The American Journal of Bioethics, 10(9): 3–11, 2010
Copyright © Taylor & Francis Group, LLC
ISSN: 1526-5161 print / 1536-0075 online
DOI: 10.1080/15265161.2010.494215

Target Article

Is Deidentification Sufficient to Protect Health Privacy in Research?

Mark A. Rothstein, University of Louisville School of Medicine

The revolution in health information technology has enabled the compilation and use of large data sets of health records for genomic and other research. Extensive collections of health records, especially those linked with biological specimens, are also extremely valuable for outcomes research, quality assurance, public health surveillance, and other beneficial purposes. The manipulation of large quantities of health information, however, creates substantial challenges for protecting the privacy of patients and research subjects. The strategy of choice for many health care providers and research institutions in dealing with this challenge has been to deidentify individual health information.

Under the current regulatory framework in the United States, studies involving deidentified health records are exempt from regulations governing research with human subjects (45 C.F.R. § 46.101(b)(4)). Similarly, deidentified health records are outside the definition of "protected health information" (45 C.F.R. § 164.514(a)) and therefore are exempt from federal privacy protections. Determining whether legal requirements for privacy protection have been satisfied for deidentified health information usually involves narrowly evaluating the technical standards used in deidentification. There is usually little or no consideration by institutional review boards and regulators of the broader issues of the risks to privacy raised by the research and whether reasonable measures have been taken to reduce the risk.

This article considers the effects on privacy and related interests of creating, using, and disclosing deidentified health information in research without the knowledge, consent, or authorization of the individual. It also evaluates other potential harms from the nonconsensual use of deidentified health information in research, including undermining of trust in research. Many of the same issues apply to the use of deidentified biological specimens in research, and the article addresses these issues as well.

The article concludes that the use of deidentified health information and biological specimens in research creates a range of privacy and other risks to individuals and groups. The current regulatory system, under which privacy protections are afforded identifiable information but no protections apply to deidentified information, needs to be revised. A range of options should be considered to extend some level of protection to deidentified information without unduly burdening research.

DEFINING "DEIDENTIFIED" INFORMATION

At the outset, it is important to acknowledge the definitional and technological thicket raised by deidentification. Deidentified information is information that has been altered to remove certain data elements associated with an individual. (The HIPAA Privacy Rule definition is discussed later.) Deidentified information is one of many intermediate degrees of identifiability between "anonymous" information, which has no direct or indirect identifiers at the time of collection and which cannot be linked to any individual, and information containing "direct identifiers," such as name or Social Security number. Current legal requirements are bimodal: If the information is "identifiable," then all of the legal protections are applicable; if the information is "not identifiable," then there are no protections whatsoever.

There is no justification for perpetuating this dichotomy. Identifiability exists on a continuum, and the range of deidentification techniques, such as pseudonymization, linking, anonymization, and single and double coding (Knoppers 2005; Weir and Olick 2004), illustrates the fundamental problem with the bimodal approach. The need for a new privacy policy becomes clearer when biological specimens are involved, either in stand-alone form or linked with individual health information.

DEIDENTIFIED INFORMATION IS UNREGULATED

The regulatory distinction between identified and deidentified information is long-standing. The Federal Policy for the Protection of Research Subjects (Common Rule) provides that research involving anonymous or deidentified information is expressly exempt from regulation under the Common Rule. Exemption 4 from the Common Rule applies to the following:

> (4) Research involving 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

Address correspondence to Mark A. Rothstein, Institute for Bioethics, Health Policy and Law, University of Louisville School of Medicine, 501 East Broadway, Suite 310, Louisville, KY 40202, USA. E-mail: mark.rothstein@louisville.edu

> that subjects cannot be identified, directly or through identifiers linked to the subjects. (45 C.F.R. § 46.101(b)(4))

Similarly, the Health Insurance Portability and Accountability Act (HIPAA) Standards for Privacy of Individually Identifiable Health Information (Privacy Rule) (45 C.F.R. Parts 160, 164) applies only to "protected health information." The Privacy Rule provides that "protected health information means individually identifiable health information" (45 C.F.R. § 164.501). Furthermore, "[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" (45 C.F.R. § 164.514(a)).

Unlike the Common Rule, which describes the conditions for exemption based on lack of identifiability in general terms, the Privacy Rule goes into great detail about the requirements for deidentification. According to the Privacy Rule, there are two ways in which a covered entity may determine that information is deidentified. First, an expert in statistical and scientific methodologies may determine "that the risk is very small that the information could be used . . . to identify an individual who is a subject of the information" (45 C.F.R. § 164.514(b)(1)). Second, because of the difficulty and expense of obtaining expert consultation, a more prescriptive method of achieving deidentification also is provided in the Privacy Rule.

