Current Medicolegal and Confidentiality Issues in Large, Multicenter Research Programs

The convenience of fast computers and the Internet have encouraged large collaborative research efforts by allowing transfers of data from multiple sites to a single data repository; however, standards for managing data security are needed to protect the confidentiality of participants. Through Dartmouth Medical School, in 1996–1998, the authors conducted a medicolegal analysis of federal laws, state statutes, and institutional policies in eight states and three different types of health care settings, which are part of a breast cancer surveillance consortium contributing data electronically to a centralized data repository. They learned that a variety of state and federal laws are available to protect confidentiality of professional and lay research participants. The strongest protection available is the Federal Certificate of Confidentiality, which supersedes state statutory protection, has been tested in court, and extends protection from forced disclosure (in litigation) to health care providers as well as patients. This paper describes the careful planning necessary to ensure adequate legal protection and data security, which must include a comprehensive understanding of state and federal protections applicable to medical research. Researchers must also develop rules or guidelines to ensure appropriate collection, use, and sharing of data. Finally, systems for the storage of both paper and electronic records must be as secure as possible. Am J Epidemiol 2000;152:371–8.

Information from medical records has contributed to data amassed in large databases for years. Cancer registries have been operating in many states for decades. More recently, other groups have designed and developed national or regional registries for childhood immunizations (1), cardiovascular surgery (2), and mammography screening (3–7). Health Employer Data Information Set performance measures derived from the databases of health maintenance organizations provide another example of computerized databases that contain potentially sensitive information. These databases are commonly aggregated for quality improvement or quality assurance purposes (2, 3, 8), as well as for research (9). Although confidentiality and integrity of data have always been a concern in research and clinical settings, technological advances in data handling and the ability to share large data sets have made the process of protecting confidentiality more challenging. Potential harm to patients whose confidential medical information is disclosed ranges from invasion of privacy to potential exposure to exploitative marketing activities. This harm is widely recognized, and the legal mechanisms available to protect patients against such disclosure are fairly well understood. In contrast, potential harm to professionals (physicians, nurses, and other care providers), such as loss of economic security and vulnerability to litigation, is not as widely recognized or understood. What is clear, however, is that the overseers and users of confidential data must protect the interests of both patients and professional research participants.

The current literature on confidentiality lacks an outline of approaches to address relevant medicolegal issues for
large computerized databases to which professional providers either knowingly or unknowingly have submitted information. We outline the approaches investigators should take to address data security and confidentiality for all research participants. This analysis is based on work conducted in eight states by investigators from seven academic research institutions, one health maintenance organization, and one state public health department. We analyzed federal and state laws as well as institutional policies intended to protect data from forced disclosure or use in litigation. We summarize the application of federal and state laws; describe essential steps for appropriate data collection, storage, utilization, and sharing; and offer confidentiality and security guidelines for data transfers from member sites to a central data depository. Our intention is to provide a clear framework for locally developed systems to protect the confidentiality of all research participants and ensure the integrity of data involved in confidential and sensitive medical research. It is critical that researchers carefully balance data use for the good of the public with a respect for the privacy and anonymity of all individuals involved.

BACKGROUND

In 1994, the National Cancer Institute (NCI) funded a mechanism that would allow mammography registries to pool their data in one centralized database, in part to respond to a congressional mandate in the Mammography Quality Standards Act (10). The speed and efficiency this pooling allows enhances our understanding of the operation and conduct of breast cancer screening in the United States. The resulting collaborative, the National Cancer Institute Breast Cancer Surveillance Consortium, is described in detail elsewhere (11). Each Consortium member site had previously established a computerized registry that collects data from designated mammographic facilities within a specific geographic region. Each site sends its data electronically to a centralized Statistical Coordinating Center for pooled analyses. The data include confidential information on mammography patients, physicians’ radiologic interpretive reports, and follow-up of abnormalities. Each mammography registry is linked to the regional population-based Surveillance, Epidemiology, and End Results registry of the NCI or similar statewide cancer registries. The linked data enable the determination of predictive value, sensitivity, and specificity of mammography as well as practice patterns. While these determinations are critically important to evaluating the performance of mammography, they necessarily involve the sharing of sensitive data, with potential risks to participants.

