INFORMED CONSENT AND PRIVACY OF NON-IDENTIFIED BIO-SPECIMENS AND ESTIMATED DATA: LESSONS FROM ICELAND AND THE UNITED STATES IN AN ERA OF COMPUTATIONAL GENOMICS

Donna M. Gitter†

This Article analyzes issues of informed consent and patient autonomy raised by advances in bioinformatics and computational genomics. The Article describes the increasing use in biomedical research of estimated data. Researchers are able to use genetic and genealogical data from research subjects, who did agree to participate in genetic testing, in order to make educated guesses about the genetic profiles of their relatives who did not volunteer to participate. This estimated data can then be combined with health records of the non-volunteers in order to conduct computational genetic research, often termed "in silico" biology, without their informed consent. The Article concludes that contributors of estimated data deserve the protection of both the law of informed consent and the right not to know their genetic risk factors. Neither research nor its results ought to be foisted upon any individual, least of all those who unwittingly participate.

"[T]he view we have today of genomes is like a world map, but Google Street View is coming very soon."

—Rebecca Skloot, Author of *The Immortal Life of Henrietta Lacks*[1]

† Professor of Law, Baruch College, Zicklin School of Business, City University of New York. B.A., Cornell University; J.D., University of Pennsylvania Law School. E-mail: Donna.Gitter@baruch.cuny.edu. The Author thanks attendees at the following conferences for their helpful comments on this work: the 2016 Annual Conference on Big Data, Health Law, and Bioethics at the Petrie-Flom Center for Health Law Policy, Biotechnology, and Bioethics at Harvard Law School on May 6, 2016; ALLDATA 2016—the Second International Conference on Big Data, Small Data, Linked Data and Open Data in Lisbon, Portugal from February 21–25, 2016; The Biology of Genomes Conference held at Cold Spring Harbor Laboratory in Cold Spring Harbor, New York from May 5–9, 2015; and the Law and Ethics of Big Data Research Colloquium held at Indiana University Bloomington from April 17–18, 2015. The Author also expresses appreciation to the Zicklin School of Business, Baruch College, City University of New York for research support for this work.

TABLE OF CONTENTS

INTRODUCTION

In the current age of bioinformatics[2] and computational genomics,[3] researchers are able to use genetic and genealogical data from research subjects, who did agree to participate in genetic testing, in order to make educated guesses about the genetic profiles of their

1 Rebecca Skloot, Opinion, *The Immortal Life of Henrietta Lacks, the Sequel*, N.Y. TIMES: SUNDAY REV. (Mar. 23, 2013), http://www.nytimes.com/2013/03/24/opinion/sunday/the-immortal-life-of-henrietta-lacks-the-sequel.html.

2 Bioinformatics is the application of computer technology to the management of biological information. Computers are used to gather, store, analyze, and integrate biological and genetic information, which can then be applied to gene-based drug discovery and development. *See* N. M. Luscombe et al., *What Is Bioinformatics? A Proposed Definition and Overview of the Field*, 40 METHODS INFO. MED. 346, 346 (2001) (defining bioinformatics as "conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying 'informatics' techniques (derived from disciplines such as applied math, computer science, and statistics) to understand and organize the information associated with these molecules, on a large-scale" [sic]).

3 Computational genomics is the practice of deciphering biology from genome sequences using computational analysis. *See* Sophia Tsoka & Christos A. Ouzounis, *Recent Developments and Future Directions in Computational Genomics*, 480 FEBS LETTERS 42, 42 (2000) (defining computational genomics as "a subfield of computational biology that deals with the analysis of entire genome sequences").

relatives who did not volunteer to participate. This estimated data[4] can then be combined with health records of the non-volunteers in order to conduct computational genetic research, often termed "in silico" biology,[5] without their informed consent. Researchers use these technologies to calculate the probability that an individual carries a particular genetic variant, without sequencing that person's DNA, thereby developing estimated data for inclusion in research databases. The increasing use of these technologies, coupled with the proliferation of genetic and medical databases, raises questions regarding informed consent and privacy with respect to individuals' health data, as well as the right not to know one's genetic risk factors. Some researchers, such as the founder of the company deCODE Genetics, Inc., consider the use of estimated data a legitimate business model with valuable applications to biomedical research. Others consider it a breach of genetic privacy.

The law of informed consent does not address the use of estimated data, given that it was not possible before the advent of computational genomics to conduct "in silico" research. This Article contends that the law of informed consent ought to be construed to apply to estimated data, in keeping with traditional norms of biomedical ethics. U.S. federal law already provides that, when informed consent is required, research participants are entitled to a description of any reasonably foreseeable risks. One such risk is the possibility that information about participants might extend to relatives or identifiable populations or groups, contributing to potential discrimination or stigmatization. Since research participants already have a right to informed consent relating to the potential revelation of their relatives' personal health information and data, then those relatives, who did not knowingly accede to participate in such research, ought to enjoy that same right of informed consent rather than being conscripted into research.

Recently proposed changes to the U.S. law of informed consent by the Department of Health and Human Services also support the notion that individuals have the right to be asked for informed consent to research with their estimated data, although these particular changes were not ultimately enacted.[6] Given that biospecimens, medical information, and the data derived from them are increasingly considered intrinsically identifiable, these proposed U.S. rule changes provided that individuals ought to be asked for their informed consent

---

4 In this Article, the term "estimated data" is used to refer to inferred genetic data for people who never participated in a genetic study, but whose relatives did participate.

5 Bernhard Palsson, *The Challenges of In Silico Biology*, 18 NATURE BIOTECHNOLOGY 1147, 1147 (2000) (describing in silico biology as "the use of computers to perform biological studies").

6 *See infra* note 88 and accompanying text.

before the use of even non-identified biospecimens or private information. Currently, no informed consent is required if researchers strip the biospecimens and information of identifiers. In fact, one particularly stringent alternative version of the proposed federal rule change went even further to require informed consent not only for use of biospecimens and identifiable private information, but also for genomic or other data, even if it is non-identified. The recently revised Genomic Data Sharing (GDS) Policy of the U.S. National Institutes of Health (NIH) does this, requiring as a condition of funding informed consent not only for use of biospecimens and identifiable private information, but also for genomic or other data, even if it is non-identified. Given increasing recognition of the ease of re-identification and discussion of the need for informed consent for even non-identified biospecimens and private information, it follows logically that those from whom estimated data are gleaned should be asked for their informed consent. Such individuals do not agree to participate in biomedical research, but rather become involved through the use of their relatives' genetic information and medical records, genealogical records, and databases, both public and private. These non-volunteers should not be conscripted into research without their informed consent.

Paradoxically, the trend toward enhanced respect for research subject autonomy and privacy, evidenced by laws and policies proposing or requiring informed consent even for non-identified specimens and private information, has been accompanied by erosion of the individual's "right not to know" his or her genetic risks. While this right has traditionally been a bedrock principle of biomedical ethics, professional associations have considered encouraging researchers to affirmatively search for genetic incidental findings and report those to research participants, absent their informed consent. This emerging view poses especially severe harms in the cases of individuals from whom estimated data has been gathered, raising the specter of individuals who have given consent neither for the use of their information, nor the return of incidental findings to them, having their estimated data used for research and then being contacted with researchers' incidental findings. This Article argues against incursions upon the right not to know, especially for those whose estimated data was inferred.

Part I of this Article reveals the controversies surrounding the use of estimated genomic data, especially as they arise in Iceland, a small island nation that, due to its largely consanguineous population and detailed genealogical records, lends itself particularly well to genetic research. These issues are not limited to Iceland, however, as evidenced by the existence in the United States of the Utah Population Database (UPDB), which is rendered particularly valuable due to its extensive

genealogical records. Part II examines the current federal law of informed consent in the United States. Proposed changes to the U.S. federal law of informed consent are analyzed in Part III, with a particular emphasis on proposed changes that would have required informed consent for non-identified biospecimens and private information. In addition, Part III considers changes in the funding policies of the NIH, which provide even stronger protection for research subjects than does the federal law. Part IV assesses the likelihood of re-identification of an individual from her non-identified specimens, private information, and/or data, given that the risk of re-identification provided the rationale for the proposed changes in the federal law of informed consent as well as the recent change in NIH policy. In light of the considerable risk of re-identification from non-identified private information, Part V analyzes the arguments against and in favor of the proposed rule changes. Part VI examines how enhanced privacy and informed consent protections have been implemented in relation to the UPDB without compromising the effectiveness of that resource. Part VII then considers the erosion of the right not to know one's genetic risk factors, a surprising development given the recognition that the privacy and autonomy interests of research participants weigh heavily in favor of requiring informed consent even for non-identified biospecimens. The Article concludes that contributors of estimated data deserve the protection of both the law of informed consent and the right not to know their genetic risk factors. Neither research nor its results ought to be foisted upon any individual, least of all those who unwittingly participate.

I. CONTROVERSIES SURROUNDING THE USE OF ESTIMATED GENOMIC DATA IN ICELAND AND THE UNITED STATES

Complex methods of computational genomics, particularly the use of estimated genetic data, are particularly effective in Iceland, an island nation with detailed genealogical records and a population of approximately 320,000 citizens who are considered to be genetically homogeneous.[7] The intimacy of this small island nation is made evident

[7] *See* Jocelyn Kaiser, *Pioneering Icelandic Genetics Company Denied Approval for Data-Mining Plan*, SCIENCE (June 20, 2013, 2:00 PM), http://news.sciencemag.org/people-events/2013/06/pioneering-icelandic-genetics-company-denied-approval-data-mining-plan [hereinafter Kaiser, *Genetics Company Denied Approval for Data-Mining Plan*] (describing the decision of the Iceland Data Protection Authority to reject a request from the company deCODE Genetics, Inc. to "allow it to apply computational methods to the country's genealogical records to estimate the genotypes of 280,000 Icelanders who have never agreed to take part in the company's research and link the data to hospital records"); *see also* David E.

by the existence of a smart-phone app in Iceland that permits individuals to determine whether they are related to another person whom they are considering dating.[8]

In light of Iceland's genetic homogeneity and the availability of detailed genealogical information, in 1996 Icelander Dr. Kári Stefánsson founded the company deCODE Genetics in order to use Iceland's population to pioneer genetic population studies.[9] In 1999, the Icelandic government granted deCODE an exclusive twelve-year license to build a Health Sector Database to hold centralized health records of its entire population.[10] The plan incited much controversy due to the presumption that citizens of Iceland would be deemed to consent to participate unless they actively opted out.[11]

In November 2003, the Supreme Court of Iceland disrupted deCODE's plans by ruling in favor of Ragnhildur Guðdmundsdóttir, an eighteen-year-old student, holding that she could prevent the transfer of her deceased father's health records to the database. The court decided that the records in the database might allow her to be identified as an individual at risk of a heritable disease, even though the data would be anonymous and encrypted. The court noted that this risk was heightened by the fact that the Health Sector Database would allow information to be linked with data from other genetic and genealogical databases.[12]

DeCODE then pursued another strategy, using estimated data to create a research database to find genetic sequences linked to diseases.[13] Using DNA and clinical data from more than 120,000 research volunteers, deCODE analyzed their DNA sequences for a selection of

---

Winickoff, *Genome and Nation: Iceland's Health Sector Database and its Legacy*, INNOVATIONS, Spring 2006, at 80, 84–86 (noting that the consanguinity of Iceland's population and its extensive genealogical records have rendered it a fertile place for genomic research). *But see* Vigdis Stefansdottir et al., *Iceland—Genetic Counseling Services*, 22 J. GENETIC COUNSELING 907, 908–09 (2013) (acknowledging that the genetic homogeneity of Icelanders has been debated and citing a study concluding that Icelanders are not more homogeneous than other European populations).

8 *See* Carol Matlack, *In Iceland, an App to Warn if Your Hookup Is a Relative*, BLOOMBERG (Apr. 17, 2013, 8:11 PM), http://www.businessweek.com/articles/2013-04-17/in-iceland-an-app-to-warn-if-your-hookup-is-a-relative.

9 *See* DECODE GENETICS, https://www.decode.com (last visited Mar. 1, 2017).

10 Alison Abbott, *Icelandic Database Shelved as Court Judges Privacy in Peril*, 429 NATURE 118, 118 (2004).

11 Winickoff, *supra* note 7, at 82–83.

12 Abbott, *supra* note 10, at 188. For an English translation of the decision, see *Icelandic Supreme Court: No. 151/2003* Ragnhildur Guðdmundsdóttir v. The State of Iceland, EPIC.ORG, https://epic.org/privacy/genetic/iceland_decision.pdf (last visited Mar. 20, 2017).