The Privacy Rule lists 17 specific provisions and one general provision regarding the types of identifiers that must be removed from health information before the information will be deemed deidentified. The following identifiers must be removed: (1) names; (2) geographical subdivisions smaller than a state except for the first three digits of a ZIP code; (3) all elements of dates (except year) that relate to birth date, admission date, and discharge date; (4) telephone numbers; (5) FAX numbers; (6) e-mail addresses; (7) Social Security numbers; (8) medical record numbers; (9) health plan beneficiary numbers; (10) account numbers; (11) certificate or license numbers; (12) vehicle identifiers, including license-plate numbers; (13) device identifiers and serial numbers; (14) URLs (web locators); (15) Internet protocol (IP) address numbers; (16) biometric identifiers; (17) photographic and comparable images; and (18) any other unique identifying number, characteristic, or code (45 C.F.R. § 164.514(b)(2)(i)). Compliance with these deidentification specifications eliminates a variety of obligations of covered entities under the Privacy Rule, including providing a notice of privacy practices, requiring an authorization for uses other than treatment, payment, and health care operations (subject to exceptions, such as public health disclosures), and restricting use of the information beyond health care. The Privacy Rule also permits covered entities to use a limited data set for purposes of research, public health, or health care operations if the recipient of the data set enters into a data use agreement specifying that the recipient will only use the information for limited purposes (45 C.F.R. § 164.514(e)(3) and (4)). The limited data set may not include "direct identifiers of the individual or of relatives, employers, or household members of the individual" (45 C.F.R. § 164.514(e)(2)). The impermissible "direct identifiers" include 16 of the 18 identifiers listed in the deidentification specifications mentioned earlier. The two categories of identifiers that may be included in a limited data set are dates, including date of birth and dates of service, and "any other unique identifying number, characteristic, or code."

Under the Privacy Rule, a covered entity may disclose protected information in a limited data set only if the recipient signs a data use agreement indicating that the information will be used only for limited purposes. In particular, the data use agreement must include the permitted uses and disclosures; indicate who is permitted to use and disclose the information; indicate that the recipient will not redisclose the information; provide that the recipient will use appropriate safeguards to prevent unapproved uses; provide that the recipient will report to the covered entity any use or disclosure not authorized by the data use agreement; provide that the recipient will ensure compliance with the agreement by any agents or subcontractors it uses; and provide that the recipient will not identify the information or contact the individuals (45 C.F.R. § 164.514(e)(4)).

The deidentification and limited data-set provisions of the Privacy Rule differ sharply from the Common Rule in both degree of detail and substance. According to a guidance document issued by the Office of Human Research Protections (OHRP), private information or specimens are "[not] individually identifiable when they cannot be linked to specific individuals by the investigator(s) either directly or through coding systems" (OHRP 2008). Furthermore, research involving only coded private information does not involve human subjects if the investigator cannot "readily ascertain" the identity of the individual because the key has been destroyed before the research begins, the keyholder has agreed not to release the key to investigators under any circumstances, there are institutional review board (IRB)-approved written policies prohibiting release of the key until individuals are deceased, or there are other legal requirements prohibiting the release of the key to the investigators until the individuals are deceased.

In its guidance, the OHRP recognized that it created a lower standard for deidentification under the Common Rule than exists under the Privacy Rule. "Therefore, some coded information, in which the code has been derived from identifying information linked to or related to the individual, would be individually identifiable under the Privacy Rule, but might not be individually identifiable under the [Common Rule]" (OHRP 2008). In the OHRP guidance, the Department of Health and Human Services (HHS) has explicitly acknowledged it has two different sets of rules regulating deidentification of health information for research. Notwithstanding the issue of whether deidentification is an adequate privacy strategy, the deidentification regulations of the Common Rule and the Privacy Rule are inexplicably and unjustifiably inconsistent. Although the Common Rule applies to other types of information besides health information, addressing the "welfare" of research subjects and

not just privacy, HHS has not attempted to harmonize these important regulations (Rothstein 2005).

THE INAPPLICABILITY OF MEDICAL CODES OF ETHICS

Deidentified health information and biological specimens are not covered by the codes of ethics applicable to physician-investigators. The Hippocratic Oath exhorts physicians to safeguard the confidentiality of patient health and other private information. "What I may see or hear in the course of the treatment or even outside of the treatment in regard to the life of men, which on no account must be spread abroad, I will keep to myself, holding such things shameful to be spoken about" (Oath of Hippocrates 1995, vol. 5, App., 2632). The American Medical Association's Principles of Medical Ethics provides that a physician "shall safeguard patient confidences and privacy within the constraints of the law" (AMA 2008, Principle IV, at xv). This means that confidential information should not be revealed without the express permission of the patient unless required by law or compelled by "overriding considerations," such as where the patient threatens to inflict serious harm on another person (AMA 2008, § 5.05).