Soon after the Consortium was formed, a working group of representatives from each site obtained copies of federal and state laws that create a privilege against disclosure in litigation and of institutional regulations that address confidentiality of data generally. Our analysis of these materials revealed remarkable variability in how states address confidentiality issues. On the basis of our findings, we outline a recommended approach that investigators participating in large, multisite research programs may take in applying a minimum set of standards for the protection of all research subjects and health care providers and the data they contribute. In presenting this information, we will address definitions of confidentiality, the responsibilities of member sites, state and federal protections, data access, and paper and computer data security.

DEFINITIONS OF CONFIDENTIAL INFORMATION

Confidential information is essentially personal information, that is, all information that links data to a specifically identified or identifiable research participant, professional or lay. Such identifying information may include physician (or patient) name, practice location name, address (including zip code), telephone number, occupation and employer, and, in some instances, rare diseases. Breach of confidentiality is the disclosure of health information without consent and without legal compulsion or legal authorization for its release (12). Table 1 outlines and defines categories of potentially sensitive information, ranked according to the severity of the potential repercussions of a breach of confidentiality.

RESPONSIBILITIES OF MEMBER SITES

The objective of each member site is to balance its research endeavors with its commitment to protect confidential information obtained and generated in the course of that research. The policy of each site should be to adhere to laws and regulations that govern the collection, compilation, use, transfer, and storage of confidential data; to protect this information from unauthorized access or use at all times; to ensure that this information will only be transferred, utilized, and stored in sanctioned and approved ways; to ensure that breaches of this policy are reported promptly and that appropriate corrective or disciplinary measures are taken; and to respond promptly to inquiries from concerned participants regarding research and other activities. The obligation to protect data from unauthorized access and release extends indefinitely, even after the patient or physician is deceased or the physician ceases practicing within the area.

Adherence to applicable laws and regulations necessarily requires familiarity with the types of protections offered by federal and state governments and institutions. Table 2 outlines these types of protections, each of which is discussed more expansively in the next section. Table 3 outlines the types of protection available for each of the eight Breast Cancer Surveillance Consortium member sites.

FEDERAL AND STATE LAWS AND REGULATIONS

The confidentiality of medical data is protected by laws and regulations at both state and federal levels. Collection, access, use, and disclosure of confidential data pertaining to study subjects entered at each member site are governed by federal and state statutes and regulations. In our medicolegal review, we focus on statutes and regulations protecting confidentiality. Although patients’ privacy rights are recognized in medical ethics, common law, and constitutional law, statutes and regulations are the primary source of protection for research subjects. These sources also define parameters for use of medical records in research. Moreover, other significant confidential-
ity protections for patients, such as physician-patient privilege, are exclusively statutory creations. Each site must comply with these laws to the fullest extent possible to meet its obligations to funding sources and to meet its commitment to ethical principles upon which human subjects regulations are predicated. While federal laws are applicable to any state, state statutes, if they exist, can vary from state to state. The strongest possible legal protection exists where there are laws to protect confidentiality of data both from use in litigation (e.g., discovery or admissibility as evidence) and from forced disclosure of identifying information.

**Federal Certificates of Confidentiality**

Federal Certificates of Confidentiality are issued in accordance with the provisions of section 301(d) of the Public Health Service Act, 42 U.S.C. section 241(d) to protect the privacy of research subjects by withholding their identities from all persons not connected with the research. Under Section 301(d), no federal, state, or local civil, criminal, administrative, legislative, or other proceedings can be used to compel disclosure of identifying characteristics of research subjects (13). This level of protection is the strongest and

