

# Confidentiality Legislation and the United States Federal Statistical System<sup>1</sup>

Joe S. Cecil<sup>2</sup>

**Abstract:** In the United States access to records maintained by federal agencies is often governed by statutes intended to address administrative uses of information. Those who seek such records for statistical purposes must respond to a system of regulation that addresses risks that do not arise with statistical uses and fails to offer protection that sensitive statistical records require. This paper demonstrates the problems of conducting statistical research with agency records under a general system of regulation intended for administrative

records (the Privacy Act of 1974) and a fragmented pattern of statutory regulation of individual statistical agencies (Bureau of the census, National Center for Educational Statistics, and the National Center for Health Statistics).

**Key words:** Statutory; administrative records; Private Act of 1974; Bureau of the Census; National Center for Educational Statistics; National Center for Health Statistics.

## 1. Introduction

Records maintained by federal agencies in the United States offer great promise as a resource for a wide range of research and statistical activities. However, obtaining access to agency records for such activities can be difficult. Records maintained by U.S. federal agencies are governed by a web of federal statutes that are "inconsistent at best and chaotic at worst" (Commission on Federal Paperwork 1977). The exchange of statistical information must conform to standards that often were

designed to guard against administrative abuses, standards that may be inappropriate for records used only for statistical purposes. As a result, researchers who seek information maintained by federal agencies often must recast their requests for access in terms of a regulatory scheme that does little to anticipate the special characteristics of statistical data.

This review of regulation of statistical records in federal agencies in the United States is intended to demonstrate the difficulty that arises when the primary means of regulating statistical records, in this case the Privacy Act of 1974, fails to distinguish between administrative uses and statistical uses of information. Such a distinction is desirable because of the difference in consequences to the individual data providers. Administrative uses of information can

<sup>1</sup> This paper was prepared for the Panel on Confidentiality and Data Access of the Committee on National Statistics. Portions of this paper appear as Chapter 5 of the panel report, *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*.

<sup>2</sup> Program Director, Federal Judicial Center, 1 Columbus Circle, N.E., Washington, D.C. 20002, U.S.A.

directly affect the individual data provider through administrative actions, such as a tax audit or determination of eligibility for benefits. The abuses of records that have lead to restrictions on exchange of information have been abuses of administrative uses of records (Flaherty 1989; Privacy Protection Study Commission 1977).

By contrast, statistical uses of information have no direct effect on the rights or benefits of individuals providing the information. Statistical uses aggregate information provided by individuals and do not employ records for making determinations about individual data providers. While the statistical use may influence policy decisions that ultimately disadvantage some data providers (e.g., research may lead to findings that cause income tax rates to change to the detriment of some who provided statistical information), the effect on individuals is indirect and mediated by a political process that is intended to be responsive to the public interest. Statistical records may include sensitive information that requires restrictions on dissemination, but the absence of determinations regarding the rights and benefits of individual data providers is assumed to allow a broader opportunity for exchanges of information for statistical uses.

This paper offers a brief review of several of the regulatory schemes that govern federal records that are of regulatory schemes that govern federal records that are of interest to researchers. First, regulation of statistical records under the Privacy Act of 1974 is reviewed. The requirements of the Privacy Act apply to records maintained by most federal agencies. Since the Privacy Act was intended to guard against abuses of administrative records its standards are ill-suited to statistical records, often impeding the release of identifiable statistical information necessary for research activ-

ities that do not disadvantage the individuals who provide the information. The Privacy Act also fails to protect sensitive information. Agencies with a need for greater protection of statistical records have obtained special protection under separate statutes that supersede the inadequate protections of the Privacy Act. The paper reviews several of these more demanding schemes of regulation as well. The thorough statutory protection extended to census records maintained by the U.S. Bureau of the Census is compared to the fragmented statutory protection of educational records maintained by the National Center for Education Statistics, and the more flexible statutory protection offered health records maintained by the National Center for Health Statistics.

