The Practice of Informatics JAMIA

White Paper ■

Multicenter Patient Records Research:

Security Policies and Tools

FRED M. BEHLEN, PHD, STEPHEN B. JOHNSON, PHD

Abstract The expanding health information infrastructure offers the promise of new medical knowledge drawn from patient records. Such promise will never be fulfilled, however, unless researchers first address policy issues regarding the rights and interests of both the patients and the institutions who hold their records. In this article, the authors analyze the interests of patients and institutions in light of public policy and institutional needs. They conclude that the multicenter study, with Institutional Review Board approval of each study at each site, protects the interests of both. ''Anonymity'' is no panacea, since patient records are so rich in information that they can never be truly anonymous. Researchers must earn and respect the trust of the public, as responsible stewards of facts about patients' lives. The authors find that computer security tools are needed to administer multicenter patient records studies and describe simple approaches that can be implemented using commercial database products.

■ **JAMIA.** 1999;6:435–443.

Medical research using patient records data requires careful design to meet the requirements of public and institutional policy as well as the requirements of the research. This article presents, first, an analysis of the policy issues of multi-institutional patient records research, and then discusses these issues in light of implementation experience in an inter-institutional ''virtual repository'' project.

The practice of examining existing patient records is not new,[1,2] but advanced information technology now makes much larger studies possible. Communication technology makes it possible to examine still larger and more diverse study populations in multi-institutional studies. Such studies are among the benefits envisioned for computer-based patient records.[3–5] This comes at a time of increasing demands to measure outcomes, for which record searches are among the most powerful tools.[6] Because of this scale, and public perceptions of the potential abuse of electronic patient records data, policy makers have expressed increased interest in this area, not all of it beneficial.[2]

Even as scrutiny increases, the policy awareness of many who conduct such research is varied,[7] and the medical and research communities would benefit from a clearer understanding of the issues. It is particularly important that informaticians and information officers understand how to apply the rules, since

Affiliations of the authors: The University of Chicago, Chicago, Illinois (FMB); Columbia University, New York, New York (SBJ).

This work was supported in part by TRP Technology Development Agreement MDA972-94-3-0047 from the Defense Advanced Research Projects Agency.

Correspondence and reprints: Fred M. Behlen, PhD, Department of Radiology, MC2026, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637.
e-mail: ⟨f-behlen@uchicago.edu⟩.

Received for publication: 5/4/99; accepted for publication: 7/20/99.

they are frequently faced with insistent requests for data from researchers who misunderstand policy and inappropriately extrapolate clinical privileges into the research domain.

Privacy and confidentiality of the patient record has attracted extensive debate and analysis,[8–13] including discussion of research.[14] Although policy issues regarding research access to public health databases have been analyzed in detail,[15] less attention has been paid to the problem of how to oversee and administer, within the framework of applicable public policy, multicenter research using privately held patient records. In addition to public policy, the policies of each participating institution must be considered.

Patient records are held by a health care institution, which has legally defined responsibilities for their use and management. The institution's patient records contain information not only about patients but also about physicians and the institution. While public policy and professional ethics set conditions and procedures for the use of patient data, the participation of provider institutions in research is strictly voluntary, and a research program must meet the criteria and serve the interests of each participating institution. Thus, a system of sharing patient records must recognize and protect the interests of both the patients and the contributing institutions.

Implementation of a multi-institution research database thus requires a set of institutional policies, agreements, and regulatory compliance mechanisms. A computer system architecture supporting these guidelines should make efficient implementation of policy possible and provide mechanisms to ensure compliance.

A system implementing a security architecture according to the principles discussed here was implemented in the Virtual Repository Project of the National Academic Medical Center Information Consortium (NAMCIC), a program supported by the Defense Advanced Research Projects Agency (DARPA). The purpose of the project is to lay the groundwork for large-scale cooperative use of patient records data in breast cancer and other research.

The next two sections address, in turn, the interests of patients and the interests of institutions. The remainder of the paper relates these discussions to tools and practicable administrative procedures and to the authors' experience in building the virtual repository. The conclusions comprise a set of guidelines for the design and administration of patient records research systems.

Interests of Patients

Patient records contain information provided by a patient and information recorded by the provider in the course of care. Patients expect that both sorts of information will be treated confidentially and used to benefit their care.

