCH, BUT OVERSIGHT OF PRIVACY PROTECTIONS IS LIMITED 13, 16 (1999), https://www.gao.gov/assets/230/226921.pdf.

75. *See* HHS, *HIPAA Privacy Rule*, *supra* note 20.

76. Comments related to informed consent drew the largest complement of the tens of thousands of comments submitted regarding the Privacy Rule proposed rule. *See* Standards for Privacy of Individually Identifiable Health Information, 65 Fed. Reg. at 82464.

77. As with informed consent under the Common Rule, a valid patient authorization under HIPAA must be written in plain language and contain several core elements. 45 C.F.R. §164.508(c) (2018) (highlighting that in addition to the core elements, a valid authorization must contain statements adequate to place the individual on notice of: (i) the individual's right to revoke the authorization in writing; (ii) the ability or inability to condition treatment, payment, enrollment, or eligibility for benefits on the authorization; and (iii) the potential for information disclosed pursuant to the authorization to be subject to re-disclosure by the recipient and no longer be protected by HIPAA requirements).

78. *See* 45 C.F.R. § 164.508(a). Also note that on December 15, 2017, the Office for Civil Rights (OCR) released further guidance on obtaining an individual authorization for the use and disclosure of PHI. *See generally Guidance on HIPAA and Individual Authorization of Uses and Disclosures of Protected Health Information for Research*, U.S. DEP'T OF HEALTH & HUMAN SERVS., OFFICE FOR CIVIL RIGHTS (June 2018), https://bit.ly/2HQ7sI4.

79. A covered entity can also disclose PHI for another covered entity's treatment or payment operations, and, in certain circumstances, its health care operations purposes. Other HIPAA requirements may apply, such as requiring that only the minimum necessary amount of PHI for the intended purposes be disclosed. *See* 45 C.F.R. § 164.506 (2018).

disclosed without patient authorization, including uses and disclosure for research.[80] Specifically, for research proposals a covered entity can seek an alteration or waiver of informed consent requirements from a HIPAA Privacy Board or IRB.[81]

E. A Growing Harmony of Privacy and Consent—Revisions to the Common Rule and Applicability to Big Data Research

The Common Rule underwent major modification as a result of a 2017 final rule[82] that, after a series of delays, went into effect on January 21, 2019.[83] These were the first major changes to the Common Rule since it was issued in 1991. These revisions were originally scheduled to take effect in 2018, and are thus referred to as the "2018 Regulations."[84] In the preamble to the final rule, HHS describes that changes were needed given dramatic changes to human subject research that have occurred since 1991, such as evolving health record technologies, the integration of multiple types of data, the Internet, the Human Genome Project, and the corresponding development of precision medicine and genomic sequencing.[85]

The preamble to the final rule also notes that research continues to grow outside of the biomedical research setting and into clinical care settings, where research and medical data are combined.[86] As a result, while the Belmont Report is cited as a continued major influence on human subjects research in the United States, OHRP recognizes that the nature of risk and benefits described in the Belmont Report has evolved, particularly as more studies involve secondary analysis of data, rather than direct involvement with research subjects.[87] In keeping with that recognition, the 2018 Requirements include a provision that the federal government must consult with "appropriate experts (including experts in data matching and re-identification)" periodically—and no less than every

80. 45 C.F.R. § 164.512(i) (2018) (highlighting specific uses and disclosures for research purposes).

81. *Id*. (listing specific criteria that an IRB or privacy board must document).

82. Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. 7149 (Jan. 19, 2017) (codified at 45 C.F.R. pt. 46 and other scattered sections).

83. Federal Policy for the Protection of Human Subjects: Six Month Delay of the General Compliance Date of Revisions, 83 Fed. Reg. 28497, 28497 (June 19, 2018) (codified at 45 C.F.R. pt. 46 and other scattered sections).

84. *See Terminology: Terms Related to the Revised Common Rule,* U.S. DEP'T OF HEALTH & HUMAN SERVS. https://www.hhs.gov/ohrp/regulations-and-policy /regulations/terminology/index.html (last updated Aug. 17, 2018).

85. Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. at 7151.