## **2. The Privacy Act of 1974**

The Privacy Act of 1974 (5 U.S.C. §552a) is the first attempt by Congress to provide comprehensive protection of an individual's right to privacy by regulating the collection, management, and disclosure of personal information maintained by governmental agencies. Together with the Freedom of Information Act (5 U.S.C §552), which specifies conditions under which disclosure of federal records may be compelled, the Privacy Act specifies a general system of regulation for identifiable federal records. Before the Privacy Act was passed, federal policy toward data management practices encouraged data sharing among agencies in order to reduce the burden and expense of reporting. This open-access policy was restricted only when statutes provided for the confidentiality of specific sensitive record systems. The Privacy Act of 1974 reversed this general policy by recognizing the right of individuals to control dissemination of informa-

tion they provide about themselves to federal agencies.

The Privacy Act was passed in response to a series of abuses of administrative records. (Flaherty 1989, pp. 306–314). But in fashioning a remedy for these abuses, Congress extended this system of regulation to records used exclusively for statistical purposes as well. The result is a system of regulation that fails to adequately protect sensitive statistical records while imposing requirements on statistical records that are more appropriate to administrative records.

### *2.1. Requirements of the Privacy Act of 1974*

The Privacy Act seeks to strike a balance, preserving individuals' interests in controlling identifiable information while recognizing the legitimate uses of this information. Briefly, the Privacy Act requires that federal agencies must (1) grant access by individuals to their identifiable records maintained by federal agencies; (2) ensure that existing information is both accurate and timely, and limit the collection of unnecessary information; and (3) limit the disclosure of identifiable information to third parties. This third provision of the Privacy Act, forbidding the disclosure of any identifiable record without the prior written consent of the individual, is most relevant to researchers' access to federal data. This prohibition is also the crux of the right of privacy provided by the Act, since an enforceable consent requirement could thwart the disclosure of identifiable information for purposes that the individual never considered and would not approve.

### *2.2. Exceptions to the consent requirement*

The protections of the Privacy Act are greatly circumscribed by twelve categories

of exceptions to the consent requirement intended to accommodate legitimate needs for identifiable information. For instance, an agency may, at its discretion, disclose identifiable records without prior written consent to officers and employers of the agency who have a need for the record in the performance of their duties. Several federal organizations have interpreted the term "agency" to be quite broad, thereby restricting the protections of the Privacy Act. For example, the entire Department of Health and Human Services has been defined as a single "agency" under the Privacy Act, thereby permitting exchange of identifiable information throughout the Department as long as there is a job-related need for such information (NCHS Staff Manual on Confidentiality 1984, p. 12). Other exemptions to the consent requirement include disclosure to the Bureau of the Census for planning or carrying out a census, survey, or related activity under Title 13; disclosure to the General Accounting Office to permit auditing of federal programs; to the National Archives; disclosure that is required by the Freedom of Information Act; and disclosure in emergency circumstances involving the health and safety of any individual. The Privacy Act also permits disclosure of identifiable information without written consent for authorized civil or criminal law enforcement activities, and pursuant to a court order – disclosures that individuals most likely would decline.

Of special interest to researchers is an exemption that permits access to records that are not individually identifiable upon receiving "written assurance that the record will be used solely as a statistical research or reporting record" (5 U.S.C. §552a (b) (5)). A "statistical record" is defined as a record "maintained for statistical or reporting purposes only, and not used in whole or in part in making any determi-

nation about an identifiable individual, [except for certain research activities by the Bureau of the Census authorized by statute]" (5 U.S.C. §552a (a) (6)). The practical benefits of such an exemption for anonymous statistical records may be questioned, since a record that is not individually identifiable is not a "record" within the definition of the Privacy Act (5 U.S.C. §552a (a) (4)) and therefore is not subject to the restrictions on disclosure imposed by the Act. Nevertheless, it is important to note that the Privacy Act will not thwart the exchange of anonymous statistical records.

Fearing that it had failed to provide for all of the legitimate needs for identifiable information that merit an exclusion from the consent requirement, Congress also included a "safety valve" exemption, permitting disclosure without consent for a "routine use" of the record (5 U.S.C. §552a (b) (3)). A routine use is a use "for a purpose that is compatible with the purpose for which it was collected" (5 U.S.C. §552a (a) (7)). Instead of obtaining individual consent prior to disclosure for such a routine use, the agency must only publish a notice of the anticipated routine uses of the record in the Federal Register and accept comments from the public for a period of 30 days. Routine uses must also be explained to individual respondents when similar information is gathered in the future.

