Confidentiality:

More than a linkage file and a locked drawer

BY MICHELE M. EASTER, ARLENE M. DAVIS, AND GAIL E. HENDERSON

A Promise of Confidentiality

Somewhere near the end of a consent form, script, recruitment letter, or brochure, researchers usually state that they will keep information "confidential." But how is this promise evaluated by the Institutional Review Board (IRB) and kept by the researchers? Given the researchers' confidentiality plan, the IRB must assess the risks related to participation, including risks created by disclosure of any information shared by subjects, and determine whether the study will be able to provide the level of confidentiality promised to subjects. In clinical research, confidentiality often concerns access to medical records or inadvertent revelation of patient status or of diagnosis. In comparison, survey research may raise fewer flags. When investigators offer confidentiality, it has frequently seemed sufficient to note that the study team will provide a locked cabinet, coded files, or confidentiality statements.

In reality, these safeguards may be inadequate or difficult to implement. Indeed, unique and important confidentiality concerns arise during the different stages of a survey research project, but are often under-appreciated in comparison to those of clinical research. According to a recent National Research Council Panel on Institutional Review Boards, Surveys, and Social Science Research, "It is likely that those involved in the human participant protection system… are paying too little attention to the ways in which technological and other changes in the research environment are increasing the risk of disclosure of the identity of participants in research."[1] What attention there is, however, has focused on statistical and administrative techniques to protect confidentiality of large-scale public-use micro data in secondary data analyses.[2] This paper examines how researchers keep their promise of confidentiality and how that promise is sometimes challenged during primary data collection and analysis for an interview study.

Studying Gene Transfer Researchers and Their Subjects

Our study of informed consent in gene transfer research (sometimes called "gene therapy") surveyed 144 investigators, study coordinators, patient-subjects, and IRB representatives involved in recent gene transfer clinical trials in order to determine how the prospect of direct benefit was understood and discussed by them. We conducted telephone interviews and taped, transcribed, and entered the data using both quantitative and qualitative analysis software. Our study team included a diverse group of scholars from law, philosophy, medicine, and sociology. Some had never before conducted survey research, but most had served on IRBs. Thus, the details of how to carry out the promise of confidentiality in our study were particularly interesting to us.

Researchers promise confidentiality because it is both an instrumental and an intrinsic value. As an instrumental value, the purpose is to protect subjects from the specific harms that might come from disclosure, e.g., loss of insurance or employment, or embarrassment. By contrast, confidentiality as an intrinsic value is a commitment to respect persons, a principle espoused by the Belmont Report,[3] regardless of possible physical, social or economic harms. Alternatively, to put it in the language of ethical theory, obligations of confidentiality can be seen as both consequentialist and deontological. In the latter sense, we have a duty to promise keeping (if we have promised confidentiality) regardless of whether a disclosure would cause measurable harms. Capron and others have described these non-measurable harms of disrespect as "dignitary harms," while others call them "wrongs."[4] Thus, our first duty as researchers was to honor the promise of confidentiality because we made it.

Regulations are often more concerned with protecting people from specific harms than upholding intrinsic values. While federal regulations for the protection of human subjects[5] state that IRBs shall determine whether there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data "when appropriate,"[6] they do not require them. Although often conflated under the

Michele M. Easter, Arlene M. Davis, and Gail E. Henderson "Confidentiality: More than a linkage file and a locked drawer," *IRB: Ethics & Human Research* 26, no. 2 (2004): 13-17.

rubric of "confidentiality," setting privacy and confidentiality out separately makes clear that they are important and discrete concepts: privacy, freedom from having private facts made public, and confidentiality, keeping shared information within a relationship. Thus, confidentiality was an additional promise we made to our subjects for its instrumental value: to protect them and minimize risks. Some studies could receive additional protection under a Certificate of Confidentiality,[7] the 1974 Privacy Act,[8] or the Privacy Rule of the 1996 Health Insurance Portability and Accountability Act (HIPAA).[9] Although our data collection predated HIPAA and did not rely on medical records, our confidentiality procedures would have conformed to HIPAA requirements.