### TABLE 1. Categories and definitions of confidential information

<table>
<thead>
<tr>
<th>Category of information</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>Information or data collected, compiled, utilized, or generated that is intended for public distribution and use or that may be obtained under freedom of information legislation. Generally, this includes aggregated data in published form, such as articles in medical journals about patterns of care, accuracy, and other related topics. This does not include confidential information.</td>
</tr>
<tr>
<td>Internal</td>
<td>Information or data collected, compiled, utilized, or generated by the organization that may be shared with employees and authorized consultants and contractors only. Authorization for external distribution or access must be obtained from the Principal Investigator. Examples of internal information include site lists, technical reports, and research proposals in stages of preparation.</td>
</tr>
<tr>
<td>Restricted</td>
<td>Confidential information collected, compiled, utilized, and/or stored by the organization that contains identifying links with specific individuals or medical practices, such as name, address, and Social Security number. Confidential registry data and reports fall within this category, as do any personal identifiers collected as part of a registry (including diagnoses that, when linked with geographic location, could identify an individual or number of patients served by a facility that could identify provider participants).</td>
</tr>
</tbody>
</table>

### TABLE 2. Types of protection offered by federal or state governments and individual institutions

<table>
<thead>
<tr>
<th>Type of protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Federal</td>
</tr>
<tr>
<td>Public health service certificate of confidentiality</td>
</tr>
<tr>
<td>IRB* requirements for protection of subjects from the risk of loss of confidentiality</td>
</tr>
<tr>
<td>State</td>
</tr>
<tr>
<td>Laws protecting the confidentiality of records used in medical research</td>
</tr>
<tr>
<td>Laws protecting cancer or mammography registries</td>
</tr>
<tr>
<td>Quality assurance or peer-review statutes</td>
</tr>
<tr>
<td>Laws regulating physician-patient privilege</td>
</tr>
<tr>
<td>Laws on Patient’s Bill of Rights</td>
</tr>
<tr>
<td>Laws governing confidentiality of patient’s medical records</td>
</tr>
<tr>
<td>Institutional</td>
</tr>
<tr>
<td>Data security</td>
</tr>
<tr>
<td>Limiting data access with key or password protection</td>
</tr>
<tr>
<td>Outlining the specifics of all data handling using a standardized protocol</td>
</tr>
<tr>
<td>Shredding unneeded paper data</td>
</tr>
<tr>
<td>Formalizing all data requests and establishing a review process for release of research data</td>
</tr>
<tr>
<td>Developing a firewall for all computer systems</td>
</tr>
<tr>
<td>Maintaining off-site backups of computerized databases</td>
</tr>
<tr>
<td>Using a specially designed encryption program to convert data before sending it over the Internet</td>
</tr>
</tbody>
</table>

* IRB, institutional review board.
most comprehensive currently offered by applicable law, and legal precedent exists to support the strength of this coverage (14). The protection extends not just to patients and other research subjects, but also to professional participants (physicians, nurses, technologists, and other health care providers) who contribute data to each member site.

A decision to obtain a Federal Certificate of Confidentiality should be based on the potential risk of loss of confidentiality and a legal analysis of the level of protection offered by state statutes, which, as mentioned, is quite variable. The coverage afforded by the Certificates provides an important layer of federal protection in addition to the variable protection offered at the state level and allows for protection of confidentiality of data crossing state lines, which is critical for sending data electronically (or otherwise) across state lines.

It is not necessary for research to be federally funded to be eligible for a Certificate of Confidentiality. However, Certificates are available only for research of a sensitive nature, such as research relating to sexual attitudes, preferences, or practices; use of alcohol, drugs, or other addictive products; illegal conduct; a situation that could, if released, be reasonably damaging to an individual’s financial standing, employability, or reputation within the community; matters that would normally be included in a patient’s record, disclosure of which could lead to social stigmatization or discrimination; or an individual’s psychologic well-being or mental health (13). Additionally, applicants for a Certificate of Confidentiality must show that they will be engaging in a systematic study “directed toward new or fuller knowledge and understanding of the subject studied” (13, p. 729). Institutional Review Board (IRB) approval is required before an application for the Certificate is submitted. To cover professional participants, evidence of their status as research subjects must be provided, and the consequences of a breach of confidentiality must be specifically outlined. Information about the Certificate and application requirements can be obtained from either the NCI or the National Institute of Mental Health.

