Patients and Biobanks

Ellen Wright Clayton
UNTIL recently, I usually saw patients in either our walk-in clinic or the hospital without access to their old charts. Given the nature of general pediatric practice in a teaching hospital like ours, I often did not know before I walked in the room whether the “five-year-old with fever” was a normal kindergartener who had entered school for the first time and was catching the obligatory viruses or a child with multiple medical problems who was taking numerous medications, not all of which were fully remembered or understood by the person who brought the child. In either event, the child’s prior chart might not appear until after the child had been sent home, if then. Things, of course, were not quite that bad; most of the children with chronic problems were well known to the clinic personnel, but still, gaps remained.

Things are different now. When a child comes in today, the person who is going to see the child, whether a student, resident or faculty, first sits down at a computer, which provides immediate access to the child’s past medical history and laboratory and radiologic evaluations. Most of these entries are actually legible, which is another bonus.

While the electronic medical record (EMR) is generally a huge help, it is a mixed blessing. Knowing a lot before you walk in the door narrows your focus, perhaps causing you to miss something that you would have seen with a more open mind. Computer screens, which are present in many of our examining rooms, enable the care team—care technicians, nurses and physicians—to enter data directly into the medical record. This efficiency, however, comes at some cost: the care provider either must sit with his or her back to the patient to enter data or the computer screen is interposed between them. The presence of the computer is a visible reminder of the reality that health information is not confined to the examining room. Nor can technology guarantee accuracy. Sometimes, as much as we hate to admit it, mistakes and even misdiagnoses creep into the record, which can lead to confusion for the care provider. While mistakes always have been present in medical records, the “cut and paste” aspects of the EMR can actually increase errors. But the most im-

* Ellen Wright Clayton, M.D., J.D., is the Rosalind E. Franklin Professor of Genetics and Health Policy, Professor of Pediatrics, Professor of Law and Co-Director of the Center for Biomedical Ethics and Society at Vanderbilt University. 
1. Some of the entries are scanned handwritten clinic notes.
2. This is particularly true for the history and physical examination, which necessarily depend on the personal recollection and observation of the person.
important aspect of this technology for the present discussion is that its accessibility is its greatest liability as well as its greatest strength.

The question about the privacy of medical information can be stated simply: To what extent can and should patients control what the medical record contains and who has access to it and for what purposes? Patients often have apparently conflicting views on this subject. On the one hand, we, as patients, say that we prize privacy and that we fear that information will be used to harm us. On the other hand, we value the benefits that come from improved communication among providers, such as having our visits covered by third party payers and advances in medical science, which often come from researchers and providers having access to large data sets. These apparently contradictory desires are not unique to health care. In many other arenas of life, we sell information about ourselves incredibly cheaply, as the grocery and drug store discount tags on our key chains attest. But areas of discomfort remain, one of which I will explore: the use of the contents of the electronic record for medical research. Evolving practices in this area raise a host of questions.

The EMR has the potential to be a godsend for research and for minimizing the risk that others would have access to identified health information. In the past, information had to be abstracted by hand, a labor-intensive process requiring an investigator to examine each page, necessarily compromising the patient’s privacy. Although the initial invasion could not be avoided, researchers could try to mask identifying information prior to photocopying or record subsets of non-specified data on separate forms. Given the burdensomeness of this process, it was often tempting to capture as much information as possible on the first pass, particularly because there often was no opportunity or desire to undertake the onerous task of examining records again at a later time to update the research files. Sometimes, patients were asked to consent to this type of research, but at other times, institutional review boards (IRBs) considered such projects to be exempt from the regulations for the protection of human subjects so long as identifiers were removed from the final data set.

With the EMR, it is possible to limit one’s search to specified fields. The resulting data sets can be encrypted to protect privacy even further. who initially puts them in the medical record. These can then be copied by other providers and adopted as their own. See Interview with Daniel Masys, Professor and Chair, Dep’t of Biomedical Informatics, Vanderbilt Univ. Med. Ctr., in Nashville, Tenn. (Oct. 18, 2005). On the other hand, the ability to copy laboratory values and radiology reports directly into progress notes clearly improves accuracy.

3. Confidentiality refers to the obligations of those who hold information to respect the privacy interests of those which the data relate. Security refers to safeguards or tools, both technical and administrative, to protect health data from access or disclosure. See James G. Hodge, Jr. et al., Legal Issues Concerning Electronic Health Information, 282 JAMA 1466, 1466-68 (1999).

It is typically easy, for example, to correlate hypertension with levels of serum creatinine in a large population, and this can be done without having a human eye look at the entire record or even at a particular patient's name. And now that we are in the genomics era, investigators would like to combine this clinical data with studies of DNA to create so-called "biobanks" to examine how genetic variation contributes to disease.