86. *Id.*

87. *Id.*

four years—to ascertain how data or samples may be identifiable.[88] OHRP states that this process responds to the growing volume of information being generated and shared in research "and evolving technology that can ease and speed the ability to re-identify information . . . previously considered nonidentifiable."[89]

The 2018 Regulations reflect a heightened sensitivity to the need for truly informed consent by research subjects. Informed consent for primary, patient-interaction research must now specify whether the data collected will be de-identified and used for other research in the future.[90] In addition, several new informed consent elements were introduced, requiring a notice about the possibility of commercial profit, whether clinically relevant research results will be returned to the subjects, and whether research activities will or might include whole genome sequencing.[91] Further, some standards under existing elements were revised. For example, the degree of information provided to a prospective subject must now be what "a reasonable person would want to have in order to make an informed decision about whether to participate" in a study.[92] Information is also required to be presented in a manner that "facilitates an understanding of why one might, or might not, want to participate."[93] OHRP has stated that its goal is to have what it recognizes to be complicated information distilled in a way that is easier for more people to understand.[94]

Conversely, the 2018 Regulations make a number of changes that indicate an effort to unclutter the path to conducting secondary research on Big Data collections. One key new exemption looks to harmonize requirements between the Common Rule and HIPAA for certain secondary research uses regulated under HIPAA (the "HIPAA Exemption").[95] The HIPAA Exemption permits the secondary research use of identifiable private information when: (i) the research involves only information collection and analysis involving the investigator's use of

88. 45 C.F.R. § 46.102(e)(7)(i) (2018).
89. Federal Policy for the Protection of Human Subjects, 82 Fed. Reg. at 7169.
90. 45 C.F.R. § 46.116(b)(9) (2018).
91. *Id*. § 46.116(c)(7)–(9).
92. *Id*. § 46.116(a)(4).
93. *Id*. § 46.116(a)(5)(i).
94. *Revised Common Rule Q&As*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2W2zDyH (last updated July 30, 2018).
95. 45 C.F.R. § 46.104(d)(4)(iii) (2018). Other exemptions include secondary research using publicly available information; secondary research using information recorded in a way that the investigator cannot readily ascertain the identity of the individuals; and secondary research conducted by or on behalf of a federal entity and involving the use of federally generated non-research information, provided that the original collection was subject to specific federal privacy protections and continues to be protected. *Id.* § 46.104(d)(4)(i), (ii), (iv).

identifiable health information; and (ii) that use is regulated under the HIPAA Privacy Rule for the purposes of "health care operations" or "research" as those terms are defined at 45 C.F.R. § 164.501, or for "public health activities," as described under 45 C.F.R. § 164.512(b).[96] As a result, for covered entities that are governed by HIPAA, or for their collaborators, the majority of proposed secondary research activities involving identifiable information will not require additional Common Rule compliance, as the HIPAA Exemption would apply. That said, recent recommendation from SACHRP clarified that the HIPAA Exemption does not apply to research involving identifiable biospecimens or genetic information (as opposed to identifiable private information).[97] SACHRP reasons that the HIPAA Exemption was intended to apply only to "information," and not human samples, as otherwise the HIPAA Exemption would "subvert the greater protection afforded to identifiable biospecimens under the modernized Common Rule."[98]

Another means by which the Common Rule attempts to facilitate the conduct of secondary research is through the introduction a new "broad consent" mechanism that allows institutions to seek upfront, broad consent from patients to permit future secondary research. Under the pre-2018 Requirement, an informed consent document had to be study-specific.[99] However, under the new broad consent mechanism—rather than requiring a specific description of the research to be conducted with an individual's data—a consent form can provide "a general description of the types of research that may be conducted" with the data "sufficient . . . such that a reasonable person would expect the broad consent would permit the types of research conducted."[100] As discussed in more detail in Part IV, while broad consent presents new options and flexibilities to research organizations to engage in secondary research, a number of significant limitations create doubt about how widely it will be used.

F. The EU's Retreat toward Individual Prerogative

While in the United States the regulatory trends have been toward greater flexibility to conduct secondary research using Big Data, the European Union (EU) has attempted the opposite tack. The EU Parliament adopted the General Data Protection Regulation (GDPR) in 2016, and it became fully implemented as law in all Member States on May 25,

96. *Id.* § 46.104(d)(4)(iii).

97. *See Attachment B – Interpretation Revised Common Rule Exemptions*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2HEnhDI (last updated Nov. 19, 2018).

98. *Id.*

99. *Attachment C – Recommendations for Broad Consent Guidance*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2QaGbFY (last updated Aug. 2, 2017).

100. 45 C.F.R. § 46.116(d) (2018).