In our study, as with most survey research, breach of confidentiality was the main risk to subjects. Recent and well-publicized events in the field of gene transfer, including the tragic death of Jesse Gelsinger and federal investigation of several research institutions,[10] made some of our subjects, especially those subjects who were gene transfer investigators, feel vulnerable to harm from potential confidentiality breaches. In contrast, several patient-subjects in our study wrongly assumed that we had access to their medical records, and thus appeared to trust us as they might a health care provider. Despite our explanations to the contrary, such subjects might believe that our research adhered to the longstanding, systematic, and more familiar standard of confidentiality in medical care, and thus possibly underestimate their risk in participation. In response to both these circumstances, we sought to offer a high standard of confidentiality—of their participation and the information they shared—to our subjects through procedures that addressed several aspects of confidential data: the sensitivity of different kinds of information, the length of time we would hold it, the usefulness of certain kinds of information, and our ability to protect information.[11]

In the end, we did what many studies would do to protect useful and sensitive information over time: we created careful recruitment letters and telephone scripts that avoided disclosure of subject participation in our study; we redacted interview transcripts; we identified interview data only by study numbers; we separated data from identifiers such as names, address, institution, or telephone numbers; and we kept all data and contact information in secure, locked files or computers. Yet, it soon became apparent to us that this seemingly thoughtful and thorough protection plan was insufficient to address all possible breaches.

Protecting Mr. "X" and His Interview

To illustrate the challenge of confidentiality, the following section traces one fictional patient-subject through our study, noting the various threats to his confidentiality and what we did to protect it. Procedural details are included to illustrate potential problems that IRB members or researchers may not anticipate.

Starting with a publicly available list from the National Institutes of Health (NIH), we contacted researchers conducting gene transfer trials. Doing our best to bypass "gatekeeper" secretaries and keep participation confidential, we recruited Dr. Jones, PI for the controversial "RareGen" study, the only gene transfer study in the world on this rare genetic disease. With the information he provided and using similar discretion, we contacted his study coordinator, Ms. Green. Both agreed to be interviewed and to forward our recruitment materials to the four patient-subjects enrolled in their phase I trial. We prepared and mailed four subject letters to Ms. Green; she addressed and mailed them to the subjects. The subject letters promised confidential participation and encouraged subjects not to reveal their participation to others as that would reveal the participation of Dr. Jones and Ms. Green.

Upon receiving the letter, Mr. X, a subject in the RareGen study, decided he was interested in our study and sent the enclosed contact information sheet to us. When we received it, we assigned him an ID number that linked him to the RareGen study in our sample. This contact information sheet was the **first linkage file** that connected our study ID number with information that could identify him as a subject: his name, address, and telephone number. We faxed this contact information sheet to the telephone calling center. This facsimile created a **second** copy of the contact information, and a **third** copy was created when the name, telephone number, and address were entered into the interviewing program for use by telephone interviewers. All three versions are linkage files, because they connect the ID number to identifying information about the individual, but each one is necessary for data collection to take place.

The interviewer called Mr. X to arrange an interview. (If necessary, the interviewer would leave a scripted message that avoided revealing Mr. X's interest in participating in our interview study and his participation in the RareGen study.) During the interview, the conversation was recorded and his answers were entered into a computer-assisted telephone interviewing (C.A.T.I.) system that enabled immediate entry of the quantitative data. We had expected most of these data to be free of potentially identifying information because answers were represented by numeric values and identified by a study ID, not his name. However, when asked to "specify" answers to

closed-ended questions, Mr. X might provide identifying details, which would become part of the C.A.T.I. dataset and copied later into yet another dataset (Stata) for quantitative analysis. Passwords were used to protect these computerized data sets and reports of data were monitored for identifying information. The cassette tape of the entire interview was labeled with his ID number and locked in a cabinet, creating a **fourth** link between ID number and identifiers, as the taped conversation was likely to include his name, those of the study team and institution, and even a confirmation of the address to which payment should be sent.