In this Essay, I will demonstrate that current regulations in the United States for the protection of human subjects and of informational privacy permit research using the EMR and biological specimens collected in the course of clinical care to be conducted without consent of the patients to whom they relate, or even IRB without review. Despite this regulatory gap, many institutions are developing elaborate systems for oversight and, in many cases, obtaining consent or at least providing greater transparency, a phenomenon I will explore in the conclusion.

The two bodies of law most relevant here are the regulations for the protection of human subjects, usually referred to as the "Common Rule," and the privacy regulations promulgated under the Health Insurance Portability and Accountability Act (HIPAA). The Common Rule exempts from its coverage the study of existing information and samples "recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects." It then goes on to define "human subject" as:

[A] living individual about whom an investigator (whether professional or student) conducting research obtains . . . (2) Identifiable private information. . . . Private information includes 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). Private information must be individually identifiable

5. DNA can be obtained from virtually any biological sample. Laboratory tests for adults typically call for between five to ten milliliters of blood. Modern technologies usually require only a small portion of that amount. The white cells contained in the remaining sample are rich sources of DNA. Modern methods of DNA amplification make it possible to obtain DNA even from urine, which in the absence of acute infection or other disease, contain very few cells at all.

6. It is important to distinguish the use of information and samples explicitly collected for purposes of research. The big questions in this context are whether investigators can use these data for other types of research and whether research participants can give valid consent for future unspecified research uses. These are vexed issues in their own right, but at least the research participants are aware that they are involved in some sort of research, which often is not true for patients seeking clinical care.

8. Id. §§ 160, 164.
9. Id. § 46.101(b)(4).
(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.\(^\text{10}\)

In 2004, the Office of Human Research Protections (OHRP) discussed this definition in a guidance document which stated that:

OHRP does not consider research involving only coded private information or specimens to involve human subjects as defined under 45 CFR 46.102(f) if the following conditions are both met:

1. 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(s) cannot readily ascertain the identity of the individual(s) to whom the coded private information or specimens pertain because, for example:
   
   (a) the key to decipher the code is destroyed before the research begins;

   (b) the investigators and the holder of the key enter into an agreement prohibiting the release of the key to the investigators under any circumstances, until the individuals are deceased (note that the HHS regulations do not require the IRB to review and approve this agreement);

   (c) there are IRB-approved written policies and operating procedures for a repository or data management center that prohibit the release of the key to the investigators under any circumstances, until the individuals are deceased; or

   (d) there are other legal requirements prohibiting the release of the key to the investigators, until the individuals are deceased.

This guidance applies to existing private information and specimens, as well as to private information and specimens to be collected in the future for purposes other than the currently proposed research. The following are examples of private information or specimens that will be collected in the future for purposes other than the currently proposed research: (1) medical records, and (2) ongoing collection of specimens for a tissue repository.\(^\text{11}\)

\(^{10}\) Id. § 46.102(f).

\(^{11}\) Office for Human Research Protections, Dep’t of Health & Human Servs., Guidance on Research Involving Coded Private Information or Biological
This makes clear that, so long as they are coded appropriately, all information and tissue collected for purposes of clinical care can be used for research without being considered to involve human subjects and hence would not be subject to the Common Rule.

The HIPAA privacy regulations, or the "Privacy Rule," take a different approach focusing on de-identification, which can be accomplished either by statistically approved methods or by removing eighteen specific identifiers. Notably, categories such as race, ethnicity and socioeconomic status


12. Information is de-identified if:
   (1) A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: . . .
   (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; . . .
   (2) (i) The following identifiers . . . are removed:
   (A) Names;
   (B) All geographic subdivisions smaller than a State, including street address, city, county, precinct, zip code, and their equivalent geocodes, except for the initial three digits of a zip code if, according to the current publicly available data from the Bureau of the Census:
      (I) The geographic unit formed by combining all zip codes with the same three initial digits contains more than 20,000 people; and
      (2) The initial three digits of a zip code for all such geographic units containing 20,000 or fewer people is changed to 000.
   (C) All elements of dates (except year) for dates directly related to an individual, including birth date, admission date, discharge date, date of death; and all ages over 89 and all elements of dates (including year) indicative of such age, except that such ages and elements may be aggregated into a single category of age 90 or older;
   (D) Telephone numbers;
   (E) Fax numbers;
   (F) Electronic mail addresses;
   (G) Social security numbers;
   (H) Medical record numbers;
   (I) Health plan beneficiary numbers;
   (J) Account numbers;
   (K) Certificate/license numbers;
   (L) Vehicle identifiers and serial numbers, including license plate numbers;
   (M) Device identifiers and serial numbers;
   (N) Web Universal Resource Locators (URLs);
   (O) Internet Protocol (IP) address numbers;
   (P) Biometric identifiers, including finger and voice prints;
   (Q) Full face photographic images and any comparable images; and
   (R) Any other unique identifying number, characteristic, or code . . . ; and
   (ii) The covered entity does not have actual knowledge that the information could be used alone or in combination with other information to identify an individual who is a subject of the information.