Two further points regarding consent for disclosure should be noted. First, the requirement of prior written consent of an individual may be avoided by inserting broad waiver provisions in the original request for information. If a person signs such a waiver, identifiable information may be released for purposes consistent with the waiver. Second, the Privacy Act places no obligation on the recipients of

information to maintain the confidentiality of the records or limit subsequent disclosure; once the records are released to a party not under the jurisdiction of the Act there is no assurance that the individual's rights will be protected.

### 2.3. Access to anonymous statistical data

Subject to the exemptions noted above, the Privacy Act prohibits disclosure by any agency of any record contained in a system of records to a person or to another agency without the written consent of the individual to whom the record pertains. The Privacy Act poses no barrier to the dissemination of anonymous information; if the research objectives can be accomplished with nonidentifiable data, rendering the data anonymous will serve to meet the standards of the Privacy Act and the information can be exchanged. Merely removing the name or individual identification number may be sufficient to permit disclosure under the exemption for statistical research if deductive disclosure is not possible (Dalenius 1986; Duncan and Lambert 1989). The exemption states that disclosure is permitted only if the record is "in a form which is not individually identifiable" (5 U.S.C. 552a (b) (5)). The guidelines for implementation of the Privacy Act interpret this phrase to mean not only that the information disclosed must be stripped of individual identifiers, but also that the identity of the individual

"cannot be *reasonably* deduced by anyone from tabulations or other presentations of the information (i.e., the identity of the individual cannot be determined or deduced by combining various statistical records or by reference to public records or other available sources of information)" [emphasis added] (Office of Management and Budget 1975).

This standard for deductive disclosure recognizes that some small degree of risk of deductive disclosure will exist in spite of efforts to remove identifying information, and demonstrates an awareness of the practical limits of such procedures that is not apparent in the language of the statute. If the statistical population is small and some of the variables are also recorded with names on publicly available lists, this guideline implies that precautions beyond the deletion of identifiers must be taken to guard against public disclosure.

#### *2.4. Access to identifiable statistical data*

Obtaining identifiable federal records to supplement statistical information that is maintained by another federal agency or maintained by researchers outside the federal government is particularly troublesome. The restrictions of the Privacy Act can bar the disclosure of identifiable records unless there is consent for the disclosure or the disclosure is brought within one of the exemptions. In an ideal situation researchers would be able to anticipate such needs and obtain the informed consent of research participants at the time the information is gathered. But even if a research participant agrees to subsequent release of agency information, the consent may be invalid if the researcher seeks access to records in a system that did not exist at the time consent was obtained.

When the need for research access to agency records was not anticipated or when the initial consent becomes invalid, a researcher may have to recontact the participants to obtain proper consent. Recontacting a participant in an earlier research study imposes special difficulties. For example, some target populations are highly mobile, so addresses and telephone numbers obtained at the initial encounter may be outdated. Recontacting such

research participants is likely to be expensive and subject to self-selection biases.

Even more formidable obstacles are faced by researchers who seek access to agency records to generate a sample of identifiable individuals to be contacted for participation in research. Since the purpose is to obtain a list of names and addresses of individuals, the researcher will be unable to contact the individuals to obtain consent for release of this information. Researchers employed by the agency maintaining the records may avoid such consent requirements by demonstrating a need for the record in the performance of their duties. But some researchers outside the agency have found the consent requirement a frustrating hurdle. At hearings of the Privacy Protection Study Commission, a number of researchers who rely on file linkage to conduct longitudinal research were sharply critical of the potential for disruption of research by the restrictions of the Privacy Act (Beebe 1981).

The following discussion focuses on the manner in which the Privacy Act impedes access to information for statistical research purposes, the specific focus of this paper. But it is important to note that the Privacy Act, applying indiscriminately to statistical and administrative records, also offer little protection from improper disclosure for nonresearch purposes of statistical records (Flaherty 1989). It is concern over the feeble protection of sensitive statistical records offered by the Private Act that has encouraged the development of some of the alternative statutory schemes of regulation mentioned below.