Another provision of the AMA Code of Ethics provides as follows: "Only physicians or other health care professionals who are involved in managing the patient, including providing consultative, therapeutic, or diagnostic services, may access the patient's confidential medical information. All others must obtain explicit consent to access the information" (AMA 2008, § 7.025). Although this provision might appear to apply to physician-researchers, it is in a section titled "Records of Physicians: Access by Non-Treating Medical Staff." There is no indication of an intent to apply the provision to research.

In accord with the AMA Code of Ethics, none of the privacy and confidentiality provisions of the ethical codes of specialty medical societies, other health professions, or scientists specifically addresses the research uses of deidentified health information or biological specimens. A likely explanation is that the issue has never been considered.

RISKS TO INDIVIDUALS AND GROUPS

Under the Common Rule and Privacy Rule individuals have no legally protected interest in their deidentified health information used for research. The question is whether individuals *ought* to have such an interest. This section explores the ways in which the use of deidentified health information poses privacy risks or otherwise might be objectionable to the individuals whose records are accessed, analyzed, and disclosed without their knowledge, consent, or authorization.

The Process of Removing Identifiers

Deidentification is the process by which the identifiability of health information is reduced by removing certain information associated with a particular individual. For paper health records, the process is slow and cumbersome; it involves the manual, individual deletion of information through "whiting out" certain entries, extracting deidentified information from a record to create a new file, or taking other steps to render identifiable information deidentified. During this process, the individuals removing the identifiers are working with identifiable information. Therefore, possibly sensitive health information is being manipulated by individuals without any contemperaneous patient care responsibilities and without the knowledge, consent, or authorization of the individual. Also, there is no regulation of what individuals may perform this function. Thus, a health care provider could contract with a business associate that merely hires a cadre of temporary employees and equips each with a large bottle of "Wite-Out."

The process of deidentifying electronic health records (EHRs) depends on the particular EHR system being used (Wellner et al. 2007). EHR systems generally have not been designed to permit "one-click" deidentification. An EHR system's ability to deidentify health information is complicated. It depends on the system's ability to remove overt identifiers in demographic and other fields; to analyze "free text" messages; to deidentify laboratory, imaging, and other reports that might be scanned into the EHR; and to meet various other technological challenges (Neamatullah et al. 2008). For some EHR systems, deidentification may still require at least some use of the crude "white out" strategies used for paper records.

The process of deidentification is rarely, if ever, considered in discussions of whether privacy protections should extend to all individual health records generated or used in clinical settings. Unless the capacity to deidentify health information is designed into an EHR system's architecture, removing all identifiable information with computer applications may involve a multistep process. For individuals with sensitive information in their health records, there is a risk that information will be observed during the process of deidentification. The Privacy Rule, however, expressly provides that no consent or authorization is needed before deidentification is undertaken (45 C.F.R. § 164.502(d)(1)), and there is no regulation of the procedures used.

Irrespective of the other recommendations in this paper, the Office for Human Research Protections at HHS (which administers the provisions of the Common Rule) and the Office for Civil Rights at HHS (which administers the HIPAA Privacy Rule) should issue joint guidance on acceptable methods and procedures for the deidentification of health information. It is not enough merely to list the types of data that must be removed.

Reidentification

Despite using various measures to deidentify health records, it is possible to reidentify them in a surprisingly large number of cases by using computerized network databases containing voter registration records, hospital discharge records, commercially available databases, and other sources (Malin and Sweeney 2004; Sweeney 2002). Indeed, it is likely that between 63% (Golle 2006) and 87% (Sweeney 2000) of the population of the United States could

be uniquely identified by using only gender, ZIP code, and date of birth. The cost of doing so, however, would vary by state, because of the different prices charged for voter registration data (Benitez and Malin 2010).

Reidentification of genomic samples in biobanks is also possible using publicly available databases, thereby raising the question of whether genetic information can ever be considered deidentified in the sense that it cannot be linked with other genetic samples (McGuire and Gibbs 2006). After a scientific paper demonstrated it was theoretically possible to identify an individual's genomic attribute data in a pooled or aggregated sample (Homer et al. 2008), the National Human Genome Research Institute immediately restricted public access to pooled genomic data.

There are no published studies on the degree of compliance of covered entities with the HIPAA-mandated measures to deidentify health information. Assuming optimal compliance with all of these requirements, Dr. Latanya Sweeney, the leading expert on deidentification and reidentification of health records, has calculated that it is possible to achieve reidentification in 0.04% of the records using only allowed demographics (NCVHS 2007, 36 n.16). Although this number might seem low, with some databases containing hundreds of thousands of records, the number of records that could be reidentified is quite substantial. In addition, in actual practice the deidentification measures often fail to achieve the theoretical optimum results and therefore the percentage of health records that could be reidentified is actually much greater than 0.04%. Finally, if a limited data set is used, which includes date of birth and date of service, a high percentage of the records can be reidentified using publicly available databases.