Federal Regulations

Public policy in the United States has recognized that, under certain conditions, patient records may be used in research, and has empowered Institutional Review Boards (IRBs) to protect the patients' interests in accordance with federal regulations set forth in the Federal Policy for the Protection of Human Subjects[16] and codified by the Department of Health and Human Services in the Code of Federal Regulations at 45 CFR 46.[17] Under these regulations, studies using patient records require one of the following:

- *Informed consent.* Informed consent is always preferred in research involving patient records. Federal code at 45 CFR 46.116(a)–(c) provides detailed requirements for informed consent of human subjects involved in research, although these consent procedures are clearly designed for more invasive or hazardous research. When large bodies of established patient records data are used, it would be impossible to obtain consent to the use of their records from every patient, but standard consents signed by patients at some medical centers may authorize some research use of patient records.
- *Waiver of informed consent.* If informed consent cannot practicably be obtained, an IRB may approve a protocol in which informed consent is waived. Federal code at 45 CFR 46.116(d) provides for waiver of consent when the IRB determines, among other things, that the research involves no more than minimal risk to the subjects, the waiver will not adversely affect the rights and welfare of the subjects, and the research could not practicably be carried out without the waiver.
- *Exempt research.* Exempt research requires neither informed consent nor waiver of consent. It is defined at 45 CFR 46.101(b)(4) as ''research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects.''

The exemption regulation follows the long-standing practices of the use of specimens and images for teaching and research purposes. An image or specimen is taken from an individual and is thereafter identified only by the medical properties it possesses. The applicability of exemption to patient medical records studies, however, is limited, because patient records contain many different facts and events that, taken together, can usually be used to identify an individual. Thus, only studies of very limited scope are candidates for exemption.

Anonymity and "Scrubbing" of Patient Data

Confusion of the legal and ethical principles and their application to patient records research is common, and further exploration of confidentiality, anonymity, and "scrubbing" of patient data is appropriate. A recurring idea is that a research database of patient data can and should be "scrubbed" of personal identifying information, and thereafter the "clean" database can be made available for research on a less restricted basis.

Much of the confusion seems to arise from inappropriate application of the requirements for exemption under 45 CFR 46.101(b)(4), which exempts studies of existing records "if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects." We argue that such complete scrubbing is not feasible, and even if it were feasible, it would not be appropriate ethically.

Setting aside for now the question of feasibility, a troublesome requirement for exemption is that of "throwing away the key" that links data to a patient. This requirement is found only in 45 CFR 46.101(b)(4), and it presents some practical, scientific, and ethical problems:

- It forecloses the possibility of benefit to the patient. In an investigation using patient records, facts may be discovered that require medical follow-up on behalf of the patient—because of an oversight in the patient's original care, for example, or the application of new knowledge to the patient's case. The principle of beneficence articulated in the Belmont Report[18] establishes an obligation to maximize possible benefits and minimize possible harms. Deliberately destroying the link to the patient eliminates all possible benefit to the patient and should be done only if the mitigation of some greater risk so justifies.
- The requirement greatly complicates the maintenance of a current database. Updating the research database with current information is made more difficult, although not impossible. (One-way cryptographic techniques[19,20] could be used, for example, but at the expense of poorer data quality resulting from the inability to correct errors.)
- The requirement eliminates some checks against scientific fraud. If data cannot possibly be traced back to authoritative sources, the auditing of questioned results becomes more difficult. This concern has humanitarian as well as administrative dimensions, since fraud in medical science may cause patients to receive inappropriate care.

These disadvantages might be tolerable if there were no other way to achieve significant research goals or if it were possible to create a valuable resource that could be opened to a much broader research community.

However, aside from the issues raised above, a database of unidentifiable patient records is infeasible. The extrapolation of the exemption rule from simple specimens to complex records ignores the strong identifying effect of linking even a small number of independent facts. Each fact identifies a subpopulation, and the linkage of facts defines the intersection of those sets. Each additional linked fact reduces the size of the intersection. When the intersection has only one element, an individual has been identified. This usually requires just a few facts. For example, the combination of birth data and residence ZIP code uniquely identifies one of the authors (F.M.B.) among the 315,000 patients whose records are in the database at The University of Chicago Hospitals.