13 Jocelyn Kaiser, *Agency Nixes deCODE's New Data-Mining Plan*, 340 SCIENCE 1388, 1388–89 (2013) [hereinafter Kaiser, *Agency Nixes deCODE*].

slight variations called single nucleotide polymorphisms (SNPs),[14] which are the most common genetic variations among individuals and some of which may prove important in the study of human health.[15]

Using a relatively new technique, deCODE geneticists calculate the probability that an individual carries a particular genetic variant without actually sequencing that person's DNA. For example, deCODE was able to use its whole genome sequencing of the DNA of approximately 2500 research participants in order to extrapolate the genomes of many more individuals. When deCODE identified a genetic variant of interest among the 2500 whole genomes, the company used the more limited SNP data that it had amassed from its 120,000 volunteers in order to impute, with ninety-nine percent accuracy, whether any among these 120,000 also carried the mutations.[16] As noted by one source, "if your mother had been in the hospital for a stroke and agreed to participate in a clinical study, while her brother had volunteered his DNA, deCODE would be able to predict *your* likelihood of a genetic disposition for stroke."[17]

While other researchers are using the same technique as deCODE, the company's unique approach is to combine the known and estimated genotypes for its research participants with its genealogical database, thereby permitting deCODE to estimate what it calls the "in silico" genotypes of close relatives of the volunteers whose SNPs were analyzed. This permits deCODE to infer data of about 200,000 living and 80,000 deceased Icelanders, who have not consented to participate in deCODE's studies. Further, it could essentially give the company genotypes for the largely consanguineous population of 320,000 people in its entirety.[18] Researchers can then determine whether a variant in the DNA sequence found by fully sequencing the DNA of a small group likewise appears in a larger population in the same proportion.[19]

DeCODE not only uses these estimated genotypes as controls in its studies, but also correlates them with health records for patients whose DNA has not been sampled, but who have participated in other types of

---

14 *See SNP*, NATURE.COM, http://www.nature.com/scitable/definition/single-nucleotide-polymorphism-snp-295 (last visited Feb. 22, 2017) (defining a SNP as "a variation at a single position in a DNA sequence among individuals" and noting that "[a]lthough a particular SNP may not cause a disorder, some SNPs are associated with certain diseases").

15 Kaiser, *Agency Nixes deCODE*, *supra* note 13, at 1389.

16 *Id.*

17 Rebecca Goldin, *Privacy and Our Genes: Is deCODE's DNA Project 'Big Brother' or the Gateway to a Healthier Future?*, GENETIC LITERACY PROJECT (June 24, 2013), http://www.geneticliteracyproject.org/2013/06/24/privacy-and-our-genes-is-decodes-dna-project-big-brother-or-the-gateway-to-a-healthier-future/#.UpzQLY5n9So.

18 Kaiser, *Agency Nixes deCODE*, *supra* note 13, at 1389.

19 *Id.*

medical studies.[20] Using estimated data, deCODE published six papers between 2011 and 2013 in the prestigious journals *Nature*, *Nature Genetics*, and the *New England Journal of Medicine*, linking specific genetic mutations to risks of diseases.[21] DeCODE's drug discovery efforts were less successful, however, and the company declared bankruptcy in 2009.[22] In December 2012, Amgen purchased the company for $415 million.[23]

In 2012, deCODE planned to use its strategy as part of a new study. Having imputed the genotypes of the close relatives of the volunteers whose SNPs had been fully catalogued, deCODE intended to collaborate with Iceland's National Hospital to link these relatives to certain hospital records for individuals, such as surgery codes and prescriptions.[24] On May 28, 2013, Iceland's Data Protection Agency (DPA) denied this request on the grounds that it would violate the relatives' privacy unless they gave their informed consent. The DPA gave deCODE until November 2013 to demonstrate that it had obtained consent.[25]

DeCODE ultimately found a means of working around the requirement of informed consent, describing their plan in a November 5, 2013 letter to the DPA. DeCODE confirmed that it had deleted all data registers containing imputed genotypes for individuals from whom consent was lacking. However, deCODE also presented the DPA with a proposal, according to which genotype data from research participants (who had consented) would be linked with genealogy data in a way that would generate statistical results as strong as those formerly achieved. According to the Iceland DPA, this would entail that a genetic imputation for those who had not consented would be generated

> in a split . . . second in the processing memory of a computer. However, this imputation would then cease to exist and would never be accessible to anyone in any form. The only accessible data would be the aforementioned statistical results, which would not in any way be traceable to individuals.[26]

---

20 Kaiser, *Genetics Company Denied Approval for Data-Mining Plan*, *supra* note 7; Kate Yandell, *All Icelandic Women with the BRCA2 Gene Can Be Found in the Database*, NEWS OF ICE. (May 13, 2013) (on file with author).

21 Kaiser, *Genetics Company Denied Approval for Data-Mining Plan*, *supra* note 7.

22 Erika Check Hayden, *Icelandic Genomics Firm Goes Bankrupt*, 462 NATURE 401, 401 (2009).

23 Kaiser, *Agency Nixes deCODE*, *supra* note 13, at 1389.

24 *Id.*

25 Yandell, *supra* note 20.

26 E-mail from Thordur Sveinsson, Icelandic Data Prot. Auth., to Donna M. Gitter, Professor of Law, Baruch Coll. (Oct. 20, 2014, 3:51 PM) (on file with author).

The DPA confirmed in a letter dated November 26, 2013, that this proposal did not give rise to objections if "all the aforementioned prerequisites were met."[27]

Most recently, deCODE published a series of papers in the journal *Nature Genetics* in May 2015 that described sequencing the genomes of 2636 Icelanders, the largest collection ever analyzed in a single human population.[28] Using the imputation technique, deCODE contends that it used the full genomes it has for about 10,000 Icelanders and the partial genetic information on 150,000 more to generate a report for genetic disease on every person in Iceland. For example, the firm can identify every person with the well-known BRCA2 mutation, which raises the risk of breast and ovarian cancer, even if the individual herself has not submitted to genetic testing.[29] Currently, this information is withheld from Icelanders, though deCODE's founder Kári Stefánsson feels strongly that "[i]t's a crime not to approach these people."[30]

DeCODE's originally proposed means of using imputed data (without deleting it) incites controversy because many, including the Icelandic DPA, consider it an invasion of patients' privacy without their informed consent. Dr. Stefánsson of deCODE counters that the practice does not violate patient privacy because it is not actually sequencing the citizens' DNA or collecting personal information, but rather forming "conjectures" or "hypotheses" about them.[31] He notes that estimated DNA sequences, unlike directly measured sequences, are not very accurate for individuals, though they are valuable at the group level.[32] Moreover, Stefánsson contends that, until now, both the DPA and Iceland's national bioethics committee have approved the use of estimated genotypes for the two-thirds of Icelanders who have not consented to its research.[33]

---

27 *Id.*

28 Daniel F. Gudbjartsson et al., *Large-Scale Whole-Genome Sequencing of the Icelandic Population*, 47 NATURE GENETICS 435 (2015).

29 Carl Zimmer, *Snapshot of Icelandic DNA Shows New Gene Mutations Tied to Disease*, N.Y. TIMES, Mar. 26, 2015, at A6.

30 *Id.*

31 Kaiser, *Agency Nixes deCODE*, *supra* note 13, at 1389.

32 *Id.* For example, as noted by Craig Venter of the biotechnology firm Celera Inc., which published the complete sequences of his genome in 2007, although his genomic data indicates an increased statistical risk of developing Alzheimer's disease, he was not surprised that his brain scan results were negative for early signs of the disease. "What works statistically for a population with genomics does not work statistically for individuals. Either you have something or you don't. You don't have 30 percent of Alzheimer's." Liza Gross, *The First Individual Genome: One Is the Loneliest Number*, PLOS: BIOLOGUE (Oct. 21, 2013), http://blogs.plos.org/biologue/2013/10/21/the-first-individual-genome-one-is-the-loneliest-number.

33 Kaiser, *Agency Nixes deCODE*, *supra* note 13, at 1389.

Geneticists disagree as to whether deCODE must obtain informed consent. Jón Jóhannes Jónsson, a geneticist with the University of Iceland, observes that deCODE is not truly doing anything new, given that geneticists routinely infer whether relatives who are not part of a particular study carry a genetic mutation.[34] What is different about deCODE's original strategy is that it invokes the DNA sequences of the entire Icelandic population. Jónsson concedes that deCODE's initial plan to use estimated data supplemented by hospital records presents a difficult case.[35] Daniel MacArthur, a geneticist at Massachusetts General Hospital, suggests that although deCODE did not actually violate the privacy of individuals, from an ethics point of view the researchers should at least attempt to obtain informed consent.[36] MacArthur laments, however, that blocking deCODE from using its estimated data presents a "tragedy" not only for the company, but the wider "complex disease genetics community."[37]

On the other hand, Yaniv Erlich and Arvind Narayanan, experts in computational biology and computer information systems, mention deCODE's efforts in an Article describing various "genetic privacy breaching strategies" that have become increasingly common in the last few years as the range of techniques to carry out such privacy breaching "attacks" has expanded.[38] In particular, they term deCODE's method a "completion technique," meaning the use of known DNA data to "enable the prediction of genomic information when there is no access to the DNA of the target."[39] There have been several high-profile breaches of privacy whereby an "attacker" has been able to infer the genomes of relatives of an individual whose genome is known.[40]

Erlich and Narayanan note that deCODE's approach is an advanced version of the completion technique, given that deCODE has access to the genealogical and genetic information of several relatives of the target, and permits genotypes of distant relatives to be inferred. They explain that it is possible to develop an algorithm that finds

---

34 *Id.*

35 *Id.*

36 *Id.*

37 *Id.*

38 *See generally* Yaniv Erlich & Arvind Narayanan, *Routes for Breaching and Protecting Genetic Privacy*, 15 NATURE REVIEWS: GENETICS 409 (2014). Erlich and Narayanan note that although most of these techniques are currently not accessible to the general public, they can be carried out by those trained in the field. *Id.* at 409.

39 *Id.* at 416.

40 *See, e.g.*, Mathias Humbert et al., *Addressing the Concerns of the Lacks Family: Quantification of Kin Genomic Privacy*, 2013 PROC. 2013 ACM SIGSAC CONF. COMPUTER & COMM. SECURITY 1141, 1141–42 (noting that the uploading of the genome of Henrietta Lacks, a famed but unwitting research subject, violated the privacy of her surviving descendants); *see also* Skloot, *supra* note 1.

relatives of a "target" who donated their DNA to the reference panel and who share a "unique genealogical path that includes the target, for example, a pair of half-first cousins when the target is their grandfather," and that "[a] shared DNA segment between the relatives indicates that the target has the same segment."[41] By studying more pairs of relatives that are connected through the target, it is possible to collect more genomic information on the target without any access to her DNA.[42]

DeCODE's use of estimated data, and the associated privacy and informed consent concerns, raises critical issues with respect to other such databases, such as the Utah Population Database (UPDB) at the University of Utah. The UPDB is the only database of its kind in the United States and one of few such resources in the world. What makes the database unique is the extensive set of family genealogies, maintained by the Church of Jesus Christ of Latter-Day Saints, in which family members are linked to demographic and medical information, analogous to the genealogical records in Iceland. Moreover, while not as consanguine as the Icelandic population, due to immigration patterns in the United States, most Utahans are descended from a common set of European ancestors. Researchers have identified the UPDB as one of the world's richest sources of detailed information useful for research on genetics, epidemiology, demography, and public health.[43] As with Iceland, advances in the fields of bioinformatics and computational genomics will permit researchers to develop estimated data about Utah residents who did not agree to contribute DNA to a research study, thereby raising questions relating to the law and ethics of informed consent and privacy. As noted by Myles Axton, Chief Editor of the journal *Nature Genetics*, Iceland's detailed genealogical records explain why the widespread use of estimated data arose first in Iceland, but a large enough U.S. database could also be used to make similar inferences.[44]

---

41 Erlich & Narayanan, *supra* note 38, at 416.

42 *Id.*

43 *See Utah Population Database*, DAILY UTAH CHRON. (Dec. 5, 2001, 12:00 AM), http://www.dailyutahchronicle.com/2001/12/05/utah-population-database; *Utah Population Database*, HUNTSMAN CANCER INST.: U. UTAH HEALTH CARE, https://healthcare.utah.edu/huntsmancancerinstitute/research/shared-resources/center-managed/updb.php (last visited Feb. 23, 2017).

44 Antonio Regalado, *Genome Study Predicts DNA of the Whole of Iceland*, MIT TECH. REV. (Mar. 25, 2015), http://www.technologyreview.com/news/536096/genome-study-predicts-dna-of-the-whole-of-iceland. While the United States lacks a national database similar to Iceland's, private companies such as 23andMe and Ancestry.com have created rough gene maps of several million people, and the NIH plans to spend millions of dollars in the coming years sequencing full genome data on tens of thousands of people. *Id.*

In order to understand fully the implications raised by genetic databases linked to detailed genealogical information, it is crucial to examine the laws of informed consent and privacy in the United States as they relate to the collection of biospecimens, identifiable private information, and genomic data.

II. THE LAW OF INFORMED CONSENT IN THE UNITED STATES

Meaningful informed consent is one of the fundamental principles of ethical research with human participants.[45] It is designed to ensure that research subjects are aware of the risks and potential benefits of their research participation and make a voluntary, informed decision about participating in the research.[46]

Historically, the law of informed consent arose from instances of unethical abuse of unwilling victims of scientific research. *The Nuremberg Code*, the seminal expression of the rights due to individual participants in medical research, was drafted as a set of standards for use during the Nuremberg War Crime Trials in order to judge physicians and scientists who had conducted horrific atrocities, in the name of biomedical experimentation, on concentration camp prisoners.[47] In the United States, *The Belmont Report*, which established the foundations for legal protections of research subjects, emerged in response to, among other abuses, the decades-long government-funded study in which poor African-American men with syphilis were denied effective and available treatment so that researchers could observe the untreated course of the disease.[48]

U.S. federal law providing for protection of human research subjects, now in its fourth decade, was developed mainly in order to prevent physical harm to vulnerable populations, ensuring that no human being would be required to participate in physically risky research against her will. In particular, legislation, titled the "Federal

---

45 *See* 2 NUERNBERG MILITARY TRIBUNALS, TRIALS OF WAR CRIMINALS BEFORE THE NUERNBERG MILITARY TRIBUNALS UNDER CONTROL COUNCIL LAW NO. 10, at 181–82 (1949), https://www.loc.gov/rr/frd/Military_Law/pdf/NT_war-criminals_Vol-II.pdf (expressing the mandate that "[t]he voluntary consent of the human subject is absolutely essential," specifying the limits of the risks that subjects should be asked to assume as part of research, and setting forth the duties that researchers owe to research participants).