State confidentiality laws

State laws protecting the confidentiality of records used in medical research can essentially be divided into five general categories: 1) laws specifically applicable to confidentiality of records used in medical research; 2) laws specifically applicable to cancer or other registries; 3) confidentiality requirements under quality-assurance or peer-review statutes; 4) laws creating a physician-patient privilege; and 5) laws generally applicable to the confidentiality of medical records. Protection afforded under all five types of legislation varies from state to state, although among most states, the first category consistently provides the most comprehensive protection for information collected for medical research. Considerations affecting coverage provided by each category of statute are briefly discussed below.

Category 1. Medical research statutes. Not all states have medical research statutes. In those that do, the adequacy of protection afforded depends upon several factors. We are unaware of any statute that specifically authorizes confidentiality protection for providers who are research subjects by virtue of reports or outcome data provided to the study, although in some states, statutory language may be expansively interpreted to provide that protection. Otherwise, the factors include whether confidentiality protection is needed for professional participants, whether the jurisdiction in which the research is conducted permits disclosure of information that identifies the participant as necessary to “further a study,” and how personally identifying information is defined. Some statutes also prohibit redisclosure of information, while others are silent on this subject.

Category 2. Registry statutes. Some states have created programs for reporting incidences of disease to state registries. For research conducted pursuant to a state-authorized registry program, fairly strong confidentiality protection may be afforded by the applicable statute. These statutes often authorize disclosure of information collected by the registry to researchers, and researchers who work with such information may be entitled to confidentiality protection by the statute. Obviously, however, such laws are useful only for protecting the confidentiality of data collected in connection with a statutorily referenced registry.

Category 3. Peer-review or quality-assurance (QA) statutes. QA statutes and the scope of protection they afford differ widely from state to state. Although many researchers assume that QA statutes provide solid confidentiality protection, in fact, they often apply only to data collected...
lected in very specific ways and for narrowly focused purposes. It may actually be possible to inadvertently waive the QA protection by using information collected for purposes that fall outside those authorized by the statute. Courts will likely find that QA statutes do not apply to protect the confidentiality of data if the following exist:

- Wrong class—The data are not within the class of information protected by the statute. Protected records typically include “records” generated by a quality-assurance, peer-review, or medical staff “committee.” Researchers may not qualify as such a committee, and their research may not constitute a record.
- Formalities not observed—In some states, the committees must be a “regularly constituted review committee” of a hospital medical staff.
- Use for improvement of patient care—Many QA statutes protect only data that are systematically used to evaluate and improve the quality of patient care.
- Statutory exemption—An exception may apply to permit disclosure of information otherwise protected. For example, QA records usually can be disclosed in suits brought to challenge the denial of medical staff privileges.
- Absence of internal controls—Persons seeking to demonstrate that information is a protected QA record must usually be able to demonstrate that internal controls exist to protect the confidentiality of the subject data.
- Waiver—Confidentiality protection for QA records will be waived if otherwise protected information is voluntarily transferred outside the hands of the statutorily designated QA committee or office.

In summary and contrary to common perception, peer review or QA statutes may not confer substantial protection from discovery (15). The value of QA statutes in protecting the confidentiality of research databases is highly dependent upon how information is handled, by whom it is handled, and whether a legal precedent exists.

Category 4. Physician-patient privilege laws. Most states, if not all, have laws that establish an evidentiary privilege for communications between a physician and a patient about the patient’s care. When the privilege applies, it prevents use of such communications in litigation. However, there are many exceptions to the privilege in most states. It is important to note that the privilege is generally said to “belong to the patient,” meaning that only the patient (and not the provider) can claim it. As a result, the patient is free to authorize disclosure of the otherwise protected information to whomever he or she chooses. Because waiver of the privilege for one purpose may be held to constitute a waiver for other purposes, it is possible for patients to unwittingly authorize much broader disclosure than intended. The privilege may also be subject to statutory exceptions. Many states provide that it is inapplicable in proceedings before professional conduct committees. In sum, the privilege does not afford any protection to professional subjects of research, and the protection it gives patients may be quite limited.