The risks of reidentification are more than theoretical. Dr. Sweeney was involved in one of the most celebrated incidents demonstrating the ease of reidentification. In the mid-1990s, in the interest of promoting health services research, the Massachusetts Group Health Insurance Commission released "anonymized" data on state employees that showed every single hospital visit. Then-governor William Weld assured the public that privacy was completely protected by removing identifiers such as name, address, and Social Security number. Dr. Sweeney, then a graduate student at MIT, obtained the hospital discharge data, compared it with publicly available voter registration information, and quickly identified the health records of Governor Weld (Shaw 2009). This disclosure led to a hasty change in state policy.

According to Dr. Sweeney, it is possible to translate identifiable data into provably anonymous data using privacy technology, but standard deidentification techniques embodied in the Privacy Rule offer no guarantee of anonymity (Sweeney, personal communication, May 20, 2009; Benitez and Malin 2010; Malin 2006). To achieve anonymity, technology has to make fine-grained decisions specific to the data. Two policy questions are raised. First, should the application of heightened deidentification technology be the goal when it might be costly and make the records less valuable for research? Second, enacting legislation prohibiting reidentification might have a deterrent effect, but are additional measures needed to address the harms associated with even completely deidentified data?

Group Harms

Deidentification customarily removes individual information from health records. It does not usually remove information about an individual's membership in certain groups defined by race, ethnicity, gender, religion, or other criteria. Consequently, research using deidentified health records could lead a researcher to conclude that members of a certain group have an X percent increased risk of developing Y health condition. If Y is a particularly stigmatizing condition, then all members of the X group could be said to suffer a "group-based harm," or what Dr. Daniel M. Hausman has termed a "group-mediated harm to individuals" (Hausman 2007, 354).

There are many ways in which group harms can be expressed, including loss of status in the majority society, self-stigmatization, and dignitary harms to the community (Freeman, Romero, and Kanade 2006, 134–139). The potential for group-mediated harms is often associated with genetic research, because certain population groups defined socially by race or ethnicity often have higher frequencies of certain genotypes based on historical patterns of migration, isolation, endogamy, founder effect, or other principles of population genetics (Hartl and Clark 2007, 519–563). Nevertheless, the potential for group-based harms from research using deidentified health records is not limited to genetic research.

An example of a group-based harm involves the Havasupai Indian Tribe, a 650-member tribe living in an isolated and remote part of the Grand Canyon. In 1989, members of the tribe approached researchers at Arizona State University, asking for help to stem the tribe's high incidence of diabetes. The tribal council approved collecting and testing blood samples to study diabetes. Allegedly without the tribal members' knowledge or consent, the researchers also studied and published articles about schizophrenia, inbreeding, and migration of the tribe. The migration studies differed from ancestral creation stories of the tribe. Thus, researchers linked the Havasupai with a stigmatizing condition, as well as undermining a fundamental cultural belief system. In both regards, the investigators allegedly performed research that exceeded the consent. Lawsuits brought by the tribe against the researchers and Arizona State University were settled in 2010. Arizona State paid $700,000 to 41 members of the tribe, and agreed to provide other assistance to the tribe (Harmon 2010).

The harm caused by associating an increased risk of a stigmatizing condition (e.g., mental illness) with a particular group, especially a minority or socially vulnerable group, attaches to each member of the group regardless of whether that person's health record or biological specimen was used in the research. Even if certain members of the group provided informed consent for research with their samples, any resulting stigma would be shared with nonparticipating members of the group. Clearly, the use of deidentified

information does not create the risk of group harm, which would exist for any research (Sharp and Foster 2007). The key point is that the use of deidentified information does not eliminate existing risks for socially vulnerable groups.

Researchers have an obligation to minimize group-based harms and to demonstrate respect for the group being studied and its members. Some of the possible ways of addressing these concerns in the context of deidentified information include the following: (1) Individuals should, at least, have the opportunity to opt out of having their deidentified information and specimens used for research; (2) information provided to individuals about the use of their deidentified information and specimens should include disclosures about possible group-based harms, even if group-based associations are not a focus of the research and would be considered incidental findings; and (3) publications and public pronouncements by researchers containing group associations should be done with extraordinary caution and precision.

Objectionable Uses

Under current law and practice, neither informed consent (under the Common Rule) nor authorization (under the Privacy Rule) is required before research is undertaken with deidentified health records or biological samples. Similarly, individuals whose deidentified records or samples are used need not be provided with notice of such use or an opportunity to opt out of the research (Clayton et al. 1995). This includes research the individual might consider objectionable from a religious, moral, ethical, or other standpoint.