46 *See* U.S. DEP'T OF HEALTH & HUMAN SERVS., THE BELMONT REPORT: ETHICAL PRINCIPLES AND GUIDELINES FOR THE PROTECTION OF HUMAN SUBJECTS OF RESEARCH (1979) [hereinafter THE BELMONT REPORT], http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html.

47 *See supra* note 45 and accompanying text.

48 THE BELMONT REPORT, *supra* note 46, at Part B.

Policy for the Protection of Human Subjects," was set forth in the Code of Federal Regulations (CFR) in 1991.[49] Known as the "Common Rule," this legislation also aims to ensure that investigators make full disclosure to participants of any risks prior to formulating an agreement for research participation, a concept termed "informed consent."[50] The central requirements of the Common Rule are that all proposed research involving human subjects must be submitted to an institutional review board (IRB) for scientific and ethical review,[51] and that informed consent of the subject must be obtained or waived.[52]

The language of the Common Rule states that it applies to "all research involving human subjects conducted, supported or otherwise subject to regulation by any federal department or agency which [sic] takes appropriate administrative action to make the policy applicable to such research."[53] Thus, this regulatory scheme applies to research that is conducted or supported by one of eighteen federal agencies that have adopted it, including the U.S. Department of Health and Human Services.[54] While the Common Rule applies to human subjects research only if it is federally funded or conducted, many privately funded

---

49 *See* 45 C.F.R. §§ 46.101–.124 (2016). "The CFR is a set of rules and regulations established by the US government to add regulatory guidance to the congressionally enacted statutes found in the United States Code." Monica J. Allen et al., *Human Tissue Ownership and Use in Research: What Laboratorians and Researchers Should Know*, 56 CLINICAL CHEMISTRY 1675, 1676 (2010).

50 45 C.F.R. § 46.116 (requiring investigators to obtain the "legally effective informed consent of the subject or the subject's legally authorized representative"). For the basic elements of informed consent, see *infra* note 84 and accompanying text.

51 The Common Rule requires that protocols for human subjects research be IRB-approved before the research can begin. The Common Rule does not require that IRBs be accredited, but it does require them to meet certain membership and review procedures. IRBs generally include volunteers who examine proposed and ongoing scientific research to ensure that human subjects are properly protected. Specifically, the Common Rule requires that each IRB have the following: at least five members; members with varying backgrounds to promote complete and adequate review of research activities commonly conducted by the institution; members that are not entirely of one profession; at least one member whose primary concerns are in scientific areas and at least one member whose primary concerns are in nonscientific areas; at least one member who is not affiliated with the institution; a membership diverse in race, gender, and cultural backgrounds, and having sensitivity to community attitudes; and, if an IRB regularly reviews research that involves a vulnerable category of subjects, such as children, prisoners, pregnant women, or handicapped or mentally disabled persons, efforts should be made to include one or more individuals who are knowledgeable about and experienced in working with these subjects. ERIN D. WILLIAMS, CONG. RESEARCH SERV., RL32909, FEDERAL PROTECTION FOR HUMAN RESEARCH SUBJECTS: AN ANALYSIS OF THE COMMON RULE AND ITS INTERACTIONS WITH FDA REGULATIONS AND THE HIPAA PRIVACY RULE, at CRS-2 to -3 (2005) (citing 45 C.F.R. § 46.107), https://www.fas.org/sgp/crs/misc/RL32909.pdf.

52 45 C.F.R. § 46.117.

53 45 C.F.R. § 46.101(a). "Research" is defined as "a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge." *Id.* § 46.102(d).

54 *See* WILLIAMS, *supra* note 51, at CRS-6.

research institutions and private firms have agreed to subject all of their research activities to its requirements.[55] Most universities follow this and apply the Common Rule to all human research, not just federally-funded studies.[56]

"The Common Rule is coordinated, interpreted, and enforced largely by the Office of Human Research Protection, which is a division of the" U.S. Department of Health and Human Services.[57] The Rule sets forth in detail the composition, function, and role of IRBs in protecting human research participants. The Common Rule also provides the requirements for obtaining informed consent from such participants.[58]

As noted by one expert, the law of informed consent is rooted in concerns about preventing physical injury and did not initially contemplate research on tissue specimens, much less the development of genetic databases.[59] Notwithstanding the fact that the federal regulations do not explicitly define research with human tissue specimens as human subjects research, if such research involves "identifiable private information," it is considered encompassed within the definition.[60] Thus, human subjects research, including research using biospecimens, medical and research records, and administrative data that is not otherwise exempt,[61] is governed by the Common Rule. This is confirmed by the document titled *Coded Private Information or Specimens Use in Research, Guidance* issued by the Office for Human Research Protections (OHRP) of the NIH.[62] According to the OHRP

55 *See id.* Privately funded research that does not voluntarily submit to the Common Rule is governed by varying state laws. *See* Gail H. Javitt, *Take Another Little Piece of My Heart: Regulating the Research Use of Human Biospecimens*, 41 J.L. MED. & ETHICS 424, 426 (2013).

56 Allen et al., *supra* note 49, at 1676.

57 *Id.*

58 *Id.*

59 Javitt, *supra* note 55, at 425–26; *see also* Lynn Sessions, *HHS to Update Common Rule for Human Research Subjects*, LEXOLOGY (Sept. 1, 2011), http://www.lexology.com/library/detail.aspx?g=b069dfe3-0a74-4e61-96df-55ec692e56a3 (noting that the U.S. Department of Health and Human Services acknowledges concerns with the federal human research subject protection in light of the increasing use of genetic information, biospecimens, medical and research records, and administrative data).

60 Javitt, *supra* note 55, at 426 ("In short, federally funded research involving identifiable human biological specimens generally is considered human subjects research for the purposes of the Common Rule, while federally funded research involving samples whose identity has been removed or has not been recorded generally is not considered human subject research according to the statutory definition.").

61 Tissue, specimens, or information may be deemed exempt if they are publicly available or if the information associated with them is recorded by the investigator in such a way that subjects cannot be identified either directly or through identifiers linked to subjects. 45 C.F.R. § 46.101(b)(4) (2016); *see infra* notes 72–75 and accompanying text.

62 *Coded Private Information or Specimens Use in Research, Guidance*, U.S. DEP'T HEALTH & HUM. SERVICES (Oct. 16, 2008) (citing 45 C.F.R. § 46.102(f)), https://www.hhs.gov/ohrp/

Guidance, federal regulations provide that, for the purpose of the definition of human subjects research, obtaining private information or identifiable human specimens includes: (l) "[U]sing, studying, or analyzing for research purposes identifiable private information or identifiable specimens that have been provided to investigators from any source; and [(2)] using, studying, or analyzing for research purposes identifiable private information or identifiable specimens that were already in the possession of the investigator."[63]

Determining whether a particular research project qualifies as human subjects research pursuant to the Common Rule requires close analysis. Research is deemed human subjects research subject to federal regulations when an investigator obtains data through interaction or intervention with a living individual, or obtains identifiable private information about a living individual.[64] Obtaining identifiable private information can occur when researchers use identifiable specimens that have been provided to them, or when researchers use identifiable specimens that were already in their possession. "Private information" is defined pursuant to federal regulations as

> information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record).[65]

In order for the private information to constitute research involving human subjects, it must be individually identifiable, meaning, for example, that "the identity of the subject is or may readily be ascertained by the investigator or associated with the information."[66]

Private information or specimens are individually identifiable only when they can be linked to specific individuals either directly or indirectly through coding systems.[67] Thus, the OHRP does not consider research involving *only* coded private information or specimens[68] to be human subjects research if the following conditions are both met: (1)

---

regulations-and-policy/guidance/research-involving-coded-private-information [hereinafter OHRP Guidance].

63 *Id.*

64 45 C.F.R. § 46.102(f).

65 *Id.*

66 *Id.*

67 OHRP Guidance, *supra* note 62.

68 The OHRP defines "coded" to mean that "identifying information (such as name or social security number) that would enable the investigator to readily ascertain the identity of the individual to whom the private information or specimens pertain has been replaced with a number, letter, symbol, or combination thereof" and "a key to decipher the code exists, enabling linkage of the identifying information to the private information or specimens." *Id.*

"the private information or specimens were not collected specifically for the currently proposed research project through an interaction or intervention with living individuals"; and (2) the investigator cannot immediately ascertain the identity of the individuals to whom the coded private information belongs.[69] Typically, this situation arises when a legal agreement or written IRB policy bars the release of the key to the code to investigators. For example, investigators and a holder of the key to the code may enter into an agreement prohibiting the release of the key to the investigators under any circumstances, until the research subjects are deceased.[70] Alternatively, IRB-approved written policies and procedures for a biospecimen repository may prohibit the release of the key to investigators until the research subjects are deceased.[71]

While certain research involving private information or specimens does not even constitute human subjects research, other human subjects research is exempt from the requirements of the Common Rule. According to the OHRP, if the investigators are not obtaining either data through intervention or interaction with living individuals, or identifiable private information, then the research activity does not involve human subjects.[72] If the activity is human subjects research, the investigator must consider whether the activity is exempt under

---

69 *Id.*

70 *Id.*

71 *Id.* The U.S. Department of Health and Human Services (HHS) regulations further provide, however, that in some cases an investigator who obtains coded private information or specimens about a living individual may unexpectedly learn the identity of one or more living individuals involved in the research or come to believe that it is important to identify the individuals. In that case, if the investigator knows, or may be able to readily ascertain, the identity of the individuals to whom the previously obtained private information or specimens pertain, then the research activity would now involve human subjects. IRB review of the research would now be required. Informed consent would also be mandatory, unless the IRB approved waiver of informed consent under HHS regulations. *Id.*

It should be noted that, until 2004, both Europe and the United States considered coded and linked anonymized samples, in which a code links the sample to its donor, as identifiable and therefore requiring participants' consent to future use. However, in 2004, the OHRP expanded the definition of non-identifiable samples to include those that have been coded, so that now only samples and data that are identifiable require informed consent for their use. *See* Bernice S. Elger & Arthur L. Caplan, *Consent and Anonymization in Research Involving Biobanks: Differing Terms and Norms Present Serious Barriers to an International Framework*, 7 EMBO REP. 661, 661 (2006) (noting the lack of international harmonization on this point); *see also* Eugenijus Gefenas et al., *Research on Human Biological Materials: What Consent Is Needed, and When*, *in* BIOBANKS AND TISSUE RESEARCH: THE PUBLIC, THE PATIENT, AND THE REGULATION 95, 96 (Christian Lenk et al. eds., 2011) (stating that "[i]f coded samples are provided to the researcher in a form that does not allow the identification of the donor, they are treated as anonymous samples in the USA, but as identifiable in many of the European jurisdictions," but noting further the lack of uniformity throughout Europe regarding the terms "identifiable" and "anonymous").

72 OHRP Guidance, *supra* note 62.

regulations of the U.S. Department of Health and Human Services.[73] The most frequent exemption involves "the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects."[74] The OHRP states that "[t]his exemption would not apply if the investigators, having obtained identifiable private information or specimens from existing records or specimens, record the data or information in a coded manner, since the code would enable subjects to be identified through identifiers linked to the subjects."[75]

In short, federally-funded research involving identifiable human biological specimens generally is considered human subjects research for the purposes of the Common Rule. In contrast, federally-funded research is not considered human subjects research where the samples have been decoupled from their identifying information or were never associated with identifying information; such samples are referred to as "de-identified" and "non-identified," respectively.[76] The tradeoff inherent in the distinction between identifiable and non-identified data should be emphasized. The removal of all identifying information from samples helps to maintain the confidentiality of research participants, but samples that cannot be identified are less useful in research because it is not possible for the researchers to update the clinical information associated with the sample over time. What is more, clinically useful results cannot be provided back to research participants if their identifiers are removed.[77]

Once it is determined that informed consent documents are required, these documents must apprise potential research subjects about many aspects of their participation, including: the purpose of the research, the procedures they will undergo, potential risks and benefits of their involvement, mechanisms to protect privacy and confidentiality,

---

73 *Id.* (citing 45 C.F.R. § 46.101(b) (2016)).

74 45 C.F.R. § 46.101(b)(4).

75 OHRP Guidance, *supra* note 62.

76 Javitt, *supra* note 55, at 426. It should also be noted that de-identified health records are also outside the definition of protected health information, and therefore exempt from federal privacy protections under HIPAA. Mark A. Rothstein, *Is Deidentification Sufficient to Protect Health Privacy in Research?*, AM. J. BIOETHICS, Sept. 2010, at 3, 4 ("[H]ealth information that does not identify an individual and with respect to which there is no reasonable basis to believe that the information can be used to identify an individual is not individually identifiable health information." (quoting 45 C.F.R. § 164.514(a))).