Category 5. Other laws generally applicable to the confidentiality of medical records. Many states have adopted a Patient’s Bill of Rights. These laws usually state that patients have the right to expect that communications and records pertaining to their care will be treated as confidential and not disclosed without their authorization. Privacy rights existing in the state and federal constitutions may also protect against disclosure of medical records in some instances. While these sources do not provide distinct protection for records collected by medical researchers, they may help bolster claims that medical information gathered by researchers is confidential. Because these laws change frequently, close surveillance is necessary by investigators who hope to access medical records for research purposes.

In addition to ensuring that the data are protected from legal discovery, researchers must be vigilant in protecting data from any use that might bring harm to the participants. This vigilance includes the establishment of both rules to prevent the misuse of data and systems to physically protect the data. These protections are discussed next.

POLICIES AND PROCEDURES FOR HANDLING DATA

The orientation and training of staff members and investigators at all sites who require access to confidential data to conduct their work should include instructions concerning the collection, maintenance, use, and release of confidential data. Developing a policy and procedures manual brings a basic level of uniformity to data handling and access. Each new staff member should be required to read the confidentiality policy and procedures manual and sign a pledge to uphold this policy. The pledges must remain in effect after cessation of employment, so sites should maintain a historical file of staff members who have signed them.

At member sites, investigators or public health officials may request access to confidential or aggregate data. All such persons given access to data should read the confidentiality policy and procedures manual and sign an agreement to adhere to the same confidentiality standards practiced by the site’s staff members.

Confidential data should not be transmitted from sites by any means (mail, telephone, electronic mail, or facsimile) without explicit authority from the Principal Investigator or a staff member to whom such authority has been delegated. The specific types of data, such as variables and date range, and those to whom they would be transmitted must be clearly communicated in writing to the staff. Because researchers often contract with computer programmers, biostatisticians, or contractors and consultants who have access to restricted information, these individuals should read the confidentiality policy and procedures manual and sign a confidentiality agreement with assurances that they will safeguard such information from unauthorized access or further disclosure. Confidential data should not be available to businesses or industries that desire to market a product or service to patients, health care providers or employees for advertising or recruitment of new patients, or insurance companies that are attempting to determine the status of individuals for any reason.
All external requests for data to be used in research should be approved by respective IRBs before submission of the request to the member site. All requests should be made in writing, preferably on a formal data request form, and should clearly state the limits of data use. Data may be used only for the exact purpose for which they are requested, must be kept confidential, and must remain in the custody of the fewest individuals possible. Applicants should specify the exact time period during which they will require access to data and should agree to provide a copy of any proposed publication or other form of public disclosure to member sites at least 30 days before release. This period will ensure adequate time to review, comment, or decide to reanalyze and provide a response or alternate explanation, if necessary.

All requests should be approved by the Principal Investigator or an advisory body, such as a steering committee made up of community physicians and members of the site’s research team. If an advisory committee is used, a description of how the committee members are chosen or elected, their length of term, and the procedures used to approve a request should be outlined, including voting criteria (majority, unanimous, quorum), time limits for responding to requests for approval, and notification and documentation requirements.

Requests requiring the use of personal identifiers should explain the necessary precautions to be taken to provide confidentiality in accordance with procedures approved by the project’s IRB, such as reporting patient, practitioner, and practice site data in sufficient aggregate to minimize the risk of identifying individuals or individual practices. When data analyses are complete, data should either be destroyed or, if needed for later reference, maintained in locked storage in the custody of an applicant for a specified period until they are no longer needed. If a central data repository is used for pooled analyses, this repository should abide by the same standards of confidentiality as all member sites. In addition, a review process for requests of pooled data should be developed.

DATA SECURITY

Paper systems

The following components can enhance data security in all areas of member site operation. Suitable locks should be installed to control access to the site, and all staff should be notified of the importance of maintaining a secure environment. A roster of persons authorized to enter the area should be maintained by the administrative personnel. Staff should be responsible for the confidentiality of all data encountered during data collection.