2018.[101] The GDPR establishes a framework to protect the privacy and personal data of individuals and is applicable to the European Economic Area ("EEA").[102]

The GDPR greatly expands an individual's rights regarding personal information in a broad range of circumstances and focuses on the protection of "personal data," which is defined as any "information relating to an identified or identifiable natural person ('data subject.')."[103] Data subjects are identifiable if they can be directly or indirectly identified, especially by reference to "an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person."[104] GDPR prohibits the processing of personal data without a lawful basis, which may be established where the data subject has given consent to the processing.[105] For consent to be valid, the consent must be freely given, specific, informed, and unambiguous.[106] Significantly for this discussion, the GDPR imposes more stringent requirements on "special categories" of Personal Data, which includes health data.[107]

While Personal Data is defined broadly, it is possible for data to be de-identified to the point of anonymization, in which case GDPR requirements no longer apply.[108] The GDPR standard for anonymization is far more difficult to satisfy than HIPAA's de-identification safe harbor, as GDPR requires an evaluation of "all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly."[109] For example, while one data point containing an individual's information may not be sufficient to identify the individual, data will not be considered de-identified if it

101. GDPR, *supra* note 25, at 87.

102. *Id.* The EU's retreat toward individual prerogative related to data privacy has ripple effects that extend out over the Atlantic. Also, as noted by HHS, "the GDPR will apply directly to, and will directly regulate, much of the U.S.-based use and processing of personal data that have been collected in the EEA for clinical and other research purposes." The problems with compliance "confront U.S.-based researchers, institutions, research funders (such as the NIH), and industry sponsors of research, including private pharmaceutical, biotechnology and medical device companies, as they seek to use personal data collected at research sites based in the EEA and transferred to the U.S." *Attachment B – European Union's General Data Protection Regulations*, U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2JtILpZ (last updated Apr. 13, 2018).

103. GDPR, *supra* note 25, at 33.

104. *Id*.

105. *GDPR Key Issues: Consent*, INTERSOFT CONSULTING, https://gdpr-info.eu/issues/consent/ (last visited Aug 1, 2019).

106. GDPR, *supra* note 25, at 34.

107. *Id.* at 6, 38–39.

108. *Id.* at 5, 39.

109. *Id.* at 5.

could be used in conjunction with another data point to identify the individual. As a result, it is difficult to accurately determine whether a particular data set has been de-identified to the extent that it may be considered anonymized, especially when the data in question is health data, which may be more likely to be individually identifiable than other data.

GDPR's adoption has caused nervousness among researchers because, in addition to the aspects discussed above, GDPR provides many individual rights, notably including the "right to be forgotten," which confers on individuals the right to compel erasure of their personal data without undue delay.[110] Further, an independent European advisory body on data protection and privacy has noted that "if a controller chooses to rely on consent for any part of the processing, they must be prepared to respect that choice and stop that part of the processing if an individual withdraws consent."[111]

However, alarm in the research community about the effects of GDPR may prove to be unwarranted. GDPR allows processing of special category personal data, such as health data, even in the absence of explicit consent where

> processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.[112]

Moreover, and quite significantly for secondary research situations, where the lawful purpose for processing personal data is for scientific research, a data subject's right to have Personal Data erased by a controller or processor may be derogated where exercise of that right is "likely to render impossible or seriously impair the achievement of the objectives of that processing."[113]

110. *Id.* at 43–44.

111. *Article 29 Working Party Guidelines on consent under Regulation 2016/679*, at 23, WP259 rev.01 (Apr. 10, 2018), https://bit.ly/2AnooDX; *see also* European Commission Press Release, The Article 29 Working Party Ceased to Exist as of 25 May 2018 (June 11, 2018), https://bit.ly/2wbEJda.

112. GDPR, *supra* note 25, at 39.

113. *Id.* at 44.

III. CHALLENGES WITH CURRENT REGULATIONS FOR BIG DATA RESEARCH

Big Data has been shown to be useful in advancing biomedical research.[114] For researchers, especially at academic medical centers where enormous quantities of useful patient data are readily available, getting past regulatory hurdles to secure access to big databases is almost always going to be worth the effort. Except where an institution or its IRB is especially risk averse, proponents of establishing large research databases will encounter few impediments, and project-specific proposals to study those databases will be approved. Nonetheless, legal and ethical problems persist.