Mr. X had recently moved, so a new address was entered as data in the C.A.T.I. program. The outdated information had already been included in the interviewer's script, but the new address was now part of a data file, making it a **fifth** link between ID number and identifying information. The calling center telephoned this new address to the research staff so that the staff could mail a payment to him. (Information on telephone and fax can be controlled more easily than on email because even password-protected email may remain on the server and be resurrected later.) A check was sent without entering his name in the check registry and when it had cleared, Mr. X's name as payee and as endorser was redacted to protect his identity; this information was deleted from the data file as soon as possible.

During transcription of the taped interview, the transcriber removed obvious identifiers, such as Mr. X's name, and those of the researchers and the institution. However, the redacted transcript still contained information that could identify participants and studies. In his interview, Mr. X (like all subjects) discussed RareGen and the unique vector used to transfer the genetic material. The disease, gene, and vector information contained in the transcript could potentially identify the study and site. Even if these had been removed, symptoms and procedures could identify a specific disease to a knowledgeable reader. Although these details would not be published, they were needed to code and analyze subject responses. Thus, the initially redacted transcript was protected as carefully as the contact information: by using locked cabinets and password-protected computers and prohibiting electronic transmission of the data. It can be seen as a **sixth** linking file, though it would identify only singular studies, and only to people familiar with the range of gene transfer research.

Transcription and revision necessarily produce multiple copies of each confidential interview and each copy must be protected. At the calling center, Mr. X's transcript was stored not only on a password-protected hard drive but in a locked cabinet as printed hard copy and floppy disk (the latter two for transfer to the main study site). The study team reviewed, edited, and further redacted the transcript, and the final approved version was saved on the password-protected hard drive at the main study site, as well as locked in a cabinet in floppy disk, and printed form. Three back-up versions of the transcript were also created and protected on- and off-site, to protect against loss through fire or other disaster. Mr. X's transcript has now been reproduced in nine ways across four sites (calling center, main study office, and two off-site storage locations).

The transcriptionist reformatted Mr. X's interview for the qualitative software program (N6). This reformatted transcript, like the standard transcript, was reproduced and stored in several ways (hard drives at calling center and main suite, floppy disk, printed copy, and 3 back-ups). The transcript was imported into N6 with a security code because of information about the rare disease, and the entire qualitative database was backed up in 3 ways.

Now there are 11 versions of the reformatted transcript: 7 versions in Word and 4 versions in N6 datasets. Despite redaction, all versions require procedures for secure storage in each location and while in transit, and for destruction once a version is no longer needed. Investigators must read the interview text in N6 reports and apply qualitative codes, so portions of his interview are printed whenever needed. Because the transcript may name RareGen and the vector, the coders are required to keep the materials in their own locked files: up to 8 copies and 8 more locked cabinets for 8 investigators. After codes are applied, data are returned to research staff for shredding or temporary storage in a designated locked cabinet.