77 Sara Chandros Hull et al., *Patients' Views on Identifiability of Samples and Informed Consent for Genetic Research*, AM. J. BIOETHICS, Oct. 2008, at 62, 63. See *infra* Part VII regarding the complex issue of providing research results back to study participants.

payment and commercialization, the release of research results to the participant, and withdrawal from the research.[78] However, researchers are not required to inform research participants regarding many issues that might be salient to such participants. For instance, under the federal regulations, researchers are not required to inform participants of all possible uses of their tissue or to disclose whether and to what extent the research may have a commercial application.[79]

While the physical risks involved in genetic and genomic research are typically minimal, the potential harms generally derive from the potential for misuse of information, which could lead to employment or insurance discrimination, stigmatization, psychological harm, and familial discord. As noted previously, because rapid technological advances have made it possible to link biospecimens and information to individuals who have provided their DNA for research, research subjects face unforeseen risks.[80] Thus, it is critical to remember that the term informed consent refers, ideally, not simply to a form, but rather to a complex process of communication.

The NIH National Human Genome Research Institute emphasizes that "[i]nformed consent involves two fundamental components: a dialogue or process, and a form."[81] The process involves interactions between a member of the research team and a potential participant, aimed at helping the potential participant make informed decisions about whether to become or remain involved in the research, and may in some cases be an "ongoing process, rather than a one-time informational session."[82] The second component of the informed consent process—the consent form—is a written summary of the research project, including the study's purpose, research procedures, potential risks and benefits, among other elements. This document also explains the individual's rights as a research participant.[83]

According to the Common Rule, the necessary elements for an informed consent form include, but are not limited to, the following: assurance that participation is voluntary; a statement of the purpose of the research; a description of the procedures, risks, potential benefits,

---

78 *See generally* Amy L. McGuire & Laura M. Beskow, *Informed Consent in Genomics and Genetic Research*, 11 ANN. REV. GENOMICS & HUM. GENETICS 361 (2010).

79 Javitt, *supra* note 55, at 426 (citing 45 C.F.R. § 46.111 (2009), which lists the requirements for informed consent).

80 *See also* McGuire & Beskow, *supra* note 78, at 366 (noting that researchers have proved capable of tracing genetic data, from a pool of hundreds of people, back to individuals who had provided their DNA for research).

81 *Informed Consent for Genomics Research: Overview*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/27026588 (last updated June 14, 2016) [hereinafter *Informed Consent*].

82 *Id.*

83 *Id.*

confidentiality, and identifiability; an explanation of financial reimbursement, costs, and commercialization; information about withdrawal from research; alternatives to participation; and an explanation of resources available in case of injury.[84]

While technological advances make it possible to gather genetic information about individuals who do not themselves agree to participate in research, but whose relatives do, there is no consensus as to whether individuals who are thus rendered de facto research participants through the use of their estimated data are entitled to informed consent. The informed consent regulations themselves clearly suggest that the use of data from research participants themselves in order to gain information about their relatives raises privacy issues. Indeed, the Common Rule provides that research participants are entitled to "[a] description of any reasonably foreseeable risks or discomforts to the subject."[85] These risks include privacy breaches due to possible re-identification or other losses of confidentiality, and the fact that information about participants in some cases might extend to relatives or identifiable populations or groups, contributing to potential discrimination or stigmatization.[86] For this reason, the Common Rule requires "[a] statement describing the extent, if any, to which confidentiality of records identifying the subject will be maintained."[87] One might argue that if potential research participants have a right to informed consent relating to the potential revelation of their relatives' personal health information and data, then those relatives, who were not knowingly part of the study at all, certainly have a right of informed consent.

Recently suggested revisions to the regulations for the protection of human research subjects proposed by the U.S. Department of Health and Human Services, which were ultimately rejected,[88] did not

84 45 C.F.R. § 46.116 (2016).

85 *Id.* § 46.116(a)(2).

86 *Informed Consent*, *supra* note 81.

87 45 C.F.R. § 46.116(a)(5).

88 In January 2017, federal officials released a final version of the revised Common Rule, which enacted some changes in the law of informed consent, but dropped controversial language that would have required written consent for broad future use of de-identified samples. Nevertheless, one former NIH official who helped draft the revised rule still believes it "was and is the right and respectful thing to do" and commented, in a personal rather than an official capacity, that the NIH can now collect evidence in order to demonstrate whether is it feasible to collect broad consent for samples. Jocelyn Kaiser, *Update: U.S. Abandons Controversial Consent Proposal on Using Human Research Samples*, SCIENCE (Jan. 18, 2017, 4:15 PM), http://www.sciencemag.org/news/2017/01/update-us-abandons-controversial-consent-proposal-using-human-research-samples.

specifically address the need for informed consent with respect to the use of estimated data, but the proposed rules did address an analogous issue: the need to obtain informed consent even for non-identified biospecimens and private information. The rationale for such rule changes was that they are needed to protect the privacy of those from whom biospecimens or other information have been taken, due to the ease of re-identifying individuals through access to those materials. Informed consent also protects a potential research subject's autonomy, permitting her to decline research participation she finds objectionable. Those whose estimated data are used should be entitled to privacy and the right to decline or limit their research participation, especially when one considers that they have a reasonable expectation of privacy in some of the tools used to estimate their data, such as hospital and prescription records.

III. PROPOSED AND RECENT CHANGES TO INFORMED CONSENT LAW AND POLICY

A. *Proposed Changes to the Common Rule*

In July 2011, the U.S. Department of Health and Human Services (HHS) Office for Human Research Protection published an advance notice of proposed rulemaking (ANPRM)[89] that envisioned changes to the Common Rule in order to better protect research subjects and render more efficient the process by which research is reviewed for compliance with informed consent regulations.[90] As explained by HHS, before the federal government implements new regulations, it must typically issue a "Notice of Proposed Rulemaking" (NPRM). The dual purpose of an NPRM is (1) "to inform the public of the specifics of the proposed regulations"; and (2) "to provide the public with the opportunity to react to and comment on the proposed regulations," so that the public's views are taken into account in developing and implementing the final regulations.[91] The federal government may issue an Advance Notice of Proposed Rulemaking (ANPRM) before the

89 Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44,512 (July 26, 2011) [hereinafter ANPRM] (to be codified at 45 C.F.R. pts. 46, 160, 164; 21 C.F.R. pts. 50, 56).

90 *ANPRM Frequently Asked Questions (FAQs) July 2011*, U.S. DEP'T HEALTH & HUM. SERVICES, https://www.hhs.gov/ohrp/regulations-and-policy/regulations/anprm-faq/index.html (last updated Mar. 18, 2016) [hereinafter *ANPRM FAQs*].

91 *Id.*

issuance of an NPRM if the government requires input from the public on various issues before proposing a rule.[92]

According to HHS, the purpose of the ANPRM entitled *Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators* was to "present[] and seek[] comment on specific issues regarding possible changes to the Common Rule" so as to "ensure a broad-based input from the multiple groups and organizations with an interest in the ethics and regulation of human subject research."[93] In keeping with its preliminary nature, the ANPRM did not include specific, proposed regulatory text for public comment, but instead set forth a broad array of possible reforms.[94] These reforms aim to update the forms and processes used for informed consent and establish "mandatory data security and information protection standards for all studies involving identifiable or potentially identifiable data."[95]

One significant proposal of the ANPRM concerned the implementation of data security protections. Currently, there are no specific data security protections for IRB-reviewed research. Regulations require IRBs to determine, for each study, "[w]hen appropriate, [that] there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data."[96] However, IRBs were not designed to evaluate risks to privacy and confidentiality, and often have little expertise in these matters. The ANPRM proposed mandatory data security protection and information protection standards that would be calibrated to the level of the identifiability of the information being collected. Setting uniform specific standards facilitates appropriate privacy and confidentiality protections for all subjects, without the administrative burden of needing a specific committee review of each study.[97]

One of the most fundamental changes proposed in the ANPRM was its language requiring informed consent for research using existing biospecimens, even if non-identified,[98] whether clinical or from prior

92 *Id.*

93 *Id.*

94 *See generally* ANPRM, *supra* note 89.

95 *ANPRM FAQs*, *supra* note 90.

96 45 C.F.R. § 46.111(a)(7) (2016).

97 ANPRM, *supra* note 89, at 44,526.

98 It should be noted that the Common Rule has historically used "the terms 'non-identified' or 'non-identifiable' . . . to signify biospecimens or data that have been stripped of identifiers such that an investigator cannot readily ascertain a human subject's identity," whereas the term "de-identified" is used only to refer specifically to the more stringent standard of non-identifiability set forth in the Health Insurance Portability and Accountability Act of

research. Presently, researchers can use biospecimens without consent by stripping them of identifiers.[99] However, in light of advances in genomic technology that have increased the amount and the nature of information about individuals that can be extracted from their DNA,[100] the ANPRM recommended treating biospecimens as "intrinsically identifiable because of the genetic information imbedded in them."[101]

Pursuant to the ANPRM's proposed change, researchers would have been required to obtain written consent for research using all existing biospecimens, whether clinical or from prior research, even those that had been stripped of identifiers. Consent would not need to be study-specific, and could be obtained using a standard short form by which a person could provide open-ended consent for most research uses of a variety of biospecimens (such as all clinical specimens that might be gathered at a certain hospital). This change would have applied only prospectively to biospecimens collected after the effective date of the new rules.[102]

The ANPRM also provided that data originally collected for one research purpose could be de-identified and then put to another use in a different research project.[103] The ANPRM thereby sought to preclude a misleading practice whereby researchers obtain informed consent to collect data for a disclosed research purpose and then, by de-identifying the data, are able to use it in additional undisclosed research projects. The ANPRM explicitly recognized that if the secondary uses are already contemplated when investigators obtain the original consent, it is

---

1996 (HIPAA). Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,942–43 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46).

99 *See supra* note 76 and accompanying text.

100 ANPRM, *supra* note 89, at 44,525. The ANPRM noted that while HIPAA Privacy and Security Rules require safeguards for individually identifiable information, "rapidly evolving advances in technology coupled with the increasing volume of data readily available" means "much of what is currently considered de-identified is also potentially identifiable data." *Id.* at 44,524. What is more, not all researchers are covered by HIPAA, which applies only to certain entities, such as health care providers and health plans, and, to a certain extent, researchers who are business associates of such entities. *Id.*

101 Barbara J. Evans, *Why the Common Rule Is Hard to Amend*, 10 IND. HEALTH L. REV. 365, 378 (2013).

102 ANPRM, *supra* note 89, at 44,515; *Regulatory Changes in ANPRM: Comparison of Existing Rules with Some of the Changes Being Considered*, U.S. DEP'T HEALTH & HUM. SERVICES, http://www.hhs.gov/ohrp/humansubjects/anprmchangetable.html (last updated Mar. 21, 2016).

103 ANPRM, *supra* note 89, at 44,520. It should be noted that the ANPRM proposed no change with respect to data originally collected for *non-research* purposes, meaning that written consent would be required only if the investigator obtains information that identifies the subjects. *Id.* at 44,519. Under the ANPRM, researchers could continue to use non-identified data originally collected for treatment or administrative (for example, insurance) purposes without consent, "just as such data can be used under the present Common Rule." Evans, *supra* note 101, at 379.

fraudulent not to disclose those uses to the research subject. The ANPRM would bar such misdirection by requiring investigators to obtain consent for all research studies that are being contemplated, even those studies intending to use the data only in a non-identified form.[104]

Professor Evans critiques the breadth of this aspect of the ANPRM, noting that "[u]nfortunately, this proposal also would require consent for future uses of de-identified data that were not contemplated at the time the original consent was procured" even though such uses were not fraudulently withheld, but instead were not anticipated at the time such data was collected.[105] Professor Evans notes that "there is a meaningful legal distinction between premeditated and unanticipated secondary uses of research data" and that "[u]nfortunately, the ANPRM did not acknowledge this distinction and simply would ban all unconsented uses of de-identified data, even those involving no fraud."[106] From the perspective of the research participant, however, the objection might not be to the fraud, but to the inability to refuse consent for the ultimate use of the specimens. Certain research may offend a research subject's religious, moral, or ethical perspective. For example, Arizona State University paid a $700,000 settlement to the Havsupai Indian tribe over an alleged breach of informed consent. While the tribe members consented to genetic testing in order to help determine a genetic variant that might be contributing to increasing rates of diabetes in the tribe, researchers conducted a number of follow-on research projects to which the tribe did not consent, such as searching tribe members for genetic variants linked to schizophrenia, and inferring the likely ancestral origins of the tribe's founders.[107] Thus, while the need to obtain additional consent presents significant obstacles in terms of re-contacting research participants, often years later, the original rationale for obtaining informed consent remains valid as future research uses arise.

Many other commentators offered critiques of the ANPRM, as summarized in the September 8, 2015 NPRM published by the HHS in the Federal Register. HHS noted that the NPRM, entitled *Federal Policy for the Protection of Human Subjects*,[108] reflected the public's input from

104 Evans, *supra* note 101, at 379.

105 *Id.*

106 *Id.*

107 *See* Dan Vorhaus, *The Havasupai Indians and the Challenge of Informed Consent for Genomic Research*, GENOMICS L. REP. (Apr. 21, 2010), http://www.genomicslawreport.com/index.php/2010/04/21/the-havasupai-indians-and-the-challenge-of-informed-consent-for-genomic-research.