The potential for such inferential identification is a well-recognized hazard of microdata, or information about individuals.[21,22] Methods developed to defeat identification of individuals in microdata include subsampling, aggregation, noise introduction, and substitution. Subsampling selects at random only a small fraction of the original data. It reduces the likelihood that any particular individual can be identified in the released data, and correlation of two subsampled databases is usually infeasible. Aggregation methods "blur" one or more of the data elements, such as the reporting of birth date only by year and of location only by region rather than ZIP code. Related to aggregation is the introduction of random noise into the data.[21] Subsampling and aggregation have been used by the Social Security Administration in their release of microdata files for public use.[23]

Substitution techniques[24] preserve much of the detail while reducing the potential for identification by replacing or permuting data elements with other values, equivalent with respect to the statistical outcome of the query, in such a way as to ensure that the number of individuals in the intersection of linked sets never falls below a certain number. "Outliers," or rare occurrences of combinations that cannot be aggregated with a sufficient number of equivalent individuals, must be dropped from the released data. The above methods have been used with success in the disclosure of statistical data.[25]

However, all these "de-identification" methods have limited applicability to patient records research. First, all methods presume some knowledge of what statistical measures need to be preserved in the de-identification process, which may be feasible for any given study but not for a general-purpose research database. Second, the very nature of much medical records research is the detection and study of outliers, rare occurrences correlated with factors that may be causal. Since medical records are not obtained under controlled experimental conditions, valid conclusions often cannot be drawn from small differences in gross statistics, so outliers can be important. Third, the interpretation of results and the assessment of statistical power are complicated by the need to include the effect of the de-identification procedure. Finally, all these methods only reduce, rather than eliminate, the potential for identification, and it is doubtful whether such quantitative reductions of identification risk will satisfy the qualitative requirements for exemption.

Because of the potential for inferential identification, studies of patient medical records data are not likely, in general, to be approved as exempt, so each study must be conducted under a protocol approved by an IRB with either informed consent or waiver of consent. It is thus unnecessary as well as inappropriate to destroy the link from patient data back to the patients. This view is consistent with positions expressed elsewhere.[2,10,26]

The question remains of to what extent patient data should be "scrubbed" in the building of a patient records research database. We have already accepted that all access to the research database must be controlled, with privileges granted on a study-by-study basis. According to the Belmont Report,[18] the "risks should be reduced to those necessary to achieve the research objective." This policy is also expressed in 45 CFR 46.111(a)(1). This principle, and not the satisfaction of a particular regulatory test, must govern the extent of "scrubbing" in building the database. A range of measures may be considered:

- Readily recognizable patient identifiers such as names, addresses (except ZIP codes), and telephone numbers are almost never needed in research and should generally be deleted in the loading of a research database.
- Patient and examination identifiers are less easily recognized and are important to the integrity of the research database, and should generally be included. Access to these identifiers should be restricted at the point of query instead.
- A further and clearly desirable measure would be the deletion of patient names and identification numbers embedded in free-text reports. Promising work has been done in this area.[27]

Deletion of names is not straightforward, since names may include aliases, nicknames, and names that match common words (e.g., Mr. Day). The occurrence of names in free text ranges from commonplace in referral letters, clinic notes, and other reports from primary care physicians, to very rare in consultation reports, such as those for radiology and surgical pathology. Until the routine deletion of patient names from free text has been proved feasible, the database will contain reports with embedded names, but investigators do not need to be granted access to these reports. The IRB review of each project plan protocol should take into consideration the likelihood that names will appear in the requested elements; it is expected that an IRB will be more restrictive of studies using these items.

Practical factors may also influence the amount of scrubbing that should be used. If a database server also performs clinical functions, such as providing failure backup for the production server, the elimination of names may not be appropriate, but it will still be possible to block access to names at the point of query. The guiding principle is that every practical opportunity to reduce disclosure should be taken.

Application of Regulations and State Laws

The federal human subject regulations govern the researcher, rather than the health care provider who creates and manages the patient records. In an integrated clinical and academic enterprise, such as that of the authors' institutions, the clinical and research units are governed by a common IRB. However, even if the provider is part of a separate organization, each institution is responsible for safeguarding the rights and welfare of human subjects in compliance with the federal human subjects regulations (45 CFR 46.114). Thus, in all research projects using patient data, the federal regulations pertain to the providers' roles in the research project, rather than their roles in patient care.