First, and fundamentally, patients expect that their right to privacy will be respected, though attitudes about privacy are complicated and evolving. The Congressional Findings and Statement of Purpose to the Privacy Act of 1974 observes that "the privacy of an individual is directly affected by the collection, maintenance, use, and dissemination of personal information" and that "the increasing use of computers and sophisticated information technology, while essential to the efficient operations of the Government, has greatly magnified the harm to individual privacy that can occur from any collection, maintenance, use, or dissemination of personal information."[115] Since the Privacy Act was written, the increased use of computers and sophisticated information technology has developed at a pace that surely was not anticipated by most of society in 1974. In response to the ever-increasing magnitude of harm that can occur from misused personal information, society's opinion on the protection of personal information has evolved and become nuanced over the past few decades.

Yet as privacy issues have increased with each passing decade, the United States has opted to address privacy concerns through piecemeal regulatory schemes that protect only certain types of information, rather than through comprehensive legislation such as GDPR. This approach has led to a collection of narrow federal laws that cover many types of personal information, including health information, school records, and financial records, among others.[116] As noted by Allan Friedman, a great example of the United States' "patchwork response" to the public's privacy concerns is the protection of video-rental records.[117] In response to a newspaper that revealed Supreme Court nominee Robert Bork's video-rental history,

114. Cheung, *supra* note 13, at 16–17.

115. *Privacy Act of 1974 and Amendments*, ELEC. PRIVACY INFO. CTR., https://www.epic.org/privacy/laws/privacy_act.html (last updated Feb. 21, 2008).

116. *See* Solove, *supra* note 69, §§ 1:4–1:5.

117. Jonathan Shaw, *Exposed: The erosion of privacy in the Internet era*, HARV. MAG., Sept.–Oct. 2009, at 38, 40–41, http://www.harvardmag.com/pdf/2009/09-pdfs/0909-38.pdf.

Congress passed a video-rental records protection law, and for nearly 10 years video-rental records had stronger privacy protections than financial records or medical records.[118] Moreover, even though the U.S. provides protection for health data, this protection is not absolute as "[m]uch of the health-related information generated today is not regulated by [HIPAA]."[119]

This panoply of privacy laws protecting personal information, combined with massive data breaches, has not resulted in a society that is confident in the protection of personal information. In 2017, a survey found that roughly half of Americans (i) are "not confident at all" or "not too confident" that the federal government is able to keep their personal information safe and (ii) believe that their personal information is less secure in 2017 than it was five years prior.[120] This lack of confidence is likely related to the fact that 64% of Americans report personally experiencing a "major data breach."[121]

Additionally, 93% of adults agreed that being in control of who can access their personal information is "[v]ery important" or "[s]omewhat important."[122] The same poll found that 90% of adults agree that controlling what information is collected about them is "[v]ery important" or "[s]omewhat important."[123] Public opinion polls show that the public understands the importance of protecting their personal information—but only 50% feel confident that they can understand requests to use their personal information,[124] and 91% of Americans "agree" or "strongly agree" that people "have lost control over how personal information is collected and used by" various entities.[125]

Moreover, in regard to health information technology ("health IT"), polling has shown a favorable shift in public opinion. In the early 2000s, respondents were concerned about exposure of private medical

118. *Id*.

119. HEALTH IT POLICY COMM., PRIVACY & SEC. WORKGROUP, HEALTH BIG DATA RECOMMENDATIONS 4 (Aug. 2015), http://bit.ly/2YJh3cq.

120. Aaron Smith, *Americans and Cybersecurity*, PEW RESEARCH CTR. (Jan. 26, 2017), https://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/.

121. *Id*.

122. Mary Madden & Lee Rainie, *Americans' Views About Data Collection and Security*, PEW RESEARCH CTR. (May 20, 2015), https://pewrsr.ch/2IUX9qo.

123. *Id*.

124. *Americans conflicted about sharing personal information with companies*, PEW RESEARCH CTR. (Dec. 30, 2015), http://pewrsr.ch/1JfqIhU.

125. Mary Madden, *Public Perceptions of Privacy and Security in the Post-Snowden Era*, PEW RESEARCH CTR. (Nov. 12, 2014), https://www.pewinternet.org/2014/11/12/public-privacy-perceptions/.

information and privacy risks associated with EMRs.[126] In a 2005 study seeking views on the prospect of medical records being housed on a nationwide system of EMRs, 70% of respondents were concerned that personal medical information could be leaked due to lack of data security, 69% expressed concern that there could be sharing of medical information without the patient's knowledge, and 69% were concerned that computer systems would lack sufficient data security.[127]

However, by 2011, the public had gained faith in health IT.[128] A 2011 paper shows that 78% of those surveyed favored the use of EMRs in doctor's offices and 78% also believed that EMRs were likely to improve healthcare.[129] While 48% of respondents indicated that they were "very concerned" about the privacy of medical records, a majority of respondents (68%) believed that EMRs were "very or somewhat secure, and 64 percent . . . agreed or strongly agreed that the expected benefit of EMRs outweigh[ed] potential risks to privacy."[130]

In addition to respect for privacy—or confidentiality—the research community should be concerned with preserving trust by ensuring that patients know whether and how their private information may be used in research. Existing mechanisms don't get us there.

