

Published in final edited form as:

*Am J Bioeth.* 2010 September ; 10(9): 21–23. doi:10.1080/15265161.2010.494222.

## Managing Patient Expectations About Deidentification

**Shawneequa L. Callier** and

Center for Genetic Research Ethics and Law

**Harald Schmidt**

Harvard School of Public Health

Mark Rothstein (2010) convincingly illustrates that deidentification is an insufficient method for protecting patients' private medical information. The current regulatory distinction between identifiable and deidentified health records is misleading, and patient health data categorized as deidentified are vulnerable to privacy breaches and other harms. Most patients know little about the subtleties made in the law, but expect privacy and confidentiality to be respected, generally assuming that this will be the case when data are deidentified. Additionally, patients might expect to benefit personally from medical research findings derived from their deidentified data without realizing that deidentification limits a researcher's ability to provide feedback that is directly relevant to them (Hull et al. 2008). The success or failure of current efforts to use health information technology as a tool for advancing medical research is heavily dependent upon closing these types of gaps between patient expectations about deidentification and the research community's ability to meet them.

Electronic medical records, public health registries, and other population-based databases can raise privacy risks if linked with deidentified genomic data and used to further scientific knowledge (Loukides, Gkoulalas-Divanis, and Malin 2010; Lowrance and Collins 2007). While this would give investigators more opportunities to relate genetic information with clinical data and to hone in on the characteristics associated with particular genetic diseases, it would also make reidentification much easier than it was when the deidentification strategy was first conceptualized. Accordingly, as Rothstein explains, "in actual practice the deidentification measures often fail to achieve the theoretical optimum results." Meanwhile, low-cost genome sequencing may help raise reidentification risks. George Church contends that "cheap sequencing technology will ... make that information more meaningful by multiplying the number of researchers able to study genomes and the number of genomes they can compare to understand variations among individuals in both sickness and health" (Church 2006, 48). To the extent that more individuals see the value in sequencing their entire genome when it becomes more affordable (Church 2006, 48) and in sharing their genetic information with researchers, family, or members of their online social networks, reidentification by third parties could become exceedingly difficult to prevent.

At the same time, as technology and methods that allow reidentification improve, researchers' ethical duty to provide feedback to patients could become more relevant. The forms of linkage that Rothstein describes were generally not considered when deidentification guidance was first formed. One reason may be that the respective organizations felt they only needed to be concerned with rules for the professional groups that they represented. Another may be that other forms of linkage were simply not viewed as realistic. These phenomena are quickly changing. Further, as Charles Rotimi and Patricia

Marshall illustrate in their commentary, “Tailoring the Process of Informed Consent in Genetic and Genomic Research” (2010), researchers will need to provide more concrete and justifiable reasons for denying patient withdrawal or feedback requests when reidentification becomes more feasible and clinically useful:

In the past, most genomic research projects did not report results back to participants. This decision was due, for the most part, to the uncertain clinical relevance of research findings. It is, however, becoming increasingly difficult to justify this position, especially in the context of large-scale medical genotyping and sequencing research studies that are likely to generate clinically relevant genetic information. Examples of this type of genomic study include ClinSeq, the Coriell Personalized Medicine Collaborative, the Framingham Heart Study and the Jackson Heart Study. However, communicating genomic results to participants requires tailored consent documents that carefully consider ethical responsibilities and social obligations to participants and their relatives. (Rotimi and Marshall 2010, 3)

Considering these advancements in reidentification and genomic technology, we should expect patients to become progressively savvier about health information technology and its capabilities related to reidentification, including electronic research systems that permit the return of research results to patients.

Indeed, recent surveys have shown that most patients want some level of control over (or at least notification about) the use of their health information for research purposes, whether identifiable or deidentified (Hull et al. 2008, 63, 65–66). If patients require an ownership interest in their health records, or at least some level of control over access to them, “and if individuals have a right to consent to or refuse research participation, then it is not clear that deidentification makes the ethical problem go away” (Miller 2008, 560–561). Further, if the public deems the research process to be unsafe, unfair, or exploitative, patients could choose to protest against and impede the use of their medical records for research, or refuse to communicate honestly with physicians about their medical conditions.

Managing patient expectations can help encourage and sustain public trust related to the use of medical records for research or for other reasons. As we explore elsewhere more fully, one important strategy would be to abandon the use of terms such as “identified” and “deidentified” and replace them with a taxonomy that is less prone to providing false certainty or being misunderstood. As a concept, deidentification fluctuates between a commonsense understanding of the term and a range of highly technical legal meanings. Moreover, internationally, there are no agreed upon uniform definitions or standards for coding health data, and in key policy documents, various different taxonomies coexist to describe deidentified information (Schmidt, Callier, and Rogers forthcoming). While institutions may have developed their own guiding principles related to deidentification, the reidentification risks associated with data are rarely static as that information is exchanged across databases and institutions. This makes it difficult for researchers and patients alike to adequately calculate the reidentification risks associated with the future use of medical data.