108 *See* Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,943 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46).

the ANPRM.[109] Citing administrative and ethical grounds, most commentators opposed the ANPRM's suggested provision requiring consent for research use of all biospecimens, regardless of identifiability. Administrative reasons included the significant costs to collect, log, and track consent status of data and biospecimens, and the administrative efforts required to keep track of the consent status. Some who opposed the suggested consent requirements mentioned "increased privacy risks to subjects arising from the need to maintain links between the consent documents and the biospecimens or data" in order to make certain that limitations on the research use of such data were observed.[110] Commentators also pointed out that proponents of the rule change had failed to present evidence of harm caused by research use of non-identified clinical biospecimens without consent, especially when considering the public health benefit of such use. They believed the principle of beneficence should trump the principle of autonomy. Furthermore, some patient advocacy organizations also expressed concerns about the consequences of requiring consent for the use of non-identified biospecimens. However, most of the comments "strongly supported consent requirements for use of their biospecimens, regardless of identifiability, and data."[111]

After issuing the NPRM, HHS held an October 20, 2015 public town hall meeting on its proposed revisions to the regulations for protection of human subjects in research.[112] The public was once again invited to submit comments on or before January 6, 2016.[113] In the summary of its major provisions, the NPRM reiterated that "informed consent would generally be required for secondary research with a biospecimen (for example, part of a blood sample that is left over after being drawn for clinical purposes), even if the investigator is not being given information that would enable him or her to identify" the specimen's donor, specifying that "[s]uch consent would not need to be obtained for each specific research use of the biospecimen, but rather could be obtained using a 'broad' consent form in which a person would give consent to future unspecified research uses."[114] The NPRM

---

109 *See* Press Release, U.S. Dep't of Health and Human Servs., HHS Announces Proposal to Update Rules Governing Research on Study Participants (Sept. 2, 2015) (on file with author).

110 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. at 54,036.

111 *Id.*

112 *See NPRM for Revisions to the Common Rule: HHS Announces Proposal to Improve Rules Protecting Human Research Subjects*, U.S. DEP'T HEALTH & HUM. SERVICES, http://www.hhs.gov/ohrp/humansubjects/regulations/nprmhome.html (last reviewed Jan. 25, 2017).

113 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 73,679 (Nov. 25, 2015) (to be codified at 45 C.F.R. pt. 46).

114 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,936 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46). While the precise template of this "broad consent"

described the advances in technology driving this proposed change, noting that "[n]ew methods, more powerful computers, and easy access to large administrative datasets produced by local, state, and federal governments have meant that some types of data that formerly were treated as non-identified can now be re-identified through combining large amounts of information from multiple sources," including publicly available sources.[115] In light of this change, "the possibility of fully identifying biospecimens and some types of data from which direct identifiers had been stripped or [which] did not originally include direct identifiers has grown, requiring vigilance to ensure that such research be subject to appropriate oversight."[116] "Most importantly," according to the NPRM, "[a] growing body of survey data show that many prospective participants want to be asked for their consent before their biospecimens are used in research."[117] Thus, the NPRM clearly prioritized an individual's right to elect or decline participation in research. This notion aligns with recognition of the right of informed consent for individuals who participate in silico biology through the use of their estimated data.

Under the NPRM, as with the ANPRM, with regard to data originally collected for one *research* purpose, a fresh consent would be required for secondary research use, regardless of whether the investigator obtained identifiers.[118] The NPRM offered a new proposal with slightly less protection to research subjects whose identifiable private information had been or would be acquired for *non-research* purposes. Under the current Common Rule, secondary research studies using identifiable private information that was collected for non-research purposes undergo IRB review and approval, often using an

was not set forth, the NPRM declared that the Secretary of HHS is in the process of drafting a broad consent template which would be released for public comment "at a later date." *Id.* at 53,969. The NPRM contemplated that there would be at least two broad consent templates developed: "One for information and biospecimens originally collected in the research context, and another for information and biospecimens originally collected in the non-research context." *Id.* at 53,974. It was contemplated that such consent would last for ten years. *Id.* at 53,973.

115 *Id.* at 53,938.

116 *Id.*

117 *Id.*

118 *Id.* at 53,963. The NPRM contemplated that this consent can be broad, and need not be study-specific. *See supra* note 114 and accompanying text. Some commenters also favored requiring IRB review and approval for specific studies involving the use of identifiable private information and identifiable biospecimens, rather than permitting the use of a broad consent for future use to satisfy the regulatory requirement for consent. These commenters believed "that IRB review of specific research studies, and the IRB's consideration of whether a study-specific informed consent should be required or whether informed consent could be waived, was more protective of human subjects than the ANPRM recommendation permitting use of a broad consent for future use." *Id.* at 53,965.

expedited review procedure. "If the activity satisfies the relevant criteria, the IRB may waive the requirement for informed consent, which IRBs typically do."[119] Under the approach proposed by the NPRM, prior notice (without a request for consent) would be required to inform individuals that their identifiable private information might be used in research, and that the identifiable private information would be used only for the specific research for which the investigator requested access.[120] Thus, according to the NPRM, "by ensuring that subjects are notified that their information may be used for research, this notice requirement may enhance subject autonomy."[121] The NPRM also proffered an alternative, even more protective proposal, which would have given the individual the right to opt out of any secondary research with their identifiable private information, rather than simply being notified that such research was going to occur.[122]

In requesting comments from the public, the NPRM noted that most commenters did not provide detailed cost estimates to collect, log, and track consent status of data and biospecimens, and the administrative efforts that would be required to keep track of the consent status, nor did they provide "estimates of the type and number of studies that could not be pursued using existing samples and data because of the absence of sufficient consent."[123] The NPRM requested quantitative information of this sort, as well the value to the public and research participants of being asked their permission for research use of their data and biospecimens.[124]

The website that gathered responses from the public to the NPRM reflected the fact that most comments came from scientists, research institutions, bioethicists, and industry groups who opposed the new consent requirements. As noted by Rebecca Skloot, author of the powerful 2010 book about human subjects research titled *The Immortal Life of Henrietta Lacks*, the general public was largely absent from this debate, given that they are less likely to know of the rule change and to wade through the complex legal language in the NPRM.[125]

One special case that the NPRM singled out is treatment of genomic sequencing data. Under the main NPRM proposal, the treatment of de-identified (or non-identified) *data* was no different than

---

119 *Id.* at 53,963.

120 *Id.*

121 *Id.* at 53,964.

122 *Id.*

123 *Id.* at 53,965.

124 *Id.*

125 Rebecca Skloot, Opinion, *Your Cells. Their Research. Your Permission?*, N.Y. TIMES (Dec. 30, 2015), http://www.nytimes.com/2015/12/30/opinion/your-cells-their-research-your-permission.html?_r=0.

under the current Common Rule. As noted by the NIH National Human Genome Research Institute, "[r]esearch with non-identified data does not constitute human subjects research and is not covered under the regulations."[126] This is true regardless of the source (e.g., biospecimen or medical record).[127] Thus, "genomic sequencing data from clinical encounters could be stripped of identifiers and used without consent (if not originally anticipated for research use)" and "secondary users of previously-generated genomic data could conduct research on non-identified data without the data security safeguards" proposed in the NPRM.[128]

The NPRM also proposed an Alternative Proposal B, which would have expanded the definition of human subjects to include research produced using a technology that created information unique to an individual. This Alternative Proposal, which was broader than the main proposal, would have required consent for not only whole genome sequencing, but also genomic sequencing of even a small portion of a person's genome, as well as other technologies that might be developed in the future that similarly generate bio-unique information.[129] In this way, the Alternative Proposal B proposed the same high level of protection of the right of informed consent for research participants as was recently implemented by the newly revised NIH data sharing policy.

B. *Recent Revisions to NIH's Genomic Data Sharing Policy*

The NPRM noted that the NIH have already changed their policies regarding genomic research so as to express the expectation that researchers funded by the NIH obtain the informed consent of study participants for the potential future use of their non-identified data.[130] The purpose of the policy is to encourage researchers to inform study participants that their data will be broadly shared for future research,

126 *The Notice of Proposed Rulemaking (NPRM) for Revisions to the Common Rule: Summary of Proposed Changes Relevant to Genomics Research*, NAT'L HUM. GENOME RES. INST., http://www.genome.gov/27563327#al-3 (last updated Dec. 22, 2015) [hereinafter *NPRM for Revisions to the Common Rule*].

127 *Id.*

128 *Id.*

129 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,945–46 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46). It should be noted the NPRM also included an Alternative Proposal A that was narrower than the main proposal, in that it would have expanded the definition of human subjects research to include "only specifically whole genome sequencing data, or any part of the data generated as a consequence of whole genome sequencing, regardless of the individual identifiability of biospecimens used to generate such data." *Id.* at 53,945.

130 *Id.* at 53,939.

given that the NIH requires data-sharing as a condition of its funding. NIH will expect informed consent not just for genomic data, but also for cell lines or clinical specimens such as tissue samples, even when they are stripped of source-identifying information.[131] The events leading up to this NIH policy change highlight the informed consent and privacy issues that can arise from the use of putatively non-identified data.

In 2007, the NIH launched two initiatives to foster sharing of genomic data while respecting the privacy and autonomy of study participants. The first, the NIH *Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies* (GWAS Policy), encouraged the sharing of data while promoting participant protections through the creation of two alternatives, either unrestricted or controlled access to GWAS data. The second NIH initiative, the database of Genotypes and Phenotypes (dbGaP), created a central repository that stores and distributes GWAS data for use by other researchers. These initiatives aimed to promote public health while honoring the principles of informed consent of research participants and avoidance of privacy risks associated with the sharing of genomic data.[132]

In 2008, researchers demonstrated that it was theoretically possible to identify an individual's genomic data in a pooled sample.[133] This led the NIH to move unrestricted aggregate genomic data sets in dbGaP into controlled access.[134] This policy covers access to sensitive data, such as those linked with medical information.[135]

Then, in 2013, "using only a computer, an Internet connection, and publicly accessible online resources," a team of researchers from the non-profit Whitehead Institute for Biomedical Research was able to identify nearly fifty individuals who had submitted personal genetic material as participants in genomic studies.[136] These researchers discovered that information from the Coriell repository, collected as part of the 1000 Genomes Project, and other publicly available

---

131 Richard Van Noorden, *US Agency Updates Rules on Sharing Genomic Data*, NATURE NEWS (Sept. 1, 2014), http://www.nature.com/news/us-agency-updates-rules-on-sharing-genomic-data-1.15800.

132 Dina N. Paltoo et al., *Data Use Under the NIH GWAS Data Sharing Policy and Future Directions*, 46 NATURE GENETICS 934, 934 (2014).

133 Nils Homer et al., *Resolving Individuals Contributing Trace Amounts of DNA to Highly Complex Mixtures Using High-Density SNP Genotyping Microarrays*, PLOS: GENETICS (Aug. 29, 2008), http:/journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000167.

134 Paltoo et al., *supra* note 132, at 937.

135 Van Noorden, *supra* note 131.

136 Matt Fearer, *Scientists Expose New Vulnerabilities in the Security of Personal Genetic Information*, WHITEHEAD INST. (Jan. 17, 2013), http://wi.mit.edu/news/archive/2013/scientists-expose-new-vulnerabilities-security-personal-genetic-information.

information could be combined to determine the individual identities of some research participants.[137] In response, the NIH requested that the Coriell repository move the relevant information about individuals to controlled access.[138]

In its latest initiative to protect the privacy and confidentiality of research participants, the NIH has stated that, for NIH-funded studies initiated after January 25, 2015, "NIH expects investigators to obtain participants' consent for their genomic and phenotypic data to be used for future research purposes and to be shared broadly" with other researchers, and that the "consent should include an explanation about whether participants' individual-level data will be shared through unrestricted- or controlled-access repositories."[139] Furthermore,

> [f]or studies proposing to use genomic data from cell lines or clinical specimens that were created or collected *after* the effective date of the Policy, NIH expects that informed consent for future research use and broad data sharing will have been obtained even if the cell lines or clinical specimens are de-identified.[140]

The NIH explained that the reason it "expects consent for research for the use of data generated from de-identified clinical specimens and cell lines created after the effective date of the Policy is because the evolution of genomic technology and analytical methods raises the risk of re-identification."[141] In addition, "requiring that consent be obtained is respectful of research participants, and it is increasingly clear that participants expect to be asked for their permission to use and share their de-identified specimens for research."[142]

The critical difference between NIH's informed consent requirement under the Genomic Data Sharing (GDS) Policy on the one hand, and the NPRM's main proposal for changing the Common Rule on the other, is that the former approach is more protective of individuals' genomic data than the latter. The GDS Policy "expects investigators to obtain consent for genomic and phenotypic data to be used in future research and to be shared broadly, irrespective of the

---

137 Paltoo et al., *supra* note 132, at 937; Fearer, *supra* note 136.

138 Paltoo et al., *supra* note 132, at 934.

139 NAT'L INSTS. OF HEALTH, NOTICE NO. NOT-OD-14-124, NIH GENOMIC DATA SHARING POLICY (2014), http://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html. Some commentators have asserted that the policy does not go far enough to protect against the misuse of data, and should use the language "required" rather than "expected" with regards to the responsibilities outlined in the document, as well as set forth penalties for noncompliance. *Id.*

140 *Id.* (emphasis added) (footnote omitted) (defining clinical specimens as "specimens that have been obtained through clinical practice").