One example of objectionable research is raised by the following scenario: Researchers at University A obtain deidentified health records and pathology specimens from University Hospital to conduct genetic research related to a particular genetic disorder. The researchers identify the responsible gene mutation, develop a test for the mutation, publicize development of the new test, and the test becomes incorporated into the battery of prenatal genetic tests routinely offered to pregnant women. One result of offering the test is to increase the number of abortions involving fetuses expressing or predisposed to this genetic condition.

An individual with the particular genetic disorder who was a patient at University Hospital might deduce that his or her health records and samples were used for research without his or her permission. If that individual considers the research and its consequences morally objectionable, he or she has been wronged by University A. It is true that the discovery might well have occurred without using any one person's sample, but an important part of what makes the conduct personally objectionable to certain individuals is knowing that their health information and biological specimens contributed, at least in part, to the discovery and subsequent uses of the genetic test. Many patients want to know the purpose of research to prevent their specimens from being used in objectionable ways (Hull et al. 2008, 66).

Although performing research only pursuant to consent would eliminate the basis of the claim of objectionable use, it is currently not legally required or customarily sought. Deidentified records and biological specimens could be used in a wide range of controversial research, including somatic cell nuclear transfer, stem-cell research, and germ-line gene therapy.

Commercial Exploitation

Research using health records and biological specimens sometimes results in discoveries with commercial value. Partly in response to some highly publicized lawsuits in which research participants have sued researchers for revenue derived from using their information and biological specimens, it has become common for informed consent documents signed by research subjects to disclaim any economic interest in possible commercial applications flowing from the research. The use of disclaimers has been criticized, however, as failing to provide an ethically acceptable level of benefit sharing with the donors of health information and biological specimens (Andrews 2005).

Research using deidentified records and specimens is even more problematic because there is no informed consent and thus no disclaimer. There is also no notice of the research or an opportunity to opt out of participation. Often, the first time an individual learns his or her information or specimen has been used is when there are media reports that Institution B, using patient-derived data, has discovered a new test or treatment for a certain disorder, which Institution B has or will patent and make commercially available. At least some of the individuals whose materials were used in the research without their knowledge or consent are likely to believe they were exploited by the researchers and Institution B.

A limited number of high-profile commercial exploitation claims have been raised where a treating physician (*Moore v. Regents of the Univ. of Cal.*, 793 P.2d 479 [Cal. 1990]) or researcher (*Greenberg v. Miami Children's Hosp. Research Inst., Inc.*, 264 F. Supp.2d 1064 [S.D. Fla. 2003]) failed to disclose a commercial interest in biological materials; where a research institution asserted ownership claims over specimens at variance with the principal investigator and the informed consent agreement (*Washington Univ. v. Catalona*, 490 F.3d 667 [8th Cir. 2007]); and where samples were obtained from indigenous peoples without informed consent (WHO 2009). It is not clear whether allegations of exploitation would follow the nonconsensual use of only deidentified health records apart from biological specimens.

Undermining Trust

For many individuals there is no difference between health care providers and researchers, especially when the providers and researchers work for the same institution and patient-based clinical records and specimens are used in the research (Katz 1993; President's Advisory Committee 1995, 468; Rothstein 2009). Therefore, any disillusionment with perceived research abuse is likely to result in a general loss of trust in clinicians, researchers, institutions, and the health care enterprise. Various negative consequences could

flow from a loss in trust, including individuals delaying or foregoing treatment (with possible adverse effects on individual and public health); utilizing ineffective, nontraditional health care providers; seeking care only at institutions that do not engage in research; refusing to participate in clinical trials; and being reluctant to support public expenditures for health research.

Trust in clinicians, researchers, and institutions has been undermined in recent years by financial incentives to undertreat or overtreat patients, conflicts of interest involving pharmaceutical companies, and limits on access to physician services attributable to managed care (Kao et al. 1998). In some minority communities, the lack of trust has resulted from notorious abuses by researchers and from discrimination and callous indifference by the health care system (Bussey-Jones et al. 2010; Skloot 2010). This has led to difficulty in recruiting minority populations in research. Trust would be further eroded if individuals learned that their health records and specimens were used without their knowledge, consent, or authorization for objectionable purposes, that they have been stigmatized by group-mediated harms, or that they have been exploited for commercial gain (Burger 2009). These individuals are unlikely to be persuaded by utilitarian arguments of the need for unfettered research or assertions that their injuries are de minimis (IOM 2009).