45 C.F.R. § 164.514(b).
are not among the identifiers that must be removed to comply with HIPAA standards. Coded information can be considered to be de-identified for purposes of HIPAA so long as the investigator does not have the key and so long as the code is not itself based on personal health information (PHI). In addition, the Privacy Rule expressly excludes three types of investigations from its coverage: studies done in preparation for research, so long as the PHI is not removed from the institution and is necessary to develop the protocol; studies involving decedents, so long as living individuals will not be investigated; and research involving "limited data sets," defined as those from which sixteen of the eighteen identifiers have been removed, so long as the investigators execute a data use agreement.

Although the provisions of these two regulatory schemes do not completely coincide, there nonetheless exists a large space in which it would be possible to create a biobank of clinical information and DNA sequences from leftover biological specimens that could support a wide array of research of epidemiologic and genetic variation research but would not be covered by either the Common Rule or the HIPAA Privacy Rule. Within this gap, neither the informed consent requirements of the former nor the authorization requirements of the latter would apply. No particular form of oversight would be required either. OHRP recognized that protocols involving such datasets would not be subject to IRB review under their guidance. They recommended instead that institutions identify someone "to determine whether research involving coded private information or specimens constitutes human subjects research" and that such a determination not be left solely to investigators themselves.

Because our laws permit the research use of these large aggregations of clinical and genetic information without oversight or consent, should we care? After identifiers are removed, do any privacy interests remain to be invaded? If the answer is yes, has the polity chosen to override those interests for the public good? After all, laws reflect our collective will.

And yet, reasons for disquiet remain. While the Common Rule has for years permitted a great deal of epidemiologic research to be conducted without individual consent, the power of current approaches vastly exceeds what was possible when that rule was initially adopted, now almost a quarter century ago. Some people find it disturbing that the institutions from which they seek health care are using information about them or

13. Id. § 164.512(i).
14. Id. § 164.514(e)(1).
15. OHRP GUIDANCE DOCUMENT, supra note 11, at 4.
16. The literature on this issue is quite complex. A number of investigators have explored patients' opinions about the use of clinical information for research. In general, all have found that a substantial percentage of patients have concerns, particularly about research conducted without specific consent. See Richard Baker et al., What Proportion of Patients Refuse Consent to Data Collection from Their Records for Research Purposes?, 50 BRIT. J. GEN. PRAC. 655, 655-56 (2000); Kalpana Nair et al., Patients' Consent Preferences Regarding the Use of Their Health Information for Research Purposes: A Qualitative Study, 9 J. HEALTH SERVS. RES. & POL'y 22,
residual biological samples\textsuperscript{17} not only to help them but also to conduct

\textsuperscript{17} Spurred by recent developments in genomics that make it possible to obtain information from residual biological specimens, public opinion surveys regarding the use of tissue samples for research have been conducted in numerous parts of the world. Not surprisingly, they show quite disparate results. People in Sweden, for example, are quite favorably disposed toward research using linked tissue samples. Klaus Hoeyer et al., \textit{Informed Consent and Biobanks: A Population-Based Study of Attitudes Towards Tissue Donation for Genetic Research}, \textit{32 Scandinavian J. Pub. Health} 224, 224-29 (2004); Åsa Kettis-Lindblad et al., \textit{Genetic Research and Donation of Tissue Samples to Biobanks. What Do Potential Sample Donors in the Swedish General Public Think?}, \textit{Eur. J. Pub. Health} (forthcoming 2006). Although a majority thought that specific consent should be required to examine medical records, a similar number was willing to defer to permission given by an ethics committee. Despite this widespread public approval, Sweden nonetheless passed a law requiring that patients provide informed consent before samples can be used for research. See Karolinska Institutet, Biobank, Frequently Asked Questions, http://www.meb.ki.se/biobank/faq.php (last visited Feb. 9, 2006).