First, as discussed in the Introduction, the concept of informed consent itself may be stretched beyond its usefulness where Big Data is concerned. Moreover, under the Common Rule, it has long been the case that secondary use of big databases can occur regardless of consent where the data are de-identified or where the IRB grants a waiver of informed consent.[131] (As discussed below, de-identification can in some ways be considered a fiction that provides false assurance to investigators and the public that confidentiality will be preserved.) An IRB may waive all or some of the required elements of informed consent where the research is of "no more than minimal risk to the subjects," "could not be practicably carried out without the . . . waiver," and the waiver "will not adversely affect the rights and welfare of the [research] subjects."[132] Thus, while it

126. Daniel S. Gaylin et al., *Public Attitudes about Health Information Technology, and Its Relationship to Health Care Quality, Costs, and Privacy*, 46 HEALTH SERVS. RES. 920, 922 (2011).

127. *See Survey: How the Public Views Privacy and Health Research*, FED. TRADE COMM'N (Mar. 2008), https://bit.ly/2LWR2oE; *see also* INST. OF MED., BEYOND THE HIPAA PRIVACY RULE: ENHANCING PRIVACY, IMPROVING HEALTH THROUGH RESEARCH 78–81 (Sharyl J. Nass et al. eds., 2009), https://bit.ly/2IODrsp.

128. Gaylin et al., *supra* note 126, at 920.

129. *Id*. at 925.

130. *Id*. at 928.

131. 45 C.F.R. § 46.116(f) (2018).

132. *Id.* In addition, where the research involves using identifiable information, the IRB must find that the research could not practicably be carried out without using the

is surely the case that IRBs usually make the right decisions about waiver and are careful to ensure that the waiver will not adversely affect the research subjects, the fact remains that the decision to allow the research is divorced from knowledge or permission of the research subjects.

The new HIPAA Exemption introduces an interesting wrinkle to the ethical considerations in Big Data research. As is explained in Part II above, the exemption provides that secondary research involving identifiable information may proceed without consent where the use is regulated under the HIPAA Privacy Rule for the purposes of "health care operations," "research," or for "public health activities and purposes."[133] Thus, identifiable information may be used in research where an authorization was signed simultaneously with the consent for treatment, meaning that the patient met with a provider motivated by two somewhat competing purposes, given that conducting research is an act taken on behalf of society and not of the individual research subject. Alternatively a provider's Notice of Privacy Practices, required under the Privacy Rule, may state simply that the provider "may use and share [the patient's] information for health research."[134] Beyond that information, a patient is unaware of research ever happening if the researcher is able to obtain a waiver of authorization—a common occurrence.

Broad consent, while admirably geared towards fuller disclosure, is unlikely to be used as currently structured, for at least two reasons. First, the information required to be disclosed to a patient may be difficult to ascertain. The research team must disclose fairly detailed information about future conditions that cannot be known, such as "[a] general description of the types of research that may be conducted" with the patient's identifiable information, the identifiable information that may be used, whether sharing of the information might occur, what types of institutions or researchers might be allowed to use the information, and the

information in an identifiable format. "Whenever appropriate," the subjects, if waiver is granted, are to be provided with "additional pertinent information after participation." *Id.*

133. 45 C.F.R. § 46.104(d)(4)(iii) (2018). SACHRP has advised that the drafters of the 2018 Regulations included "health care operations and "public health activities and purposes" as bases for the exemption because under the Privacy Rule a project is considered research only if the primary purpose of the activity is the creation of generalized knowledge—as opposed to the Common Rule, where a purpose but not the primary purpose must be the creation of generalized knowledge. Given that health care operations and public health activities could include as a secondary purpose the creation of generalized knowledge, it was appropriate to include those purposes in the exemption. *See Attachment B – Recommendations on the Interpretation and Application of §_.104(d)(4) the "HIPAA Exemption*," U.S. DEP'T OF HEALTH & HUMAN SERVS., https://bit.ly/2VFTBL6 (last updated on Dec. 15, 2017).