AUTONOMY AND BIOMEDICAL RESEARCH

The current regulatory frameworks of the Common Rule and Privacy Rule emphasize privacy interests, but they overlook the autonomy interests of individuals whose health information and biological specimens are used in research without their knowledge, consent, or authorization. Autonomy "encompasses at a minimum, self-rule that is free from both controlling interference by others and from certain limitations such as an inadequate understanding that prevents meaningful choice" (Beauchamp and Childress 2009, 99). Currently, individuals have no control over the use of their deidentified information and specimens. The most likely justification for such a policy is that individuals do not have a protectable interest in their records or specimens, the risk of harm from the research is insubstantial, or individual interests are outweighed by the beneficial uses of the information. Just as competent adults have the right to decide what is done to their bodies, they should have reasonable control over their health information and biological specimens generated by medical encounters, regardless of whether their information and specimens have been deidentified.

Recent survey research confirms the importance of autonomy in individuals' opinions about the use of their health information and biological specimens in research (Goldenberg et al. 2009; Kaufman et al. 2009; Westin 2007). Most individuals will agree to the use of their information and specimens for research, but they want to be asked. Furthermore, autonomy is only one part of the broader concept of respect for persons, which also includes "attention to important subjective experiences, persons' existence as part of communities, and considerations of comportment" (Dickert 2009, 311).

Survey research also clearly demonstrates that the public does not follow the regulatory distinction between identifiable and deidentified samples and information. Hull and colleagues (2008) surveyed 1193 patients recruited from general medicine, thoracic surgery, and medical oncology clinics at five academic medical centers. When asked whether "anonymous" biological samples could be used for research, 57% said that researchers should be required to obtain permission and 43% said that notification was sufficient. (Note: Neither permission nor notification is currently required.) Patients were more likely to support permission if they had more education, were Black, were less religious, were in better health, were more private, and were less trusting of researchers.

Similar findings have been reported in surveys of the public with regard to health records. Westin (2007) found that 38% of respondents would require consent for research with their deidentified health records and 13% did not want their records used under any circumstances. Only 28% said that no individual consent was needed for research with deidentified records; 20% said they were unsure. Thus, the public does not share the distinction between research using identifiable versus deidentified samples and records used by both the Common Rule and Privacy Rule. In both situations, most members of the public want to control the use of their samples and information as an element of personal autonomy.

ANTICIPATING THE LIKELY CRITICISMS

At a time when many researchers favor lessening existing regulations protecting research participants (IOM 2009, ch. 5; Wartenberg and Thompson 2010), a proposal to consider extending protections to research involving deidentified health information and biological specimens is not likely to be well received by the research community. Among the likely criticisms are that the proposal is infeasible, burdensome, time-consuming, unnecessary, expensive, and would lead to selection bias in research. Most importantly, it probably would be asserted that compliance will delay and perhaps prevent the development and introduction of medical innovations, pharmaceutical products, and medical devices to improve the health of millions of people.

These concerns should not be dismissed lightly. On the other hand, the interests of patients and the public also deserve respect and consideration. At present, there is an insufficient empirical basis to assert that adding some level of privacy and autonomy protection to deidentified health information and biological samples will invariably and unreasonably disrupt biomedical research. Ultimately, the policy question may well be whether some degree of inconvenience or burden to researchers and some level of imprecision in research methodology is an acceptable price to pay for safeguarding the privacy and autonomy of

individuals at a time of increasing computerization of health information and greater scale of genomic and other research technologies.

This paper raises numerous complicated and contentious issues, but it attempts to resolve only one. That issue is whether all deidentified health information and biological specimens should be categorically excluded from the regulatory domain and the bioethics discourse. It concludes that utilitarian concerns about burdens on research are insufficient to justify dispensing with any consideration of the possible effects of the research on the individuals from whom the information and specimens were obtained.

Unquestionably, research often benefits the individual sources of the information and specimens as well as society as a whole. Nevertheless, the world of research has changed substantially from the time the Common Rule was developed. The increased scale of research and new computer technologies demand a more nuanced assessment of the risks and benefits of research using a range of deidentified information and biological materials.

CONCLUSION: GOING FORWARD

By itself, the current strategy of deidentifying health records and biological specimens is insufficient to protect privacy and respect autonomy in research. It is indefensible from technical, ethical, and policy standpoints to continue drawing a bright-line regulatory distinction between identifiable and deidentified health information (Ohm 2010). The National Bioethics Advisory Commission (NBAC), in the context of research involving human biological materials, stated: "Generally, it is NBAC's view that when it is feasible to conduct human biological materials research that is in accordance with the usual protections for research subjects, it is preferable to do so, rather than to unlink the samples in order to circumvent those protections" (NBAC 1999, vol. 1, 61). Sole reliance on deidentification of health records and biological specimens to protect privacy and autonomy similarly represents an unacceptable circumvention of the essential principles of research ethics.

The limitations of deidentification as a solitary privacy strategy do not mean that it should be discarded entirely as a tool to protect privacy. Deidentification should be considered a necessary but insufficient means of protecting health privacy. In accordance with this view, health information should be collected, maintained, disclosed, and used in the least identifiable form consistent with the purpose of the information. Although the analogous principle of "minimum necessary" is a cornerstone of the Privacy Rule (45 C.F.R. § 164.502(b)), "least identifiable form" has yet to be adopted explicitly or implemented as a regulatory strategy.