141 *Id.*

142 *Id.*

source of the data," which are expected to be non-identified.[143] In contrast, while the proposed revisions of the NPRM would have required consent for all secondary use of biospecimens and identifiable private information for research, the NPRM did not require consent for genomic (or other) *data* obtained from clinical encounters that are subsequently non-identified, nor were there expectations set forth in the NPRM for broad data sharing.[144] Thus, with respect to secondary research with data (as opposed to biospecimens or other identifiable private information), the NPRM did not require researchers to obtain informed consent to use the data once it had been stripped of identifiers.

The need for changes in the law that would require informed consent for non-identified biospecimens, private information, or data depends in large measure on the actual likelihood of re-identification of the source of those specimens, information, or data. Because medical progress depends on such resources, imposing unnecessary informed consent requirements could impede public health.[145] Experts disagree strongly on the degree of re-identification risk. Nonetheless, it is clear that technology increasingly enables the re-identification of biospecimens and some types of data from which direct identifiers have been stripped or which did not originally include direct identifiers, requiring vigilance to ensure that such research be subject to appropriate oversight.

IV. THE RISK OF RE-IDENTIFICATION

Much of the debate regarding the likelihood of re-identification of individuals from partial data has centered on completion attacks through the use of publicly available data. Professor Ohm maintains that the concept of anonymization is a fiction, given the proliferation of easy, cheap, and powerful re-identification technologies. He warns that researchers have "cast[] serious doubt on the power of anonymization, proving its theoretical limits and establishing . . . the easy reidentification result," leading inexorably to the rejection of anonymization as a "privacy-providing panacea."[146]

---

143 *NPRM for Revisions to the Common Rule*, *supra* note 126.

144 *Id.*

145 *See, e.g.*, Jane Yakowitz & Daniel Barth-Jones, *The Illusory Privacy Problem in* Sorrell v. IMS Health, TECH. POL'Y INST. 1, 1 (May 2011), https://techpolicyinstitute.org/wp-content/uploads/2011/05/the-illusory-privacy-problem-i-2007545.pdf (referring to non-identified health data as "the workhorse driving numerous health care systems improvements and medical research activities").

146 Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1716 (2010).

On the other hand, some experts contend that the risks arising from re-identification of data are smaller than commonly believed because of the effort needed to achieve re-identification. Professor Barth-Jones contends that

> [e]ven with considerable computer assistance with the requisite data management, there simply isn't enough human time and effort as would be needed to track, disambiguate and verify the ocean of messy data required to clearly re-identify individuals in large populations—at least when proper de-identification methods have already made the chance of success very small.[147]

Professors Barth-Jones and Yakowitz further contend that "de-anonymization attacks do not scale well because of the challenges of determining the characteristics of the general population."[148] Because "[e]ach attack must be customized to the particular de-identified database and to the population as it existed at the time of the data collection," such an attack "is likely to be feasible only for small populations under unusual conditions."[149]

Professors Barth-Jones and Yakowitz critique many studies that attempt to re-identify the contributors of biomedical data. They note that the majority of such studies use the source data that produced the de-identified data to create auxiliary information that a would-be data intruder could use in a re-identification attack. Thus, "the study authors created a perfectly clean population register which allowed re-identification to be performed error-free."[150] Professors Barth-Jones and Yakowitz commended what they deemed the only recent de-anonymization study conducted under realistic conditions that a real data-intruder would face, and which verified the re-identification. In this study, performed for the HHS Office of the National Coordinator for Health Information Technology (ONC), the team started with a set of approximately 15,000 patient records that had been de-identified in accordance with HIPAA. The ONC sought to match those de-identified records with identifiable records in a commercially available data repository and conducted manual search through external sources such as the InfoUSA database to determine whether any of the records in the identified commercial data would match up with anyone in the de-identified data set. The result was that the team accurately re-identified

147 Daniel Barth-Jones, *Re-Identification Risks and Myths, Superusers and Super Stories (Part II: Superusers and Super Stories)*, CONCURRING OPINIONS BLOG (Sept. 6, 2012), http://concurringopinions.com/archives/2012/09/re-identification-risks-and-myths-superusers-and-super-stories-part-ii-superusers-and-super-stories.html#more-65800.

148 Yakowitz & Barth-Jones, *supra* note 145, at 7.

149 *Id.*

150 *Id.* at 4.

two of the 15,000 individuals, for a match rate of 0.013%.[151] Thus, Barth-Jones and Yakowitz conclude that even after "extraordinary effort," the risk was "very small."[152]

The work of Professors Barth-Jones and Yakowitz seeks to demonstrate that in order to prove that a person within a given biomedical data set is the only person in the larger population who has a set of combined characteristics (known as "quasi-identifiers") that could potentially re-identify that person, a re-identification attempt must be able to create a complete and accurate population register. Without knowing a complete and accurate listing of the entire population, one cannot be certain that a given individual is the only person in the entire population with that set of characteristics.[153] Professor Barth-Jones thus aims to disprove what he terms the "myth of the perfect population register,"[154] noting that creating a complete and accurate population register "is a tremendous challenge for even the U.S. Census Bureau and would typically be far beyond the likely abilities of a hypothetical data intruder."[155] He further notes that because "[d]isclosure risk scientists themselves usually cannot afford to complete this final exhaustive step when making re-identification risk estimates," they "wisely skip this last essential task and instead make easily obtained, but highly conservative, estimates of the true re-identification risks."[156] Barth-Jones cautions that "is an appropriate practice as long as everyone who interprets the results understands that we've left out the hardest part of the equation and chosen to err strongly on the side of caution in order to protect privacy."[157] Professor Barth-Jones argues care in this balancing of interests, urging that de-identification policy must "achieve an ethical equipoise between potential privacy harms and the very real benefits

151 *Id.* at 4–5.

152 *Id.* at 5.

153 *See* Daniel C. Barth-Jones, The 'Re-Identification' of Governor William Weld's Medical Information: A Critical Re-Examination of Health Data Identification Risks and Privacy Protections, Then and Now 5–6 (July 2012) (unpublished manuscript), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2076397 (noting that, when using voter registration data in order to identify an individual via his or her health records, the individual must be listed in both data sets in order to be at risk of re-identification; it is very rare to achieve perfect population registers because, using voter registration rolls as an example, a large proportion of voting age individuals do not register; and, the absence of individuals from the voting registry confounds the ability to re-identify others who share those quasi-identifiers and are indeed registered to vote).

154 *Id.* at 3.

155 *Id.* at 9.

156 *Id.*

157 *Id.* at 9–10.

that result from the advancement of science and healthcare improvements which are accomplished with de-identified data."[158]

Relating this argument to the use of estimated data and the smaller populations studied in such research, Professor Barth-Jones acknowledges that it "seems reasonable to presume that data intruders might be able to create near-perfect population registries for small or isolated populations for limited time periods aided by their personal knowledge of the population within a specific location."[159] He recognizes that his conclusions apply to larger populations, reemphasizing that given the challenges of determining that there are not individuals missing from the population register and that the quasi-identifier information was correct in both the data source and population register, "any realistic assessment of a lone data intruder's ability to accurately create population registries which include time-dynamic quasi-identifiers (such as patient locations) for populations numbering in the tens of thousands should include some healthy skepticism about the purported 're-identifications.'"[160] Nevertheless, it must be remembered that the ability to create an accurate population register increases dramatically with the use in research of small, genetically homogeneous populations associated with detailed medical and genealogical records, such as is being conducted in Iceland and Utah. What is more, given the research and commercial benefits of achieving re-identification, it should be expected that such activities will be pursued by sophisticated entities, not simply "lone wolves."

Moreover, even if de-identification were truly possible, individuals who contribute biospecimens or other information may desire or expect to be consulted as to their disposition. Indeed, research ethics dictate that individuals ought to have the right to decide whether to participate in biomedical research, and to withdraw at any time if they do participate. Thus, it is necessary to consider societal views as to whether research participants ought to be able to object to the use of their biospecimens, even if they are de-identified. This requires weighing the relative importance of public health versus individual autonomy. With respect to estimated data in particular, it would seem that it is not tenable, under a revised Common Law rule, to accord autonomy to individuals from whom an investigator obtains data through interaction or intervention, yet deny the same rights to individuals from whom data was estimated via a process of in silico biology.

---

158 *Id.* at 13; Daniel Barth-Jones, *The Debate over 'Re-Identification' of Health Information: What Do We Risk?*, HEALTH AFF. BLOG (Aug. 10, 2012), http://healthaffairs.org/blog/2012/08/10/the-debate-over-re-identification-of-health-information-what-do-we-risk.

159 Barth-Jones, *supra* note 153, at 10.

160 *Id.*

V. REACTIONS TO THE PROPOSED CHANGES TO THE LAW OF INFORMED CONSENT

As it currently stands, the federal Common Rule provides that research involving non-identified information is not human subjects research.[161] As noted by Professor Rothstein, current legal requirements are "bimodal," meaning that if information is identifiable, then all of the legal protections are applicable, whereas if the information is "not identifiable," then no legal protections exist whatsoever.[162]

Survey research indicates that the public does not recognize the regulatory distinction between identifiable and non-identified samples and information. One 2008 survey included 1193 patients recruited from general medicine, thoracic surgery, and medical oncology clinics at five U.S. academic medical centers from 2002 to 2003.[163] Most respondents stated that it was moderately to very important for them to be informed that research would be performed on their samples: seventy-two percent when the data was anonymous versus eighty-one percent when identifiable.[164] Only twenty-three percent of respondents differentiated between the two scenarios, including seventeen percent who felt it was moderately or very important for them to know about the identifiable scenario and not about the anonymous scenario—which tracks the requirements of the Common Rule—and six percent felt the opposite.[165] Of those who wanted to be informed about either or both scenarios, as many as fifty-seven percent would require their permission to be sought before their samples were used, whereas the other forty-three percent would accept notification only.[166] For anonymous samples, neither is required under the current Common Rule.[167] The authors of this study concluded that "[f]ew patients expressed preferences consistent with the regulatory distinction between non-identifiable and identifiable information," which should "cause policy-makers to question whether [this] distinction is useful in relation to research with previously collected" samples.[168]

There are many reasons that individuals may object to the use of their non-identified data, even if it is estimated data. First, individuals may decline on ethical, religious, or other personal grounds to

161 *See supra* note 76 and accompanying text.
162 Rothstein, *supra* note 76, at 3.
163 Hull et al., *supra* note 77, at 63–64.
164 *Id.* at 66.
165 *Id.* at 65.
166 *Id.* at 66.
167 *Id.*
168 *Id.* at 62.

participate in certain controversial forms of research, such as somatic nuclear cell transfer, stem cell research, and germ-line gene therapy. In addition, individuals may reject research that purports to establish a genetic link between the members of a specific ethnic group and a particular medical problem. As noted in the Human Subjects Research NPRM, "[a] more participatory research model is emerging in social, behavioral, and biomedical research, one in which potential research subjects and communities express their views about the value and acceptability of research studies."[169] Second, research participants may object to commercial exploitation of discoveries developed through the use of their non-identified data. Largely in response to some highly publicized lawsuits in which research participants have sued researchers for revenue earned from using their information and biospecimens, it has become common for researchers to present research participants with informed consent documents that disclaim any economic interest in possible commercial applications flowing from the research. Research using non-identified records is highly problematic in that there is no informed consent and, therefore, no disclaimer.[170] Since past legal cases contesting researchers' commercial interests in biological materials involve the use of actual biospecimens, not non-identified data,[171] it is not clear whether allegations of commercial exploitation would lie where researchers make use without consent of non-identified health records and data, without corresponding biological specimens.[172]

Just as there are many valid arguments in favor of expanding informed consent protections for research participants, and even for

---

169 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,938 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46). According to the NPRM,

> [t]his participatory model has emerged alongside a broader trend in American society, facilitated by the widespread use of social media, in which Americans are increasingly sharing identifiable personal information and expect to be involved in decisions about how to further share [it], including health-related information that they have voluntarily chosen to provide.

Id. Indeed, "over the past half-century, rather than being passive recipients of health advice and treatment, patients have gradually become more active in decisions about their health and health care," which signals a critical "shift from a paternalistic research environment to one where participants are active partners in biomedical and behavioral research." *Id.*

170 Rothstein, *supra* note 76, at 7.

171 *See, e.g.*, Wash. Univ. v. Catalona, 490 F.3d 667 (8th Cir. 2007) (holding that research institution asserted ownership claims over specimens in a way that contradicted the claims of the principal investigator and the informed consent agreement); Greenberg v. Miami Children's Hosp. Research Inst., Inc., 264 F. Supp. 2d 1064 (S.D. Fla. 2003) (finding that researcher failed to disclose an interest in biological materials); Moore v. Regents of the Univ. of Cal., 793 P.2d 479 (Cal. 1990) (holding that treating physician failed to disclose a commercial interest in biological materials).