In addition to clearer terminology, more appropriate consent strategies are also needed to better inform patients about the plausible scenarios in which linkage and reidentification could be achieved in a particular instance; the patient’s options for placing limits on the “scope, type, or location of the research” (Girod and Drabiak 2008, 226); and the potential for participants to seek to “benefit from any screening tests, treatments, or medications developed through research using their donations” (Girod and Drabiak 2008, 228). Through the use of more precise language, researchers can better inform patients when obtaining consent, and can mitigate some of the ethical challenges raised by wanting to respect patient autonomy while also needing to override some patient preferences.

The consent process and any discussion of the risks to individuals or groups, however, should be carefully formulated. As Sharona Hoffman reasons in a separate response to Rothstein's article, "An overabundance of caution about privacy may be ill-advised if it leads to skewed data. The danger is that large segments of the population or subgroups of individuals ... will decide to opt out of inclusion in databases" (Hoffman 2010). Prudent consent documents will avoid language that might unnecessarily encourage widespread opposition to participation. And more emphasis would be placed on empowering the patients of the future through patient and community education programs. Public education could encourage positive attitudes toward research participation, even when patients are unable to control every aspect of the research process.

Patient education schemes should be accompanied by robust privacy laws, as well as adequate guidance on and access to timely legal remedies. As Rothstein argues, federal law should expand to protect those forms of deidentified information vulnerable to reidentification. And the law should provide quick and efficient remedies if patients' privacy expectations regarding their health data are not met. In addition, public education campaigns should ensure that patients are aware of such laws. At the same time, funding institutions should continue to promote the development of software that can allow the tracking of shared data, the identification of parties responsible for unauthorized uses, and other programs that will secure data and hold violators accountable.

Rothstein advocates a "detailed process of public engagement, pilot projects, and careful study" before extending regulatory coverage to deidentified health information (Rothstein 2010). We suggest also funding educational programs that highlight the benefits, in addition to the risks associated with participating in research. And most importantly, patients must understand the limits of institutions, data-sharing systems, and fledgling research projects that are incapable of accommodating patient interests but have the potential to greatly advance clinical medicine. Health records will play an increasingly important role in health care research, especially as vast data sets are analyzed in highly distributed computer networks. The more members of the public are content for their data to be used on a reasonable basis, the more useful the databases will be, and the more sustainable the public's trust will be.

## Acknowledgments

Partial support for this essay came from the Postdoctoral fellowship program (Callier) of the Center for Genetic Research Ethics and Law, through the National Institutes of Health grant P50-HG003390 from the National Human Genome Research Institute. Harald Schmidt carried out this work while being supported by a Harkness Fellowship from the Commonwealth Fund. The views presented here are those of the author alone and should not be attributed to The Commonwealth Fund or its directors, officers, or staff or to CGREAL.

## References

- Church GM. Genomes for all. *Scientific American*. 2006; 294(1):46–54. [PubMed: 16468433]
- Girod J, Drabiak K. A proposal for comprehensive biobank research laws to promote translational medicine in Indiana. *Indiana Health Law Review*. 2008; 5(217):226–228.
- Hoffman S. Electronic health records and research: Privacy versus scientific priorities. *American Journal of Bioethics*. 2010; 10(9):19–20. [PubMed: 20818550]
- Hull SC, Sharp RR, Botkin JR, Brown M, Hughes M, Sugarman J, Schwinn D, Sankar P, Bolcic-Jankovic D, Clarridge BR, Wilfond BS. Patients' views on identifiability of samples and informed consent for genetic research. *American Journal of Bioethics*. 2008; 8(10):62–70. [PubMed: 19003716]

- Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies. *Proceedings of the National Academy of Sciences of the United States of America*. 2010; 107(17):7898–7903. [PubMed: 20385806]
- Lowrance WW, Collins FS. Ethics: Identifiability in genomic research. *Science*. 2007; 317(5838):600–602. [PubMed: 17673640]
- Miller FG. Research on medical records without informed consent. *Journal of Law, Medicine and Ethics*. 2008; 36(3):560–566.
- Rothstein MA. Is deidentification sufficient to protect health privacy in research? *American Journal of Bioethics*. 2010; 10(9):3–11. [PubMed: 20818545]
- Rotimi CN, Marshall PA. Tailoring the process of informed consent in genetic and genomic research. *Genome Medicine*. 2010; 2(3):20. [PubMed: 20346094]
- Schmidt H, Callier S, Rogers C. How anonymous is ‘anonymous’?—Some suggestions towards a coherent universal coding system for genetic samples. *Forthcoming*.