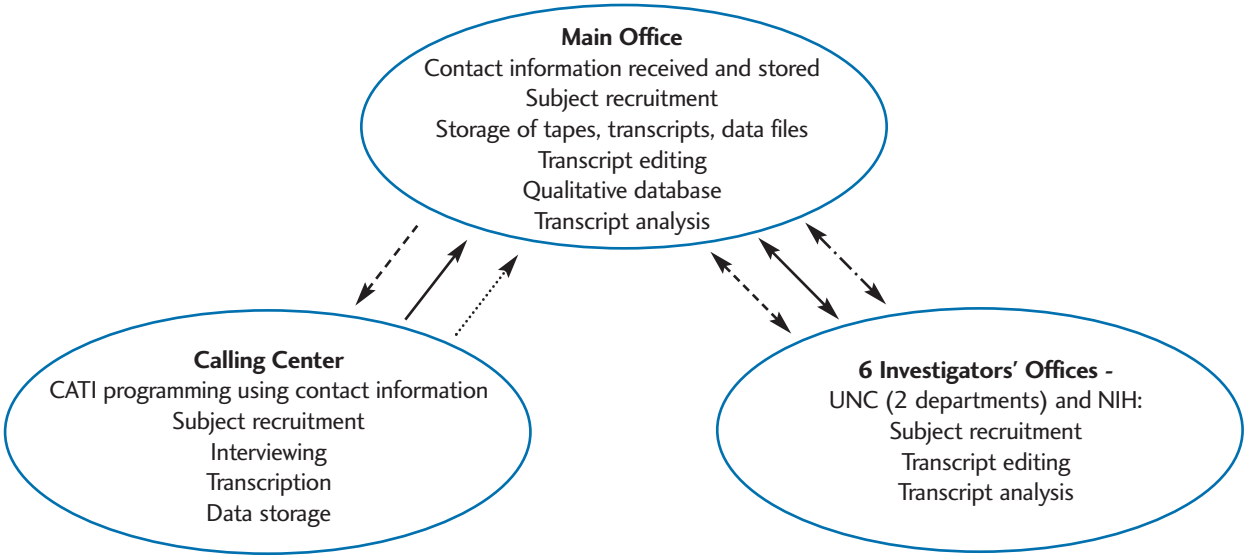
Rote Confidentiality: One Size Does Not Fit All

Our project's IRB application had described our confidentiality plan in fairly broad and standard language:

> Only authorized, study-related personnel will have access to any data, whether computerized or in hard copy. Hard copy and back-up disks will be stored in locked file cabinets. All interview data will be identified only by study numbers. Identifiers will be retained in a separate linkage file only for purposes of re-contact and subject payment, and the linkage file will be destroyed within six months after completion of data collection activities... [A]ll project files will be secured physically and electronically, and all personnel associated with the project will pledge confidentiality.[12]

Many of our confidentiality procedures were invented as the study progressed. We did not expect fulfillment of these promises to require as much attention and detail as it did. The surprising proliferation of de

Figure 1. Transfers of Confidential Information, Activities at Each Location



facto "linkage files" necessary for data collection required three locked cabinets, three password-protected computers, and procedures for confidential transfer between sites and for back-up storage.[13] As an interview went from tape to transcription to redaction to qualitative software and coding, it could exist in more than twenty manifestations constantly under control until most versions of transcript and linkage file could finally be destroyed. Figure 1 shows the various activities performed at each site along with the types of information that were transferred between sites.

Because some of the challenges described above were unique to our study, IRBs might not often encounter them. First, our cluster design necessitated disclosure of researcher participation to their patient-subjects, and exposed researchers to the possibility that their subjects might tell others. Second, the rarity of the diseases and technology studied in gene transfer research meant that redacted transcripts still contained sensitive information that could be identifiable and required careful security. Third, the public scrutiny of gene transfer research caused researchers to perceive an elevated risk of harm to themselves or their institutions and to ask for reassurance of their confidential study participation.

Other challenges are not unique to our study. The use of a separate calling center to conduct interviews is common in survey research. This arrangement means that identifying information and linkage files must be at more than one location, and therefore protected to the same degree as information at the main study site. Although our subcontractor was accustomed to conducting research in a confidential fashion, it was necessary to tailor a new confidentiality plan to meet our special requirements for the transfer of contact information and data; one procedure does not fit all studies. Finally, any research study that produces transcribed interviews will need to protect both identifying and identifiable information in the contact information, interview tapes and the many versions of transcripts created over time and at different sites, as well as the back-up versions of the transcripts and software programs. Although the sensitivity of the information often drives these protections, other factors may be important, such as the need to hold some useful information for a specific duration in order to re-contact researchers or pay patient-subjects and then account for that payment.

Suggestions For Researchers and IRBs

Devising study-specific protections for the confidential participation of our subjects and the information they shared with us required time and diligence that we did not initially anticipate, despite our knowledge as researchers and IRB members. We needed to look comprehensively at our confidentiality procedures on a regular basis and readjust them as we moved through the different stages of our research project.