172 Rothstein, *supra* note 76, at 7.

those from whom estimated data has been gleaned, there are numerous legitimate concerns voiced by the research community in their opposition to the extension of research protections, whether for non-identified or estimated data. First, it is not feasible to contact each individual from whom data has been de-identified or estimated in order to request that person's informed consent. Even if it were possible, it would be very time-consuming and costly. Each individual's contribution to the research is so small, perhaps as to be dispensable, yet would require the full process of informed consent. Most importantly, and flowing from these reasons, the necessity of such informed consent might delay, and perhaps even preclude altogether, the development and introduction of medical advances.[173] Furthermore, it is not only researchers, but also patient advocacy groups, who warn of these dangers. As noted by these critics, in the context of requiring informed consent for the use of non-identified data, requiring such consent "might inappropriately give greater weight to *The Belmont Report*'s principle of autonomy over the principle of justice, because requiring consent could result in lower participation rates in research by minority groups and marginalized members of society," though "most of the comments from individual members of the public strongly supported consent requirements for use of their biospecimens, regardless of identifiability."[174]

Indeed, it can undermine trust in the medical establishment when individuals learn that their data, whether non-identified or estimated, is used without their consent. As noted by Professor Rothstein, in discussing the need for informed consent for non-identified data, individuals are likely to conflate their health care providers and researchers, particularly "when the providers and researchers work for the same institution and patient-based clinical records and specimens are used in the research."[175] He mentions, among the list of possible consequences of denying informed consent with respect to non-identified data: individuals delaying or foregoing treatment, seeking care only at institutions that do not conduct research, refusing to participate in clinical trials, and declining to support public expenditures for health research.[176] Indeed, the Human Subjects Research NPRM stated that "failure to acknowledge and give appropriate weight to this distinct autonomy interest in research using biospecimens could, in the end, diminish public support for such research, and ultimately jeopardize

173 *Id.* at 8.

174 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,943–44 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46).

175 Rothstein, *supra* note 76, at 7 (citations omitted).

176 *Id.* at 7–8.

our ability to be able to conduct the appropriate amount of future research with biospecimens."[177]

It is clear that the trend, as reflected in the proposed Human Subjects Research NPRM and NIH policy, is toward the requirement of informed consent for the use of non-identified data. The question then arises whether there is a meaningful distinction between non-identified data and estimated data in terms of the need for informed consent for the use of such data. It should be noted that neither non-identified nor estimated data requires any direct interaction with the individual about whom data is gathered. Indeed, the Common Rule specifies that human subject research occurs when an investigator obtains data either "through intervention or interaction with the individual," or obtains "identifiable private information."[178] The regulation provides that "[p]rivate information must be individually identifiable (i.e., the identity of the subject is or may readily be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects."[179] It is this condition of individual identifiability that deCODE Genetics seeks to avoid when it declares to the Icelandic Data Protection Authority that the data will be individually identifiable only for a split second and then deleted from the computer memory.[180] This argument fails, however, if data is as easily identifiable as Yaniv and Erlich have described.[181]

The main difference between non-identified data and estimated data is that the latter are not accurate at the individual level, but only at the group level.[182] While this fact may adequately address the privacy issue, it does not resolve the matter of autonomy, meaning a person's ability to decline to participate in research, either totally or as a means of rejecting the specific research proposed. Given the difficulties inherent in the use of estimated data, it is useful to consider two different models that users of the Utah Population Database implemented in order to link that data and make it more useful, while simultaneously privileging the preferences of research participants.

177 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. at 53,942.
178 *See supra* note 64 and accompanying text.
179 45 C.F.R. § 46.102(f) (2016).
180 *See supra* notes 26–27 and accompanying text.
181 *See supra* notes 38–42 and accompanying text.
182 *See supra* note 32 and accompanying text.

VI. DATA LINKAGE USING THE UTAH POPULATION DATABASE (UPDB)

The UPDB has the potential, if not administered properly, to generate informed consent and privacy problems relating to estimated data, given the detailed genealogies upon which the source is based and the ability to link to other databases in order to gain more information about research subjects. The UPDB is one of the world's richest sources of demographic and family history information on more than 7.2 million individuals whose data can be used to support research on genetics, epidemiology, demography, and public health.[183] Created in the mid-1970s from genealogy records, it has greatly expanded over the years to include records from a number of contributors,[184] including the Utah Department of Health, which tracks: births, deaths, marriages/divorces, hospitalizations, and ambulatory surgery; cancer registry data; family history records; driver's licenses; and voter registration records. "The UPDB contains a master subject index that allows for cross linkage with health care administrative records of all patient encounters maintained in electronic data warehouses."[185]

Because of its size and the many sources from which it draws, the UPDB represents most families living in Utah. Researchers have used this resource to discover links between genetic variants and human disease, familial risk associated with heritable diseases, and quantification of other disease risk factors.[186] The data access model is different from the one used by deCODE Genetics.

Access to the UPDB is governed by the Utah Resource for Genetic and Epidemiologic Research (RGE),[187] which contracts with UPDB data contributors, such as hospital systems and the Department of Health, and sets conditions for data use and reviews requests to access UPDB data. UPDB staff can use personal information to match individuals

---

183 *See* Linda S. Edelman et al., *Linking Clinical Research Data to Population Databases*, 62 NURSING RES. 438, 442 (2013).

184 *See* Scott L. DuVall et al., *Evaluation of Record Linkage Between a Large Healthcare Provider and the Utah Population Database*, 19 J. AM. MED. INFORMATICS ASS'N e54, e55 (2012).

185 Edelman et al., *supra* note 183, at 439.

186 DuVall et al., *supra* note 184, at e55.

187 Created by Executive Order of the Governor and functioning since 1982, the Utah Resource for Genetic and Epidemiologic Research (RGE) does not conduct research, but holds, maintains, and improves data used by research projects, which it obtains through contracts with data contributors. Each contract sets forth the conditions for use of the data and requires that the data contributors approve projects that use their data. Jean E. Wylie & Geraldine P. Mineau, *Biomedical Databases: Protecting Privacy and Promoting Research*, 21 TRENDS BIOTECHNOLOGY 113, 113 (2003).

from two or more datasets, and then remove it from the final linked dataset provided to researchers.[188]

Requests for access are reviewed by the RGE Review Committee, which includes representatives of the data contributors and others knowledgeable about the data and the research uses of them. An IRB must also approve the project. Data use is project specific, such that researchers may not use the data for any other project or purpose and, on completion, data must either be destroyed or returned to the RGE. "Project requests for information that identifies individuals must be made specifically and must be justified. To protect the privacy of individuals represented in RGE-held data, research projects wishing to contact individuals . . . through those data for information and/or biospecimens must adhere to a specific protocol" as follows.[189] First, potential subjects are contacted by representatives of the data contributor or its designee about interest in the study. Second, "identifying information is provided to investigators only for individuals who agree to be contacted."[190] Third, "information and/or biospecimens are collected only after consent and only as part of the specific research study."[191] Thus, the "confidentiality of the information about an individual is protected by providing that individual with the opportunity to decline contact with the researcher."[192]

The UPDB model accepts the premise that research by for-profit companies, while controversial, is a typical feature of the U.S. landscape. While for-profit organizations are not granted direct access to RGE-held data, they may partner with a university or non-profit entity. Commercial entities, however, may participate only in research involving no identifying information.[193]

Researchers contend, however, that there is a clear need for retaining identifying information when data from multiple sets is linked. Thus, the UPDB has moved away from linking records to the creation of "person-oriented information." This is longitudinally linked data that, when considered in concert, convey a significant amount of information about an individual.[194]

UPDB data has been used by researchers retrospectively to identify individuals with different cancer phenotypes and genotypes, as well as to link research subjects to relatives with data in the UPDB to study the

188 Edelman et al., *supra* note 183, at 439.
189 Wylie & Mineau, *supra* note 187, at 113.
190 *Id.* at 114.
191 *Id.*
192 *Id.*
193 *Id.*
194 *Id.*

effects of genetic mutations on female fertility. In addition, UPDB data has been linked to data from an autism prevalence study in order to research mortality and causes of death.[195]

A 2013 study published in the journal *Nursing Research* using UPDB data demonstrates the feasibility of linking research-participant data to data from population databases in order to study long-term post-study outcomes. The purpose of the research was to combat the fact that "the cost of following research participants over time is often prohibitive for all but the largest and best funded of studies. Therefore, the majority of clinical research studies are limited in the length of follow-up time, as well as the long-term outcomes measured."[196]

In the *Nursing Research* study, participants were linked from a completed oncology nursing research trial to outcomes data in two state population databases, the Utah Population Database and the Utah Emergency Department Database. The nursing research study, titled *Energy Conservation and Activity Management (ECAM) for Patients with Cancer-Related Fatigue* was used as the archetype. The ECAM trial was a multicenter, multistate study funded by the NIH that "tested a nursing intervention focused on helping patients to conserve energy as a strategy for managing fatigue during cancer treatment."[197] The study took place between September 1999 and October 2001 at the Huntsman Cancer Institute (HCI) in Salt Lake City, Utah and Fox Chase Cancer Center in Philadelphia, Pennsylvania.[198]

For this study, the ECAM participant data were linked to the UPDB, with three UPDB databases providing most of the data for the study. The Utah Cancer Registry contains information from Utah residents diagnosed with cancer since 1966. The vital records death certificates database maintains cause of death information. The hospital discharge database contains population-based information about hospitalizations throughout the state of Utah.[199] Another database used was the Emergency Department Database (EDDB), which is maintained by the Department of Health and contains information on all emergency department visits to Utah hospitals. The data includes visit information such as date, length of stay, discharge status, and treatment outcomes.[200]

The researchers complied with existing ethics and privacy laws. RGE and Utah Department of Health IRB approvals were obtained, and

195 Edelman et al., *supra* note 183, at 442.
196 *Id.*
197 *Id.* at 439.
198 *Id.*
199 *Id.*
200 *Id.*

due to study investigators never having access to identifying information other than what they had access to in the original ECAM study dataset, the University of Utah IRB approved the study as a minimal risk study. Only database administrators had access to personal health information contained within the UPDB and EDDB that was used to identify and link records.[201] Software was used to link UPDB records to all EDDB records for visits after the date of cancer diagnosis. Pairs of UPDB and EDDB records that had a ninety percent or greater probability of being a true match were linked.[202] When the final study dataset, which was a result of the linkage of the ECAM, UPDB, and EDDB databases, was complete, all identifying information was removed before returning the linked data to the study investigators.[203]

The final dataset contained demographic, cancer diagnosis and treatment, and baseline data from the oncology study linked to post-study long-term outcomes from the population databases. Ultimately, 129 of 144 (89.6%) study participants were linked to their individual data in the population databases. Of those, seventy-three percent were linked to hospitalization records, sixty percent to emergency department visit records, and twenty-eight percent were identified as having died.[204] Thus, the study's investigators concluded that their work demonstrated "the feasibility of linking completed oncology research participant data to large population databases to answer questions about long-term outcomes."[205]

The investigators in the *Nursing Research* study identified means of protecting research participants' rights to informed consent and privacy while leveraging the value of the project data. Most importantly, including post-study linkage plans in the initial study design "creates the opportunity to obtain explicit consent from participants to use specific data items for linkage with population databases."[206] For example, researchers should determine at the outset which databases they may want to link study participant data to in the future, and request permission to do so in the study consent form.[207] This study pursued an approach that protected patient autonomy and confidentiality, while employing population databases as a robust and cost-effective source of data on the long-term outcomes of participants in clinical research studies, thereby contributing to medical research.

---

201 *Id.*
202 *Id.* at 440.
203 *Id.* at 439.
204 *Id.* at 438.
205 *Id.* at 442.
206 *Id.* at 443.
207 *Id.*

Another recent study, published in the *Journal of the American Medical Informatics Association* in 2012, demonstrates the importance of creating an index, called a master subject index (MSI), between institutions that link their medical information. This enables each institution to maintain control and confidentiality of its own information.[208] This study linked records from the UPDB and the enterprise data warehouse (EDW) maintained by Intermountain Healthcare, a nonprofit healthcare delivery system.[209] Intermountain is the largest healthcare system in Utah and operates multiple hospitals, outpatient clinics, ambulatory surgery centers, laboratories, and health insurance plans covering Utah and southeastern Idaho.[210]

The Pedigree and Population Resource at the Huntsman Cancer Institute, University of Utah maintains the UPDB and is responsible for linking resources to the UPDB. The University has substantial experience with linking diverse datasets.[211] The UPDB demographic fields used in record linking include: full name (including maiden name); sex; date of birth; multiplicity (to identify twins and other multiple births); death date; social security number; and residence history (street address, city, state, and zip codes). Also available in the UPDB are names, social security numbers, and residential history of parents, siblings, and spouses.[212]

"The family structure that is available in UPDB was used to calculate the depth of the pedigree for each linked EDW record," which do not contain familial information.[213] These relationships are measured in "pedigree quality," meaning an "indication of how useful a record is for genetic and familial analysis."[214] Pedigree quality levels for records that linked to UPDB were assigned in the following manner: "[N]o family relationships[;] parent-child set or siblings with parents who had only name information[;] two-generation family with four or more members[;] multi-generational pedigree with three or more generations. Some pedigrees have as many as 11 generations."[215]

The wide range of information available through the UPDB from its various source records added considerable value to the more than 3.4 million records in the EDW that linked (out of more than five million total EDW records). While the EDW itself has no familial information,

208 DuVall et al., *supra* note 184, at e54.

209 INTERMOUNTAIN HEALTHCARE, https://intermountainhealthcare.org (last visited Mar. 4, 2017).

210 *Id.*; *see also* DuVall et al., *supra* note 184, at e54–55.