By contrast, people who were already enrolled in studies were often quite likely to agree to have their tissues and information used for future research. Donna T. Chen et al., \textit{Research with Stored Biological Samples: What Do Research Participants Want?}, \textit{165 Archives Internal Med.} 652, 652-55 (2005); Thomas Malone et al., \textit{High Rate of Consent to Bank Biologic Samples for Future Research: The Eastern Cooperative Oncology Group Experience}, \textit{94 J. Nat'l Cancer Inst.} 769 (2002); Geraldine M. McQuillan et al., \textit{Consent for Genetic Research in a General Population: The NHANES Experience}, \textit{5 Genetics Med.} 35 (2003); Birgitta Stegmayr & Kjell Asplund, \textit{Informed
research, especially without their consent. Many people worry about seeking health care for fear that they will be used as "guinea pigs." To be sure, most of these fears reflect concerns that they will be subjected to experimental medications or surgeries, or that they will be cared for by learners in the context of academic medical centers, but the fine points of biobanks may fail to persuade those who already distrust the health care system. More to the point, these regulations identify a large hole in the legally protected control individuals have over information regarding their health and in the oversight of research.

Interestingly, even though the law does not always require it, many institutions are seeking informed consent and putting in place mechanisms for oversight of biobanks. Some IRBs are reviewing biobanks at their own institutions and the ways they are used even when those projects are structured to be technically exempt under current OHRP guidance and HIPAA regulations. Even though some commentators urge that seeking consent creates bias or is unnecessary, many institutions still obtain consent. The Marshfield Clinic, for example, engaged in an elaborate process of public education and consultation and sought informed con-


19. It is more difficult to ascertain how many institutions are taking advantage of the haven created by the Office of Human Research Protections (OHRP) guidance and Health Insurance Portability and Accountability Act (HIPAA) exclusions.

20. See, e.g., Mary Terrell White & Jennifer Gamm, Informed Consent for Research on Stored Blood and Tissue Samples: A Survey of Institutional Review Board Practices, 9 ACCOUNTABILITY RES. 1, 1-16 (2002) (noting that practices were quite variable). The IRB at Vanderbilt University Medical Center has exercised a great deal of oversight over the DNA databank that is being developed.


sent from each of its patients within the two catchment areas covered by its efforts before launching its personalized medicine project. Following the example of the International HapMap Project and the policies of the Coriell Repositories, some institutions are establishing community advisory boards as well, which may serve simply as conduits of information about the projects to the broader population or may play a more active role in defining research questions and processes.

So why are some health care institutions not taking advantage of the OHRP Guidance and HIPAA provisions to conduct research on clinical records and samples without patient consent or IRB or public oversight? One can only speculate because health care institutions rarely reveal their innermost thoughts, but several hypotheses seem plausible. One is that hospital legal counsel tends to be conservative, for good reason. This is a hot topic. OHRP only recently clarified its position, after many years of declining to provide guidance, and the HIPAA privacy regulations are an even more recent development. During the same period, numerous commentators, including scholars and patient advocates, argued that informed consent should be required before clinical information and biological specimens are used for research.

Good lawyers, of course, can...
look past commentary to see what the regulators are actually saying at present, which is that there are no legal requirements in this area. And they could take comfort in the knowledge that while someone could bring a legal challenge to OHRP's interpretation of the regulations, overcoming the deference given to the government's position would be a heavy burden.

Ultimately, I suspect that the most important reason that health care institutions are engaging in numerous forms of public involvement when creating biobanks is not lack of awareness of the scope of the legal safe haven, but rather the recognition that non-disclosure and lack of oversight are risky in terms of public perception. It is hardly as if the American public is universally enthusiastic about its hospitals and clinicians. The health care system is under fire from a number of directions, from concerns that too many people are falling through the cracks or being bankrupted to allegations of fiscal mismanagement, fraud and poor quality of care. As part of this sea of concerns, health care institutions know that at least some patients are concerned about how information about them and specimens from them are used. They are all too aware that just one person complaining to the media can create a fire storm. These entities understand that what is needed is the public's trust and that this can be achieved only through transparency and engagement with individuals and communities. Ultimately, the silent creation of biobanks from clinical information and specimens turns out simply to be a bad idea, no matter what the law says.

---

e.g., J. Cassell & A. Young, Why We Should Not Seek Individual Informed Consent for Participation in Health Services Research, 28 J. MED. ETHICS 313 (2002) (arguing that informed consent requirements undermines mission of organizations conducting research in interests of entire population); cf. Klaus Hoeyer et al., The Ethics of Research Using Biobanks: Reason to Question the Importance Attributed to Informed Consent, 165 ARCHIVES INTERNAL MED. 97, 97-100 (2005) (suggesting that informed consent is inadequate measure of public acceptance of biobank-based research).

29. See supra notes 16-17 and accompanying text.