134. Jennifer Kulynych & Henry T. Greely, *Clinical genomics, big data, and electronic medical records: reconciling patient rights with research when privacy and science collide*, 4 J. L. & BIOSCIENCES 94, 111 (2017).

periods of time during which the information might be stored and used.[135] In regard to the last requirement, the periods of time may be indefinite,[136] but disclosure that information will be stored and used for an indefinite period is not likely to be viewed favorably by the person being asked to consent. If the point of broad consent is to facilitate ongoing secondary research with biomedical data, requiring these elements will either lead to informed consent disclosures that are inaccurate or are so vague as to be useless. In neither case is the patient's best interest served. A second, more critical impediment is that when broad consent is refused by a patient, the information clearly cannot be used—but in addition an IRB cannot waive consent for the storage, maintenance, or secondary research use of the information.[137] Setting up a system to flag the records of patients who have declined to give broad consent and to prevent their identifiable information from being used through a waiver will require significant deployment of IT resources. In addition, SACHRP has cautioned institutions against de-identifying data in response to a patient's refusal to provide broad consent, noting that while such an action technically may not be a regulatory violation, it would "offend accepted ethical precepts of human subjects research."[138] The likely result is that few institutions will make use of broad consent.[139]

As an aside, it is worth noting that both under HIPAA and the Common Rule, assumptions that data claimed to be de-identified truly are de-identified may be suspect in an era where so much information about people is readily discoverable. Data scientists have long urged that a dichotomous view of data as either identifiable or not is inaccurate and unhelpful.[140] More accurate is the view that "[i]dentifiability exists on a continuum, and the range of deidentification techniques, such as pseudonymization, linking, anonymization, and single and double coding, illustrates the fundamental problem with the bimodal approach."[141] Notorious examples abound of people being identified by linking

135. 45 C.F.R. § 46.116(d).

136. *Id*.

137. *Id.* § 46.116(f).

138. *Attachment C – Updated FAQs on Informed Consent for Use of Biospecimens and Data*, U.S. DEP'T OF HEALTH & HUMAN SERVS., http://bit.ly/2ZhOJz9 (last updated Apr. 13, 2018).

139. *See, e.g.*, *Revised Common Rule*, U.N.C. RESEARCH (Dec. 6, 2016), https://research.unc.edu/human-research-ethics/common-rule/ ("*Due to these implications, the requirement for Institutional tracking, and lack of available guidance from OHRP currently, UNC will not implement the use of "Broad Consent" at this time.*").

140. Mark. A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, AM. J. BIOETHICS., Sept. 1, 2010, at 3, 3.

141. *Id.* (citing Bartha M. Knoppers, *Biobanking: International Norms*, 33 J.L. MED. & ETHICS 7 (2005); ROBERT F. WEIR ET AL., THE STORED TISSUE ISSUE: BIOMEDICAL RESEARCH, ETHICS, AND LAW IN THE ERA OF GENOMIC MEDICINE (2004)).

deidentified data to publicly available data. One involved then-Massachusetts Governor William Weld, who assured the public that a state agency has protected patient privacy when it deleted some identifiers—but not ZIP codes, birth dates, and sex—before making available for researchers every Massachusetts state employee's hospital records.[142] Eighty-seven percent of the American population can be identified through use of ZIP code, date of birth, and sex.[143] Weld found his own medical records identified by Latanya Sweeney, then a graduate student and now a Professor of Government and Technology at Harvard University. Sweeney knew that Weld lived in Cambridge, MA and combined the purportedly deidentified records with the voter rolls of the city of Cambridge, a database that included the name, address ZIP code, birth date, and sex of every voter in the city. [144] Only six people in Cambridge shared Weld's birth date; three of those people were men, and only one—Weld—lived in his ZIP code. "In a theatrical flourish, Dr. Sweeney sent the governor's health records (including diagnoses and prescriptions) to his office."[145] Sweeney went on to conduct several influential studies involving reidentification, in one of which she demonstrated that newspaper stories about hospital visits in the State of Washington, where anonymized health records are made available, led to identifying matching health records 43% of the time.[146]

The unsettling findings from these and similar studies do not mean that deidentification should be abandoned as a goal or promise, and it would be unthinkable for regulators to take the position that sensitive personal data could not be accumulated at all because of the risks of identification, even where data were deidentified. What is indicated is honesty about deidentification in the information age. Otherwise HIPAA authorization and informed consent under the Common Rule become shams.

142. Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1719 (2010) (citing Henry T. Greely, *The Uneasy Ethical and Legal Underpinnings of Large-Scale Genomic Biobanks*, 8 ANN. REV. GENOMICS & HUM. GENETICS 343, 352 (2007)).