Of the key unresolved issues surrounding deidentified information, many involve the degree of protection to be afforded different types of information. These issues include the following: (1) the proper coordination of deidentification with other measures to protect privacy; (2) the degree to which rules for protecting the privacy of deidentified information should align with those applicable to identifiable information; (3) the specific means by which privacy should be protected in deidentified information, such as notice provisions, consent management tools, or opt-out strategies; (4) the degree to which individual control over biological specimens should be consistent with control over health information; (5) whether deidentified health information for nonresearch uses, such as quality assurance and public health, also should be regulated; and (6) whether additional protection should be afforded all deidentified information or only deidentified health information. The regulatory and ethical obligations of researchers might vary based on the type of research, degree of deidentification, vulnerability of the individuals, and other factors, but specific proposals to address these issues are beyond the scope of this article.

Both the Common Rule and the Privacy Rule took years to develop, and they included input from researchers, research administrators, research sponsors, experts on research ethics, advocates for patients and research subjects, government officials, other stakeholders, and members of the public. A detailed process of public engagement, pilot projects, and careful study is needed before any type of regulatory coverage should be extended to deidentified health information and biological specimens. In the interim, responsible researchers should consider whether, in the context of their particular research, additional measures are needed to protect deidentified health information and biological specimens and demonstrate respect for the individuals from whom the information and specimens were obtained.

Those who engage in research ought to be as thoughtful and meticulous about their relations with the human subjects of their research as they are about designing their experiments and analyzing their data. The ethical implications of research on human subjects go beyond a single protocol. As Hans Jonas cautioned:

> Let us also remember that a slower progress in the conquest of disease would not threaten society, grievous as it is to those who have to deplore that their particular disease be not yet conquered, but that society would indeed be threatened by the erosion of those moral values whose loss, possibly caused by too ruthless a pursuit of scientific progress, would make its most dazzling triumphs not worth having. (Jonas 1969, 245)

REFERENCES

American Medical Association. 2008. *Code of medical ethics 2008–2009 edition*. Chicago: American Medical Association.

Andrews, L. B. 2005. Harnessing the benefits of biobanks. *Journal of Law, Medicine and Ethics* 33: 22–30.

Beauchamp, T. L., and J. F. Childress. 2009. *Principles of biomedical ethics*. 6th ed. New York: Oxford University Press.

Benitez, K., and B. Malin. 2010. Evaluating re-identification risks with respect to the HIPAA privacy rule. *Journal of the American Medical Informatics Association* 17(2): 169–177.

Burger, J. A. 2009. What is owed participants in biotechnology research? *Chicago-Kent Law Review* 84: 55–89.

Bussey-Jones, J., J. Garrett, G. Henderson, et al. 2010. The role of race and trust in tissue/blood donation for genetic research. *Genetics in Medicine* 12: 116–121.

Clayton, E. W., K. K. Steinberg, M. J. Khoury, et al. 1995. Informed consent for genetic research on stored tissue samples. *Journal of the American Medical Association* 274: 1786–1792.

Dickert, N. W. 2009. Re-examining respect for human research participants. *Kennedy Institute of Ethics Journal* 19: 311–338.

Freeman, W. M., F. C. Romero, and S. Kanade. 2006. Community consultation to assess and minimize group harms. In *Institutional review board management and function*, eds. E. A. Bankert and R. J. Amdur. 2nd ed., 134–139. Sunderland, MA: Jones and Bartlett.

Goldenberg, A. J., S. C. Hull, J. R. Botkin, et al. 2009. Pediatric biobanks: Approaching informed consent for continuing research after children grow up. *Journal of Pediatrics* 155: 578–583.

Golle, P. 2006. Revisiting the uniqueness of simple demographics in the U.S. population. In *Workshop on privacy in the electronic society*, 77–80. New York: Association for Compute Machinery. Available at: www.trustste.org/wise/articles2009/articleM3.pdf (accessed March 8, 2010).

Harmon, A. 2010. Tribe wins fight to limit research of its DNA. *New York Times* April 22, 2010, A1.

Hartl, D. L., and A. G. Clark. 2007. *Principles of population genetics*. 4th ed. Sunderland, MA: Sinauer Associates.

Hausman, D. M. 2007. Group risks, risks to groups, and group engagement in genetics research. *Kennedy Institute of Ethics Journal* 17: 351–369.

Homer, N., S. Szelinger, M. Redman, et al. 2008. Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genetics*. Available at: http://dx.doi.org/a0.1371%2Fjournal.pgen.1000167 (accessed February 19, 2010).