211 *Id.* at e55.

212 *Id.*

213 *Id.* at e56.

214 *Id.*

215 *Id.*

more than half of the linked records have multi-generational family information, thereby furnishing the ability to detect and localize genetic traits. "Since 78.7% of all linked records have at least some family information, parent-child and sibling pairs can be analyzed when varying amounts of pedigree information is available."[216]

This study noted that "[o]ur methodology allows the record linking activity to be completed using patient demographic information without exposing any medical information" due to the creation of an MIS rather than a combined database.[217] When research projects request use of the new information, the investigator is required to obtain approval from RGE and the institutional review boards from each institution, and it is only at this point that information from both institutions is accessed and combined. Thus, "[t]he MSI allows each institution to maintain control of their information and protects the confidentiality of the individuals within each institution."[218] This study therefore demonstrates the possibility of database linkage that preserves the privacy of the research participants. Such examples are increasingly important as research becomes more reliant on large datasets, which raise complex issues of privacy and autonomy.

VII. FUTURE ISSUES RAISED BY THE USE OF ESTIMATED DATA: THE RIGHT NOT TO KNOW

As demonstrated by the studies describe above and explained by Professor Evans, twenty-first century research is "inherently collective" in its nature, "[u]nlike the randomized, controlled clinical trials for which the Common Rule was primarily designed," and "sometimes require extremely large, inclusive datasets free of the biases that can creep in if people self-select for research participation."[219] The cost of protecting privacy and autonomy of research subjects may be too high, in that other people may suffer death or disease.[220] Gísli Pálsson, an anthropologist with the University of Iceland, believes that traditional notions of medical ethics are now in direct conflict with biomedicine, and that standards will need to be adjusted in fundamental ways in the future, so as to emphasize public health at the expense of individual privacy rights.[221]

---

216 *Id.* at e55, e57–58.
217 *Id.* at e55.
218 *Id.*
219 Evans, *supra* note 101, at 386.
220 *See id.*
221 Regalado, *supra* note 44.

On the other hand, perhaps society will decide collectively that protecting people's right to informed consent ought to take precedence over the potential for saving human life.[222] It is also true that medical research often does not lead directly and rapidly to a cure, or necessarily lead to a cure at all. This would mean society would have to suborn definite breaches of informed consent in the hope that they would ultimately lead to a public good, absent any assurance or guarantee that they would. By such logic, a proportion of the population (perhaps the sickest among them, those who are already under medical treatment) ought to be included in research, even without their informed consent, in order to serve as human subjects for research designed to protect society at large. It is just such thinking that the Nuremberg Laws were designed to combat.

The use of estimated data portends even more thorny issues in the future, including a person's right not to know his or her genetic risks, particularly when that person has never agreed to participate in biomedical research in the first place. In the context of genetic testing, the "right not to know" (RNTK) refers to the idea "that adults should be permitted to control whether they receive genetic information—particularly information about the risk of future illness."[223] The two principles underlying the right not to know are respect for an individual's decisional autonomy, as well as the principle of "protecting individuals from receiving unwanted and potentially harmful information."[224]

The RNTK faces increased scrutiny in an era of computational genomics. Because deCODE is able to estimate the DNA profile of nearly all Icelanders, it can now identify approximately 2000 people with the BRCA2 mutation, which is associated with greatly increased risk of breast and ovarian cancers. The company has been in negotiations with health authorities regarding whether to alert those individuals. According to Kári Stefánsson, the founder and CEO of deCODE: "We [can] save these people from dying prematurely, but we are not, because we as a society haven't agreed on that . . . . I personally think that not saving people with these mutations is a crime."[225] The Icelandic Ministry of Welfare has formed a special committee to

222 *See* Evans, *supra* note 101, at 386–87.

223 Benjamin E. Berkman & Sara Chandros Hull, *The "Right Not to Know" in the Genomic Era: Time to Break from Tradition?*, AM. J. BIOETHICS, Mar. 2014, at 28, 28–29.

224 *Id.* at 29.

225 Regalado, *supra* note 44. Alerting individuals as to their BRCA2 status, which is related to prostate as well as ovarian cancer, is particularly important given that these cancers are particularly amenable to prophylactic treatments, such as mastectomy. *Id.*

regulate such "incidental" findings and is planning to propose regulations in the future.[226]

The RNTK genetic information about oneself was traditionally a generally accepted principle.[227] As noted by one commentator, however, the "right not to know" has become more controversial in recent years, due to evolving professional practice guidelines.[228] In 2013, the American College of Medical Genetics and Genomics (ACMG) issued a highly controversial recommendation that when a report is issued for clinically indicated genome sequencing, "a minimum list of conditions, genes, and variants should be routinely evaluated and reported to the ordering clinician."[229] The ACMG advised these incidental findings be reported even "without seeking preferences from the patient and family."[230] The ACMG acknowledged that this approach "may be seen to violate existing ethical norms regarding the patient's autonomy and 'right not to know' genetic risk information," but emphasized their view that "clinicians and laboratory personnel have a fiduciary duty to prevent harm by warning patients and their families about certain incidental findings and that this principle supersedes concerns about autonomy."[231] This recommendation was very surprising because it not only imposes on the medical establishment a "duty to hunt," but clearly moves from a regime that respects the RNTK to one that imposes the obligation to learn one's genetic risks.

The ACMG reversed its position the following year, due to criticism from many groups, including a federally appointed bioethics panel. Now, the ACMG recommends that patients having their genome sequenced consult with their doctors to decide whether they want genetic testing for an array of genetic disorders.[232]

---

226 *Id.*

227 *See, e.g.*, Council of Europe, Convention on Human Rights and Biomedicine, art. 10, ¶ 2, Apr. 4, 1997, 36 I.L.M. 817, E.T.S. No. 164 ("Everyone is entitled to know any information collected about his or her health. However, the wishes of individuals not to be so informed shall be observed.").

228 Benjamin E. Berkman, *Should a Patient Have a Right Not to Know Genetic Information About Him or Herself?*, HARV. L. PETRIE-FLOM CTR.: BILL OF HEALTH (Nov. 19, 2015), http://blogs.harvard.edu/billofhealth/2015/11/19/should-a-patient-have-a-right-not-to-know-genetic-information-about-him-or-herself; *see also* Effy Vayena & John Tasioulas, *Genetic Incidental Findings: Autonomy Regained?*, 15 GENETICS MED. 868 (2013).

229 Robert C. Green et al., *ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing*, 15 GENETICS MED. 565, 573 (2013).

230 *Id.* The ACMG chose conditions for which diagnoses could be confirmed; which preventive measures and/or treatments were available; and individuals with pathogenic mutations might be asymptomatic for long periods of time. *Id.* at 567.

231 *Id.* at 568.

232 *See* Rina Shaikh-Lesko, *The Right to Not Know*, SCIENTIST (Apr. 2, 2014), http://www.the-scientist.com/?articles.view/articleNo/39614/title/The-Right-to-Not-Know. It should be noted that one critique of the right not to know is that it disproportionately affects the

Despite the ACMG's reversal, its original viewpoint seems to be shared by many IRB members and staff, according to the first extensive national survey of IRB professionals, published in 2015.[233] An overwhelming majority of respondents, ninety-six percent, endorsed the right of research participants not to know their genetic incidental findings. However, when asked about a case where a specific patient has chosen not to receive clinically beneficial incidental findings, only thirty-five percent indicated that the individual's RNTK should absolutely be respected and twenty-eight percent responded that they would "probably" honor the request not to know.[234] The percentage of respondents who do not support the RNTK increased from two percent at baseline to twenty-six percent when presented with the specific case, and the percentage of people who are unsure likewise increased from one percent to eleven percent.[235] As noted by one of the principal investigators, "[t]hese data demonstrate that support for a strong RNTK is soft; while autonomy and the RNTK may seem sacrosanct in isolation, forcing people to confront the tradeoffs inherent in real world cases changes many minds."[236]

The proposed NPRM addressed issues relating to the return of research findings to those being studied. In the case of research involving the use of biospecimens or identifiable private information that have been stored or maintained for secondary research use via "broad consent," the NPRM would have provided that broad consent would not suffice in cases where the investigator anticipated that individual research results would be returned to the subjects. In such instances, IRB review and approval would be required for a plan to return those research results to subjects. The NPRM explained that when a series of genetic analyses were performed, investigators would learn information that was not necessarily related to their studies, but that would be significant to subjects in terms of health care decisions;

disenfranchised, by permitting doctors to skip important testing for those who may need it but do not advocate for themselves. *See id.* One way of dealing with this issue might be better informed consent policies, which offer and explain testing to individuals who may need it, as opposed to mandating such testing and its return to individuals who may prefer not to undergo it.

233 Catherine Gliwa et al., *Institutional Review Board Perspectives on Obligations to Disclose Genetic Incidental Findings to Research Participants*, 18 GENETICS MED. 705, 705 (2015).

234 Benjamin E. Berkman, *The Right Not to Know*, HEALTH CARE BLOG (Dec. 15, 2015), http://thehealthcareblog.com/blog/2015/12/15/the-right-not-to-know (discussing his published research).

235 *Id.*

236 *Id.*

one example would be a woman learning of a genetic mutation significantly increasing her risk of breast or ovarian cancer.[237]

Pursuant to the proposed NPRM, a researcher could have alternatively stated in its consent documents that researchers would not provide individual results to research participants. This could have a negative impact on research participation, however, as individuals would be less inclined to consent to research when investigators were not making any commitment to return to them important information that unexpectedly arises. As a result, some investigators would have been inclined to include in in their research protocols provisions for returning individual results to subjects, thereby requiring IRB review. The NPRM noted that "many IRBs do not have any particular unique expertise in making these determinations about returning results," which "could lead to inappropriate variability in disclosure from study to study, and would seem to be in conflict with the ethical goal of justice."[238]

The NPRM mentioned that one option that has been considered would be the creation of "a federal panel of experts to make determinations about which unexpected findings should be disclosed to human subjects in research, and what information should be given to subjects about themselves," rather than full IRB review of these protocols.[239] If this option were implemented, no informed consent would be required even if the researchers proposed to return results to subjects, as long as disclosures were made in conformance with the rules announced by the federal panel. On the other hand, it is not certain that such a panel's guidance would be superior to that of IRBs.[240]

The application of the RNTK becomes even more complex with the estimated data gleaned from individuals who are not research participants. The potential limitation of this right raises the specter of individuals who have given consent neither for the use of their information, nor the return of incidental findings to them, having their estimated data used for research and then being contacted with researchers' incidental findings. This paternalistic approach conflicts deeply with longstanding norms of biomedical ethics.

---

237 Federal Policy for the Protection of Human Subjects, 80 Fed. Reg. 53,933, 53,967 (Sept. 8, 2015) (to be codified at 45 C.F.R. pt. 46).

238 *Id.*

239 *Id.*

240 *Id.*

CONCLUSION

As stated in proposed rule changes to the U.S. federal law of informed consent, biospecimens are increasingly viewed as intrinsically identifiable. What is more, armed with bioinformatics and computational genomics techniques, along with public and private databases, researchers can accurately impute the genetic sequence information of individuals without access to their biospecimens. While this can yield new discoveries and vital data for improving diagnostics, it also raises complex questions regarding the need to obtain informed consent from research participants about whom data is imputed via in silico research. The law of informed consent, codified before the development of powerful current technologies, does not address issues arising from the use of estimated data.

Proposed changes to U.S. informed consent regulations, which were ultimately defeated, would have provided protection for research subjects by requiring informed consent for the use of even their non-identified biospecimens, whether clinical or from prior research. Presently, researchers can use non-identified specimens without consent by stripping them of identifiers. The NIH already indicates that, as a condition of its funding, it expects researchers to obtain informed consent from research participants not just for the use of their cell lines or clinical specimens, such as tissue samples, but also for genomic data, even when stripped of information that directly identifies the source. These recently proposed and current changes reflect the view that researchers ought to respect the privacy and autonomy of research participants in an era where re-identification of research subjects has become easier to achieve. While a liberal reading of the proposed federal rule changes supports the notion that those from whom estimated data is gathered are entitled to the same rights of informed consent, privacy, and autonomy as conventional research subjects, the proposed rule changes contemplated only research subjects who contribute biospecimens or identifiable private information, whether wittingly or not.

Paradoxically, notwithstanding what seems to be a growing recognition that research subjects need enhanced protection from re-identification through their biospecimens, there appears to be a decline in the acceptance of an individual's RNTK her genetic risk profile. Increasingly, professional societies and IRBs see advantages in requiring medical professionals to test for certain genetic disorders and convey those findings to those who were tested, even without their informed consent. Taken together, these developments raise the troubling possibility that individuals will be involved in genetic research without their explicit consent, and then informed against their will of the results

of such research. Further policies must be developed to protect research participants from a system that would conscript them into research and then foist the results upon them. There are many reasons, whether emotional, religious, cultural, or even pragmatic (i.e., avoidance of discrimination) that individuals reject participation in genomic research and the incidental findings that it might provide to them about their own health. These individual choices not to participate and not to learn incidental findings deserve legal protection, whether the data is accessed conventionally or via advanced computational methods. The fundamental precepts upon which informed consent rest, including the Nuremberg Code, suggest that human dignity requires no less.