A site-developed mail-tracking system should be used to protect confidential data. The physical security of confidential data stored on paper documents, computer printouts, microfiche, and other media forms from member sites should be ensured. Confidential documents to be destroyed should be kept in a secure environment until they are shredded and disposed of properly.

If member sites produce QA reports for practitioners or other facilities at designated intervals, those receiving the reports should be informed about appropriate and inappropriate methods of handling them and should comply with applicable QA statutes. While legal protection from discovery is necessary to ensure that no harm comes to those contributing data to a database, the same individuals have an equal responsibility to protect the confidentiality of data they receive from member sites.

QA reports may contain identifying information about providers or patients. Any report that contains identifiable information must be treated as confidentially as any medical record. Encrypted codes may be generated when appropriate each time a report is created to protect the identity of a receiving practice location or radiologist. These codes should never link participant identifiers to actual study data. To provide extra protection when preparing report mailings, a two-step process may be used. Here, two individuals are responsible for report handling within a site, with one kept blind to the encrypted code and having access to the database for report production while the other, who applies the encrypted code for processing and ultimate mailing, is kept blind to the report content. Practitioner or patient data should be reported only in aggregate sufficient to minimize the risk of identifying individuals or individual practice groups. Thus, any cells that have a small number of cases (which may identify an individual or a practice location) should be suppressed in those reports. The purpose of the reports should be clearly printed on them or on accompanying information.

Computer systems

Computers should be located in a locked facility with no access to public traffic. Computer security safeguards are outlined below.

1. Participant identifiers and demographic information should be stored in files that contain no other information. Other data should be stored in separate computer files in the database. They should be linked by a scrambled code that can be accessed only by authorized personnel.

2. Password protection should be required for the computers, applications, and databases of each member site. All users accessing the database should have a unique identification code and password. Passwords should be changed on a regular basis. A user’s identification and password should be invalidated when the individual no longer requires access to the database. Precautions should be taken for both physical and electronic security of confidential data sent on magnetic or electronic media. Secure telephone data transmission should be accomplished by using an unlisted telephone number, password access to the bulletin board systems, and restricted use of facsimile technology for the transmission of confidential data.

3. Backup disks or tapes should have no identification on them other than numbers or codes and a generic office
Managing Confidentiality in Large Research Programs

The major risk to participants is disclosure of potentially sensitive information to federal program auditors and mortgage holders (20). National legislative activity has increased significantly in this area. On the national level, a comprehensive federal policy on confidentiality of medical records can be expected in the year 2000. The United States Congress has considered at least two recent legislative proposals that deal directly with attempts to ensure privacy of identifiable information, such as the medical record (20). Issues concerning informed consent, disclosure, and physical security, as well as who would be the oversight body, are under consideration.

It is important for the public to understand and recognize the difference between utilizing medical information for the good of the public, such as is done in medical research, and medical record misuse that occurs outside the protection of the federal and state regulations discussed in this paper. For research studies to gain the participation needed by the public, the confidentiality of research data must be honored and protected. Otherwise, it will be impossible to conduct research such as that being done by our Consortium. It is equally important for researchers who intend to collect data for research purposes to rely on current laws and to monitor pending legislation that may affect their ability to conduct their research. The strongest legal mechanism of protection that currently exists is the Federal Certificate of Confidentiality. Its strength lies in the geographic coverage...
it affords, the relative paucity of exceptions to its coverage, and the legal precedent that already exists regarding its use to protect the confidentiality of research subjects. Notwithstanding this valuable mechanism, researchers should be familiar with the specific confidentiality and privacy protections that may exist within their own jurisdictions and apply them when appropriate. To maximize protection, researchers should obtain a Certificate of Confidentiality; research legal precedents in their state and take advantage of the protection available; and institute measures to minimize the chance of unauthorized or inadvertent disclosure of confidential information in databases, data reports, and research information. Through these actions, researchers can fulfill their ethical and legal obligations by protecting confidential information to the maximum extent possible under existing law, while continuing their research.

REFERENCES