Hull, S. C., R. R. Sharp, J. R. Botkin, et al. 2008. Patients' views on identifiability of samples and informed consent for genetic research. *American Journal of Bioethics* 8(10): 62–70.

Institute of Medicine. 2009. *Beyond the privacy rule: Enhancing privacy, improving health through research.* Washington, DC: National Academies Press.

Jonas, H. 1969. Philosophical reflections on experimenting with human subjects. *Daedalus* 98: 219–247.

Kao, A. C., D. C. Green, A. M. Zaslavsky, et al. 1998. The relationship between physician payment and patient trust. *Journal of the American Medical Association* 280: 1708–1714.

Katz, J. 1993. Human experimentation and human rights. *St. Louis University Law Journal* 38: 7–54.

Kaufman, D. J., J. Murphy-Bollinger, J. Scott, et al. 2009. Public opinion about the importance of privacy in biobank research. *American Journal of Human Genetics* 85: 643–654.

Knoppers, B. M. 2005. Biobanking: International norms. *Journal of Law, Medicine & Ethics* 33: 7–14.

Malin, B. 2006. Re-identification of familial database records. *Proceedings of the American Medical Informatics Annual Symposium* 2006: 524–528.

Malin, B., and L. Sweeney. 2004. How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems. *Journal of Biomedical Informatics* 37: 179–192.

McGuire, A. L., and R. A. Gibbs. 2006. No longer de-identified. *Science* 312: 370–371.

National Bioethics Advisory Commission. 1999. *Research involving human biological materials: Ethical issues and policy guidance*. Rockville, MD: National Bioethics Advisory Commission.

National Committee on Vital and Health Statistics. 2007. *Report to the Secretary of Health and Human Services on enhanced protections for "secondary uses" of electronically collected and transmitted health data*. Available at: www.ncvhs.hhs.gov/071221lt.pdf (accessed February 19, 2010).

Neamatullah, I., M. M. Douglas, L. H. Lehman, et al. 2008. Automated de-identification of free-text medical records. *BMC Medical Informatics and Decision Making*. Available at: http://www.biomedcentral.com/1472-6947/3/32 (accessed February 19, 2010).

Oath of Hippocrates. 1995. *Encyclopedia of Bioethics*, ed. W. T. Reich. New York: Simon & Schuster Macmillan.

Office of Human Research Protections. 2008. *Guidance on research involving coded private information or biological specimens*. Available at: http://www.hhs.gov/ohrp/humansubjects/guidance/cdebiol.htm (accessed February 19, 2010).

Ohm, P. 2010. Broken promises of privacy: Responding to the surprising failure of anonymization. University of Colorado Research Paper No. 09-12. Available at: http://ssrn.com/abstracts 145006.

President's Advisory Committee on Human Radiation Experiments. 1995. Subject interview study. In *Final Report of the Advisory Committee on Human Radiation Experiments*. Washington, DC: U.S. Government Printing Office.

Rothstein, M. A. 2009. Improve privacy in research by eliminating informed consent? IOM report misses the mark. *Journal of Law, Medicine & Ethics* 37: 505–512.

Rothstein, M. A. 2005. Research privacy under HIPAA and the common rule. *Journal of Law, Medicine & Ethics* 33:154–159.

Sharp, R. R., and M. W. Foster. 2007. Grappling with groups: Protecting collective interests in biomedical research. *Journal of Medicine and Philosophy* 32: 321–337.

Shaw, J. 2009. Exposed: The erosion of privacy in the Internet era. *Harvard Magazine* September–October: 38–43.

Skloot, R. 2010. *The immortal life of Henrietta Lacks*. New York: Crown.

Sweeney, L. 2002. K-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness, & Knowledge-Based Systems* 10: 557–570.

Sweeney, L. 2000. *Uniqueness of simple demographics in the U.S. population.* Available at: http://privacy.cs.cmu.edu.html (accessed March 8, 2010).

Wartenberg, D., and D. Thompson. 2010. Privacy versus public health: The impact of current confidentiality rules. *American Journal of Public Health* 100: 407–411.

Weir, R. F., and R. S. Olick. 2004. *The stored tissue issue: Biomedical research, ethics, and law in the era of genomic medicine.* New York: Oxford University Press.

Wellner, B., M. Huyck, S. Mardis, et al. 2007. Rapidly retargetable approaches to deidentification in medical records. *Journal of the American Medical Informatics Association* 14: 564–573.

Westin, A. F. 2007. *IOM project survey findings on health research and privacy*. Available at: www.iom.edu/~/media/Files/Activity%20Files/Research/HIPAAandResearch/AlanWestinIOMsrvyRept.ashx (accessed February 19, 2010).

World Health Organization. 2009. *Indigenous peoples & participatory health research.* Available at: www.who.int/ethics/indigenous_peoples/en/index6.html (accessed December 2, 2009).