Physicians are also bound by standards of professional ethics.[28] Providers may also be restricted under state laws,[14,29,30] but many of these laws relate more to protecting the privacy of the records against court-ordered disclosure than to restricting their use by physicians. Forty-one states recognize this doctor–patient privilege,[30] and federal courts have recently affirmed it.[31] Physicians in 37 states are required to maintain the confidentiality of medical records, and health care institutions in 32 states are under a similar obligation.[30] Only nine states extend this duty to non-health-care institutions in possession of medical records,[30] although courts in many states have recognized a fiduciary duty to maintain confidentiality, the breach of which may be actionable.[32] No general federal law exists; the federal Privacy Act of 1974 applies only to records in the possession of federal government entities.[33] Similarly, all states have some form of legal protection for state-held patient records data.[15] Only a dozen states explicitly authorize disclosure for research, subject to various requirements.[15] Even in the absence of such authorization, however, research performed *by the health care provider* is not clearly addressed by state laws, and no simple rule answers the question of whether the provider is the researcher or is disclosing patient records to a researcher. In this heterogeneous environment, the local IRB is best qualified to determine whether a patient records research project is in compliance with both federal human subjects regulations and local legal requirements.

The challenge in regulating the research use of patient records data is to find a clear boundary between the research and clinical uses of this information, a boundary that can then be effectively policed. It is tempting to define the boundary as the point where raw data are extracted from the clinical data stream and moved into a ‘‘research computer.’’ However, this is not a satisfactory boundary, for a number of reasons. First, the division between ‘‘research’’ and ‘‘clinical’’ computers is not clear: A computer used primarily for research may also serve clinical functions, such as quality assurance or cross-patient searches in clinical care, or it may double as a standby system to back up the ‘‘clinical’’ computer. Second, much of the computational load of a research project is the extraction of the study sample from the much larger patient data archive. This is precisely the computational burden that the clinical production systems must be spared, so for practical reasons the entire database would have to be released to the researcher for each study.

As noted, any patient records database suitable for multiple projects will contain some sensitive data and so must be guarded. Transfer of patient records data to a research database server does not mean that any possible use of that data is therefore legitimate. Thus, *the clean boundary requiring regulated research access is the point of query*, not the point of data transfer to a research server. We assert that transfer of patient records data to a secure research database server within the institution does not itself constitute research use of the data. Instead, *every instance of access to the research database constitutes research use and must be specifically authorized by IRB procedures.*

Most institutions, including ours, require that the IRB office review *all* requests for use of human subjects and determine whether exemption is appropriate or a protocol is required. While not explicitly part of federal regulations, the procedures requiring that eligibility for exemption be determined by the IRB office are included in our institutions’ Multiple Project Assurance (MPA) filed with the Department of Health and Human Services Office for Protection from Research Risks (OPRR). Such policies are undoubtedly widespread, since the policy language is included in the Sample Language[34] provided by the OPRR as a starting point for each institution’s MPA, and any variance of an institution’s MPA from the Sample Language must be justified in the MPA filing.

For simple studies using only a few elements of the patient record, where inferential identification is impossible and anonymity can be enforced, IRB officials may find that exemption is appropriate. For more complex studies, a protocol with waiver of consent (or with a finding that the hospital’s standard consent has been satisfied) issued by the full IRB will be required. In either case, a formal approval must be obtained from the institution’s IRB office; in the rest of this article we refer to all such approved research procedures as ‘‘protocols.’’ The application for an IRB protocol must include explicit statements of the goals of the research, the methods to be used, the anticipated benefits, the need for use of the data, and the methods used to obtain patient consent.

Implicit throughout 45 CFR 46 is the ‘‘investigator’’ who performs the research. A common feature of institutional implementations of 45 CFR 46 is the designation of a principal investigator, who applies for the IRB protocol and is responsible for the conduct of the research in accordance with the protocol. Institutional policies define the eligibility requirements for principal investigators conducting human subjects research. At our institutions the principal investigator must be a faculty member with the rank of assistant professor or higher. As we discuss later, the role of the principal investigator is key to the security system we propose.