143. Latanya Sweeney, *Simple Demographics Often Identify People Uniquely* 2 (Carnegie Mellon Univ., Working Paper No. 3, 2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf.

144. Ohm, *supra* note 142, at 1719–20.

145. *Id.* at 1720.

146. Latanya Sweeney, *Only You, Your Doctor, and Many Others May Know*, TECH. SCI. (Sept. 29, 2015), https://techscience.org/a/2015092903/.

IV. TOWARDS A MIDDLE GROUND

A. Societal Advantage vs. Individual Autonomy

While everyone agrees that privacy is important and most people say that their privacy rights are critical, how people actually act may tell a different tale. The Pew Research Center has reported that the percentage of Americans using social media platforms grew from 5% to 69% between 2005 and 2018, with growth across all age groups.[147] During the same period of growth in social media usage, a 2014 Pew survey found 91% of Americans

> "agree" or "strongly agree" that people have lost control over how personal information is collected and used by all kinds of entities. Some 80% of social media users said they were concerned about advertisers and businesses accessing data they share on social media platforms, and 64% said the government should do more to regulate advertisers.[148]

However, the study also found that people overwhelmingly continue to use social media platforms despite having concerns about the implications for their privacy.[149]

Some commentators argue that data privacy is based on data responsibility, and that the United States should understand protecting data privacy to be a responsibility of individuals as well as governments and businesses.[150] While framing this responsibility as a moral duty, as one scholar does,[151] may be a stretch for many, it does seem reasonable that individuals should take ownership over how their data is disclosed and used. The problem is that exercising such ownership becomes impracticable if doing so means that one is unable to participate in society and its benefits (such as health care) and pleasures—(such as social interactions via Facebook, Snapchat, etc.) Further, protecting one's own privacy can be an impossibility in the age of Big Data, where methods of collection and analysis can overcome even the most informed consumer's capacity to shield confidential information.[152] One can quit Facebook, but not an electronic medical record.

Security breaches involving medical records will continue to occur for a variety of reasons, such as poor oversight, inadequate security

147. Lee Rainie, *Americans' Complicated Feelings about Social Media in an Era of Privacy Concerns*, PEW RESEARCH CTR. (Mar. 27, 2018), https://pewrsr.ch/2UlroJ9.
148. *Id.*
149. *Id.*
150. Allen, *supra* note 4, at 72.
151. *Id.* at 72–73.
152. *Id.* at 73.

training, and susceptibility to phishing.[153] Happily, however, most people favor letting researchers use their personal data.[154] Considered together with the concurrent concerns for autonomy regarding personal data, that tolerance for research use of data may suggest possibilities for a middle ground approach to broad consent that is truly informative and does not over-promise individual autonomy.

B. Toward a "True" Broad Consent

Current regulatory schemes generally promise research subjects more autonomy than they actually provide. Of the existing mechanisms, the IRB waiver is probably the simplest to administer and, assuming IRBs have the best interests of research subjects at heart, may well provide enough protection for the subjects without impeding Big Data research. Some scholars argue for blanket consent, where patients consent to any future research on their data or biospecimens, without having been provided information on future uses and with no further oversight.[155] Others advocate for open consent, such as has been used in the Personal Genome Project, where donors are asked to consent to their data being included in an open-access database with no privacy guarantees. The latter approach seems workable only in a society where prohibitions against discrimination on the basis of genetics and medical conditions are available and rigorously enforced.

Broad consent is a salutary idea but, as demonstrated, is not practicable in its current form. It is simply too difficult to predict the types of research that may be conducted with the patient's identifiable information, the identifiable information that may be used, whether sharing of the information might occur, what types of institutions or researchers might be allowed to use the information, and the periods of time that the information might be stored and used. Any statement of these conditions will either limit the usefulness of a Big Dataset or be so vague as to utterly fail to ensure that the consent is truly informed consent. Moreover, the current requirement that if broad consent is refused an IRB is perpetually barred from waiving consent as to that patient creates too great a logistical and compliance challenge to make broad consent worth

153. *See, e.g.*, Baker Hostetler & Eric Packel, *What Can We Learn from the Healthcare Data Breach "Wall of Shame"?*, JD SUPRA (Feb. 4, 2019), http://bit.ly/2IxOutw (citing John Jiang & Ge Bai, *Evaluation of Causes of Protected Health Information Breaches*, 179 JAMA INTERNAL MED. 265, 266 (2019) (finding that more than half of breaches listed on the HHS OCR website (the "Wall of Shame") between 2009 and 2017 could be attributed to internal mistakes or neglect, as opposed to outside causes).