Based on our experiences, we offer the following suggestions for researchers and IRBs. Studies that do not focus on medical research, collect sensitive data, or claim to protect the privacy of subjects or maintain the confidentiality of their information may not require the detailed procedures described in this paper.

The following suggestions apply to studies that do, and address some aspects of confidentiality that may be overlooked.

- Review the confidentiality promise of the consent process and assess its components in terms of the sensitivity of the information, the length of time the information is held, the usefulness of the information collected, and the ability to protect the information.
- Review the procedures related to protection of information across multiple sites and during transfer. Identify and limit the number of people who have access to the information, locations where information may be stored, and transfers across locations. When transfers are necessary, remember that the transferred information is often reproduced in another format, such as a facsimile, photocopy, a qualitative data report, or fields entered into a computer program. Make provisions for duplication and destruction when transferring information.
- Discuss agreements with calling centers and other subcontractors and their impact on the collection and storage of data. Address confidentiality issues with subcontractors from the beginning. Provide subcontractor and staff with written procedures so everyone knows what is expected.
- Include confidentiality issues in the continuing IRB review of relevant studies, enabling both researchers and IRB members to reexamine the sensitivity of collected information and the success of protection efforts, and to evaluate any need for modification of the confidentiality plan. If significant concerns arise about the collected information or the researchers' ability to protect it, the IRB may elect to review the project more frequently, creating shorter periods of continuing review, at intervals of less than twelve months.
- Provide for the education of study staff, researchers, and IRB members regarding data management and protection procedures.

Acknowledgments

The procedures described in this paper were part of a study supported by Nat-ional Human Genome Research Institute grant R01-HG02087, 1999-2003, "Social Construction of Benefit in Gene Transfer Research" (Gail E. Henderson and Nancy M.P. King, co-PIs). We are grateful to Larry R. Churchill, Daniel K. Nelson, Benjamin S. Wilfond, and especially Nancy M.P. King for their contributions to this paper.

■ **Michele M. Easter** is a Social Research Associate in the Department of Social Medicine, School of Medicine, University of North Carolina at Chapel Hill; **Arlene M. Davis** is an Assistant Professor in the Department of Social Medicine, School of Medicine, University of North Carolina at Chapel Hill and **Gail E. Henderson** is a Professor in the Department of Social Medicine, School of Medicine, University of North Carolina at Chapel Hill.

References

1. Constance FC, Ilgen DR, and Marrett CB, eds. *Protecting Participants and Facilitating Social and Behavioral Sciences Research*. Washington, DC: The National Academies Press, 2003;13.

2. Duncan GT, Jabine TB, and de Wolf VA, eds. *Private Lives and Public Policies*. Washington, DC: National Academies Press, 1993; Doyle P, Lane JI, Theeuwes JJM, Zayatzv LV, eds. *Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Amsterdam: Elsevier, 2001; Mackie C, Bradburn N, eds. *Improving Access to and Confidentiality of Research Data*. Washington, DC: The National Academies Press, 2000.

3. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: U.S. Government Printing Office, 1979.

4. Capron AM. Protection of research subjects: Do special rules apply in epidemiology? *Law, Medicine & Health Care* 1991;19(3-4):184-90.

5. 45 CFR 46 (1991).

6. 46.111(a)(7).

7. 42 CFR Part 2a; 42 USC 241(d).

8. 5 USC 552A (1974).

9. 45 CFR Parts 160 and 164 (2002).

10. Smith L, Byers JF. Gene therapy in the post-Gelsinger era. *JONA's Healthcare Law, Ethics, and Regulation* 2002;4(4):104-110.

11. Dale Hammerschmidt, MD, Director of Education in Human Subjects Protection at the University of Minnesota, often discusses these aspects of confidentiality. Confirmed in email correspondence, September 19, 2002.

12. 2000 application to Office of Human Research Studies at the University of North Carolina School of Medicine. (National Human Genome Research Institute grant R01-HG02087, 1999-2003, "Social Construction of Benefit in Gene Transfer Research").

13. Interviews with IRB representatives would have gone through more locations, since they were conducted from investigators' offices, not the calling center.