Federal code defines the conditions under which an IRB *may* approve a protocol with a waiver of informed consent, but it does not specify when the IRB *should* do so. Federal regulations correctly recognize that the complex issues of human subjects cannot be reduced to a list of rules to be applied administratively, and they thus rely on the exercise of judgment by the IRBs, guided by principles set forth in the regulations and in such documents as the Belmont Report.[18] The institutions and their IRBs are free to impose additional restrictions on their participation in such research.

Interests of Institutions

The three main institutional issues are protection from liability, privacy of business data, and fair sharing of any commercial benefits of the research.

Providers could be exposed to tort liability or official sanctions if information harmful to a patient is disclosed.[8] Research projects have not been a significant source of such disclosures in the past; the major breaches have been by health care workers with access to information during clinical care.[35] Still, the institutions have an interest in ensuring that research systems have security mechanisms sufficient to prevent abuse. Strict adherence to IRB procedures is the best protection from this risk.

Furthermore, patient records could also be processed to extract information about the hospital operations, such as operating procedures or service volumes and their trends, which could be used to the advantage of competing health care providers. The mechanism of IRB approval effectively eliminates the risk of this form of abuse, in several ways:

- Any protocol compiling hospital operations information would have to be sponsored by a principal investigator and approved by the IRB of each contributing institution.
- Each institution could establish additional administrative restrictions on acceptable protocols, which in all cases are required to demonstrate proper medically motivated reasons for the research.
- Any faculty sponsor conducting such espionage in violation of an IRB protocol would be taking enormous personal risk in return for a small benefit to his institution's business operations.

These disincentives are more than adequate without further controls.

Finally, a question of how to share commercial benefits of the research remains. Medical research using patient data may yield tangible commercial benefits. An example would be the searching of patient data to assist in a drug evaluation.[36] Medical records data are collected and maintained by provider institutions at considerable cost, and the institutions reasonably expect a fair share of the fruits of such research. Significant institutional support is often provided as part of projects funded by commercial entities such as pharmaceutical companies. A system is needed for fairly sharing the benefits of such research between the contributing institutions.

One can imagine fairly elaborate systems for sharing such benefits. The problem of codifying a sharing formula is difficult at the outset, since the economic benefits are often part of complex agreements that include support for collateral activities and are thus difficult to define. The simplest and most flexible structure may be that provided as a byproduct of IRB approval of each study. Each study will require a faculty sponsor at each institution, who can judge whether participation in the study is justified in relation to the effort required and the benefit to the institution. For smaller projects, informal reciprocal participation can be arranged. For large projects, inter-institutional agreements, including funding from the project sponsor, can be negotiated.

The key feature of this approach is that details of such cost- or benefit-sharing arrangements do not have to be specified at the outset. Instead, the requirement for a faculty sponsor at each institution means that such arrangements can be developed when needed.

Tools for Protocol Enforcement

As we have argued, if a study is to analyze patient data from multiple institutions, then the IRB at each institution must approve a protocol for that study. The protocol must be sponsored and supervised by a qualified principal investigator from that institution. The IRBs are accustomed to dealing with such multicenter protocols.

However, most multicenter trials are of a much larger scale and have a much larger budget than studies of patient records. For a multicenter clinical trial of a new therapy, the budget at each participating institution is sufficient to justify manual preparation of the protocol submission to the IRB and considerable effort of the faculty sponsor at each site to ensure that the protocol is properly followed. Furthermore, most of the activity in a clinical trial is visible and subject to scrutiny.

In contrast, the site sponsor for a multisite patient records study may be only minimally involved in the actual conduct of the study, and the data retrieval, communication, and analysis activities are not plainly visible. Site sponsorship by faculty sponsors is expected to be done for the most part on a reciprocity basis, with minimal direct effort by the site sponsor (other than the one originating the study) once the protocol is reviewed and approved.