154. Evans, *Power to the People*, *supra* note 3, at 247.

155. *See* Effy Vayena & Alessandro Blasimme, *Health Research with Big Data: Time for Systemic Oversight*, 46 J.L. MED. & ETHICS 119, 122 (2018).

pursuing when other options, though less transparent to the research subject, are readily at hand.

What's needed is a form of consent that provides patients with a reasonable amount of information about how their data will be used and honors the altruistic motivation to contribute, through facilitating research, to the greater good. Though broad consent may not be feasible or desirable for all types of biomedical Big Databases, one can envision a form of broad consent that honored patient autonomy in reasonable ways without creating unworkable barriers to the conduct of secondary research with medical Big Data. To achieve that goal, providers would need to be honest and straightforward about risks and benefits, and the limitations of data security measures, and the condition—which we advocate—that refusal to grant broad consent would not prevent one's data being used in a study through waiver.

Assuming that the type of database was appropriate for broad consent—and some data, such as substance abuse treatment records, might be deemed too sensitive—a key component of broad consent should be patient education. People being asked to provide a broad consent should be told, at a minimum and in addition to the general requirements for informed consent:

(1) That their medical data will be aggregated into a large database for reseasrch use in one or more studies.
(2) That research may be conducted by employees of the primary custodian or by researchers from other entities.
(3) That no researchers will be allowed access to the data without agreeing to terms of use, including strict observation of confidentiality provisions.
(4) That it cannot be guaranteed with 100% certainty that no data breach will occur, though every effort will be made to prevent such an occurrence, and that they will be notified in the event of a breach.
(5) That the nature of the research conducted using their data, while it will meet ethical standards and will be for improvement of health and well-being, may vary widely.
(6) That broad consent can be revoked at any time. That if they decline to give broad consent they might be contacted to see if they will consent to their data being used for a specific research study.
(7) That it is always the case that in certain circumstances reviewing research ethics boards might conclude that it is permissible to conduct a study using their medical data without their consent, but only if the study is of no more than minimal risk to the

> subjects, could not be practicably carried out without the waiver, and the waiver will not adversely affect the rights and welfare of the research subjects.

Patients armed with the foregoing details, given in plain and unadorned terms, would then be asked to answer one question: "Do you agree at the present time to have your medical data made available for one or more research studies under the terms just described to you?" If the answer were yes, and appropriate signatures followed, broad consent would be put into place and remain in effect until revoked.

One can reasonably question why broad consent should be pursued at all if refusal can be circumvented by waiver. The simple answer is that more information in such matters is generally better. Removing the heavy technological burden that accompanies refusals of broad consent under the current Common Rule, along with other streamlining measures, would encourage rather than discourage its use. Balancing a refusal with waivers in appropriate circumstances does not put the patient at risk for unwarranted harm,[156] and, again, the patient should be told that a waiver might be granted.

V. CONCLUSION

The growth of secondary research projects driven by big data carries significant benefits to society, but also strains research institutions' resources. An appropriate compliance function includes verifying that research will occur under a permissible consent 'pathway;' principally: obtaining informed consent, de-identifying patient data, or obtaining an appropriate IRB or Privacy Board waiver or modification of consent requirements. Increased research activities strains each of these three mechanisms—more consents are needed, or more data must be de-identified, or a greater number of IRB/Privacy Board evaluations and (if appropriate) waivers will be required.

As the pace of big data research increases, research oversight mechanisms must react in parallel. Regulators will continue to struggle to balance respect for individual autonomy with society's needs for research to advance through the use of exciting new resources such as Big Data collections. The recent Common Rule changes reflect an incomplete solution to these challenges. In particular, the Broad Consent mechanism does not go far enough to bring meaningful changes to the informed consent process for big data research projects and is a procedure that fails

156. *See* 45 C.F.R. § 46.116(f)(3) (2018) (providing that an IRB may waive informed consent only if, *inter alia*, the research involves no more than minimal risk to the subjects and the waiver will not adversely affect the subjects' rights and welfare).

to align with its underlying goals. Rather than punishing institutions that follow this path by eliminating the waiver mechanism option if broad consent is pursued, true broad consent, we think, would strike an improved balance, allowing greater individual involvement in how data is used and lessening reliance on an IRB/Privacy Board waiver process, while still preserving waiver as an option when appropriate.