For these reasons, the site sponsor needs tools to ensure that each project is conducted within the bounds of the approved protocol. Electronic record studies using protocol enforcement tools can reduce disclosure of individual identities, replacing manual studies in which researchers see patient names routinely. A computer-based patient records research system can thus improve patient privacy by reducing the use of readily recognizable identifiers in the course of medical research. However, no automated tools can substitute for human supervision and responsible conduct. All such programs of research must rely on the public's trust in the scientific community to use personal information responsibly and for the benefit of society.[37]

Shifting Ground

This discussion has been based on the current regulatory framework of 45 CFR 46. However, the legal basis for all uses of patient records may be significantly affected by recent and pending federal legislation. The Health Insurance Portability and Accountability Act of 1996 (HIPAA), which was signed into law August 21, 1996, requires the Secretary of Health and Human Services to promulgate patient records privacy rules by February 2000 if Congress does not act to define them legislatively. During 1997, several agencies and groups released reports and recommendations regarding the confidentiality of health records.[11,14,26,35,38] As required by HIPAA, the Secretary of Health and Human Services delivered to Congress on September 11, 1997, a comprehensive set of legislative recommendations covering the confidentiality of individually identifiable health information.[39] Also, bills regarding the privacy of medical records have been introduced in both the Senate and the House of Representatives, but the timetable for their adoption is uncertain.

While it is always hazardous to predict the course of public policy, some probable outcomes of the current changes can be inferred from an apparent consensus on the research use of patient records. The basic structure of the present IRB-regulated system is unlikely to change, although the general level of awareness and scrutiny will increase. Rules for waiver of consent may be reviewed and possibly tightened. Of special note, new criminal penalties for violations may be introduced. Among the criminal violations may be the act of intentionally identifying a person using records disclosed in a form not intended to be individually identifiable. The additional criminal penalties alone would be certain to increase the rigor of IRB supervision and enforcement.

Thus, the likely result of pending changes will be somewhat tighter regulation within the same IRB-based administrative structure. The protocol-based approval and enforcement structure described here is appropriate for the present environment and will be more important in the more stringent regulatory environment that is likely in the future.

Implementation

We constructed a ‘‘virtual repository’’ of breast cancer data, with data residing at two physical sites—Columbia University[40] and The University of Chicago.[41] This project investigated tools for regulating access and ensuring that research queries do not exceed the scope authorized by the IRB. Technical details of that implementation will be published separately.[42] Once the appropriate policies were defined, as described above, the implementation proved to be remarkably simple and was achieved using the security features of a standard commercial database management system.

The key concept is to manage access to the database in terms of individual IRB protocols. For each approved protocol, a database user account is created, which has privilege to access only specified columns of specified tables in the database. More restrictive protocols can be enforced by creating views (virtual tables) showing only specific information, and giving the account access to only those views. Access to each database account is controlled by a password, which is known only to the principal investigator and staff members designated by him or her. Each account is active only for the time period specified by the protocol and must be renewed on regular basis (e.g., annually), as required by the IRB.

Using this approach, unrestricted access to the database is never allowed; each query is posed in the context of a specific protocol that has been previously approved by the IRB and defined in the site's database server. The database still presents a standard relational (table-based) interface to the user, enabling researchers to exploit the full power of commercial database query tools.

More extensive security tools, which employ a security officer in the supervision of requests from a broader range of users, have also been described.[43,44] Our work has narrower scope and can be implemented by most sites using more readily available tools. More powerful and comprehensive security tools will undoubtedly be needed as the health care information infrastructure expands.

Conclusions

From the foregoing analysis of practical and policy issues, we have drawn the following key conclusions:

- *Security must be provided at the point of access,* not at the point of data entry into a research database.
- *The above requirement cannot be avoided by making the database "anonymous.''* All databases that link many facts about individuals are individually identifiable, even without names or other traceable identifiers. The linkage of many facts relevant to individual medical histories is central to the utility of a patient records database. The only way to make such a database unidentifiable is to randomize the links, thereby destroying the value of the database as a longitudinal patient record.
- *Names and addresses should, wherever feasible, be deleted when a research database is built,* but this does not make the data unidentifiable. The research database and security mechanisms must reduce the visibility of patient-identifying information to the minimum necessary to achieve the objectives of the research project.
- *Institutional Review Board approval is required for each study. Exempt research must still be approved by the institution's IRB office.* An authorized principal investigator at the provider institution must submit a proposal to the IRB office for each study. The proposal must state the goals of the research, the methods to be used, the anticipated benefits, the need for use of the data, and the methods used for obtaining patient consent or the justification for waiving consent.
- *For a multisite study, an IRB approved research procedure is required at each site* and, thus, a responsible principal investigator is required at each site as well.
- *The requirement for an approved research procedure and principal investigator at each site also protects the interests of each site institution* while satisfying public policy requirements for the protection of the interests of patients.
- *Principal investigators at participating sites need automated enforcement tools* to ensure that the research is conducted within the bounds of the IRB-approved procedure.

Computer-based patient records can fulfill their promise to yield knowledge to improve the public health, without harm to the privacy of patients. This can be done cleanly within the framework of present and likely future public policy, but only if practitioners understand and address the legal and ethical issues.

The authors thank Judith Argon, W. Dean Bidgood, Jr., and Charles E. Kahn, Jr., for helpful discussions and review of the manuscript.

References ■

1. Huffman EK. Medical Record Management. 6th ed, revised by the American Medical Record Association. Price E (ed). Berwyn, Ill: Physicians' Record Co., 1972.
2. Melton LJ. The threat to medical records research. N Engl J Med. 1997;337:1466–70.
3. National Institute of Medicine (U.S.), Committee on Improving the Patient Record. The Computer-based Patient Record: An Essential Technology for Health Care. Dick RS, Steen EB (eds). Washington, DC: National Academy Press, 1991.
4. National Institute of Medicine (U.S.), Committee on Improving the Patient Record. The Computer-based Patient Record. Rev ed. Dick RS, Steen EB, Detmer DE (eds). Washington, DC: National Academy Press, 1997.
5. Fitzmaurice JM. Computer-based patient records. In: Bronzino JD (ed). The Biomedical Engineering Handbook. Boca Raton, Fla: CRC Press, 1995:2623–34.
6. Epstein AM. The outcomes movement: Will it get us where we want to go? N Engl J Med. 1990;323:266–70.
7. Lazaridis EN. Database standardization, linkage, and the protection of privacy. Ann Intern Med. 1997;127:696.
8. Gostin LO. Health information privacy. Cornell Law Rev. 1995;80:451–528.
9. Gostin LO, Turek-Brezina J, Powers M, Kozloff R, Faden R, Steinauer DD. Privacy and security of personal information in a new health care system. JAMA. 1993;270:2487–93.
10. Institute of Medicine (U.S.), Committee on Regional Health Data Networks. Health Data in the Information Age: Use, Disclosure, and Privacy. Davidson MS, Lohr KL (eds). Washington, DC: National Academy Press, 1994.
11. National Research Council, Committee on Maintaining Privacy and Security in Health Care Applications of the National Information Infrastructure. For the Record: Protecting Electronic Health Information. Washington, DC: National Academy Press, 1997.
12. Rindfleisch TC. Privacy, information technology, and health care. Communications of the ACM. 1997;40:93–100.
13. Goslin LO. Health care information and the protection of

personal privacy: ethical and legal considerations. Ann Intern Med. 1997;127:683–90.

14. Lowrance WW. Privacy and Health Research. Washington, DC: Department of Health and Human Services, 1997.
15. Gostin LO, Lazzarini Z, Neslund VS, Osterholm MT. The public health information infrastructure: a national review of the law on health information privacy. JAMA. 1996;275: 1921–7.
16. Federal Policy for the Protection of Human Subjects, 56 Federal Register 28003 (Jun 18, 1991).
17. Protection of Human Subjects, 45 CFR §46 (1991).
18. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont Report. Washington, DC: Department of Health, Education and Welfare, 1979.
19. Rivest RL. The MD4 message-digest algorithm. Internet Engineering Task Force Web site. Network Working Group Request for Comments (RFC) 1320. Apr 1992. Available from: http://www.ietf.org. (Also available at: ftp://ftp.isi.edu/in-notes/rfc1320.text.) Accessed July 23, 1999.
20. Rivest RL. The MD5 message-digest algorithm. Internet Engineering Task Force Web site. Network Working Group Request for Comments (RFC) 1321. Apr 1992. Available from: http://www.ietf.org. (Also available at: ftp://ftp.isi.edu/in-notes/rfc1321.text.) Accessed July 23, 1999.
21. Federal Committee on Statistical Methodology, Subcommittee on Disclosure-avoidance Techniques. Report on Statistical Disclosure and Disclosure Avoidance Techniques. Statistical Policy Working Paper, vol 2. Washington, DC: Government Printing Office, 1978.
22. Boruch RF, Cecil JS. Assuring the Confidentiality of Social Research Data. Philadelphia, Pa: University of Pennsylvania Press, 1979.
23. Alexander LA, Jabine TB. Access to Social Security microdata files for research and statistical purposes. Social Security Bull. 1978;41:3–17.
24. Sweeney L. Guaranteeing anonymity when sharing medical data: the Datafly system. Proc AMIA Annu Fall Symp. 1997: 51–5.
25. Mueller W, Blien U, Wirth H. Identification risks of microdata. Sociol Methods Res. 1995;24:131–57.
26. Association of American Medical Colleges. Health Data Security, Patient Privacy, and the Use of Archival Patient Materials in Research. Washington, DC: AAMC, 1997.
27. Sweeney L. Replacing personally identifying information in medical records: the Scrub system. Proc AMIA Annu Fall Symp. 1996:333–7.
28. American Medical Association, Council on Ethical and Judicial Affairs. Code of Medical Ethics: Current Opinions with Annotations. Chicago, Ill: AMA, 1997.
29. Alpert SA. Health care information: access, confidentiality, and good practice. In: Goodman KW (ed). Ethics, Computing, and Medicine. Cambridge, England: Cambridge University Press, 1998:75–101.
30. Gostin LO, Lazzarini Z, Flaherty KM. Legislative Survey of State Confidentiality Laws: Final Report Presented to the U.S. Centers for Disease Control and Prevention. Atlanta, Ga.: U.S. Centers for Disease Control and Prevention, 1996.
31. Jaffee v. Redmond. 116 Sup Ct 1923 (1996).
32. Smith RE. Guidelines and mechanisms for protecting privacy in medical data used for research. In: Chapman AR (ed). Health Care and Information Ethics. Kansas City, Mo: Sheed & Ward, 1997:279–308.
33. Schwartz PM, Reidenberg JR. Data Privacy Law: A Study of United States Data Protection. Charlottesville, Va: Mitchie Law Publishers, 1996.
34. Department of Health and Human Services (DHHS), Office for Protection from Research Risks. Sample language for a DHHS Multiple Project Assurance (MPA) for compliance with DHHS regulations for the protection of human subjects (45 CFR 46), in accordance with the federal policy (effective August 19, 1991) and as may otherwise be amended. Rockville, Md: DHHS OPRR, 1998.
35. National Committee on Vital and Health Statistics. Health Privacy and Confidentiality Recommendations. Washington, DC: Government Printing Office, 1997.
36. Classen DC, Evans RS, Pestotnik SL, et al. The timing of prophylactic administration of antibiotics and the risk of surgical wound infection. N Engl J Med. 1992;326:281.
37. Lowrance WW. Modern Science and Human Values. New York: Oxford University Press, 1985.
38. President's Advisory Commission on Consumer Protection and Quality in the Health Care Industry. Confidentiality of health information. Consumer Bill of Rights and Responsibilities, chap 6. Washington, DC: Government Printing Office, 1998.
39. Secretary of Health and Human Services. Individually Identifiable Health Information: Recommendations of the Secretary of Health and Human Services, Pursuant to Section 264 of the Health Insurance Portability and Accountability Act of 1996. Rockville, Md: DHHS, 1997.
40. Johnson SB. NAMCIC Milestone 7 Report, Columbia Site Project. Arlington, Va: Defense Advanced Research Projects Agency, 1997.
41. Behlen FM. NAMCIC Milestone 7 Report, Chicago Site Project. Arlington, Va: Defense Advanced Research Projects Agency, 1997.
42. Behlen FM, Johnson SB. Security architecture for multisite patient records research. Proc AMIA Annu Fall Symp. 1999: 476.
43. Wiederhold G, Bilello M, Sarathy V, Qian X. A security mediator for health care information. Proc AMIA Annu Fall Symp. 1996:120–4.
44. Lin TY, Qian S (eds). Database Security XI: Status and Prospects. Proceedings of the 11th International Conference on Database Security; Aug 10–13, 1997; Lake Tahoe, Calif. New York: Chapman & Hall, on behalf of the International Federation for Information Processing (IFIP), 1998.