The most promising opportunity for disclosure of identifiable records for statistical purposes is found in the exemption that permits disclosure of an identifiable record for a routine use. The definition merits reiteration: routine use is a use "for a purpose

which is compatible with the purpose for which [the record] was collected" (5 U.S.C §552a (a) (7)). The manner in which the routine use exemption has been employed represents one of the greatest failings of the Privacy Act of 1974 (Flaherty 1989, pp. 324-4; 332). Such ambiguity in statutory language suggests that an agency may choose to define "statistical analysis" as a routine use of all or a selected portion of agency record systems, permitting researchers outside the agency to have access to identifiable records without gaining the consent of the individuals to whom the records pertain.

In fact, a great many agency notices allow for disclosure involving statistical research programs as a routine use. the Department of Health and Human Services has been particularly thorough in identifying record systems that have research potential and publishing notices permitting statistical research as a routine use (O'Neill and Fanning 1976). One version of the routine use notice requires an assessment of the risk and potential benefits of the research and requires the recipient to sign an agreement to protect the records from subsequent disclosure. This is one instance in which the discretion delegated to agencies by the Privacy Act has been used to fashion a specific set of standards to permit data sharing while providing safeguards that exceed those extended to administrative records. The need, however, to rely on the routine use exemption to overcome the failure of the statute to provide for research and statistical access to identifiable records is an awkward solution to the problem. With no explicit policy concerning access to federal records for statistical purposes, individual agencies are free to develop inconsistent regulations that may either be too restrictive or fail to offer adequate protection to the identified individuals.'

In summary, the Privacy Act's failure to distinguish research and statistical uses from administrative uses in restricting access to identifiable records poses a major obstacle for researchers who seek identifiable information. Regulation of administrative records is based on the awareness that the records may be used to make decisions regarding individuals, such as the award or termination of benefits. Such a system of regulation does not recognize dissimilar needs of statistical records in which the information is not used to make decisions regarding individual research participants. Researchers and statisticians have become adept at framing their research needs and protection in the standards developed for administrative records. In some instances it may be possible to anticipate the research purpose and obtain consent for disclosure of identifiable information at the time the information is solicited. Otherwise, the researcher must structure a request for access to identifiable information to fit within one of the exceptions to the consent requirement of the Privacy Act, such as the routine use exemption.

### **3. Bureau of the Census**

Unlike the lax protection of statistical records under the Privacy Act of 1974, the statutory protection of statistical information collected by the Bureau of the Census under Title 13 is so rigorous that many worthwhile studies by researchers outside the bureau can be thwarted even though the research poses virtually no risk of disclosure to research participant. In recent years the Bureau of the Census has gone to considerable lengths to develop standards and procedure that would permit greater access to information maintained by the bureau (Gates 1988; Courtland 1985; Zeisset 1985). But the statutory standards,

and interpretation of those standards by the Supreme Court, still pose a formidable barrier to the release of information outside the bureau in a form that permits the range of statistical analyses that are commonly permitted by other agencies.

### *3.1. Protection of data collected under Title 13*

Unlike agencies that must rely on the confusing and ill-fitting standards of the Privacy Act for regulation of both administrative and statistical records, the Bureau of the Census is governed by legislation that specifically addresses disclosure of census records. The Bureau of the Census is permitted to: (1) use census information only for statistical purposes; (2) publish data only in a way that prevents the identification of individuals; and, (3) prohibit anyone from examining information that identifies an individual unless they take an oath to uphold the confidentiality of provisions of Title 13 (13 U.S.C. §9 (a)). The bureau also may furnish "tabulations and other statistical materials which do not disclose the information reported by, or on behalf of, any particular respondent" (13 U.S.C. §8 (b)).

These statutory standards are among the few that have the benefit of an interpretation by the Supreme Court. In *Shapiro v. Bandridge* (1981), the Court considered the extent to which master address lists, compiled as part of the census, can be made available outside the bureau. Several cities challenged the 1980 census count of their populations, contending that the census had erroneously counted occupied dwellings as vacant, and sought to compel disclosure of a portion of the address lists used by the bureau in conducting its count in their respective jurisdictions. Although the case addressed access to this information for purposes other than research, in rul-

ing on the case the Court offered an interpretation of Title 13 that clarifies the limits of the discretion of the bureau to release statistical information that is individually identifiable as well.

The district courts had ordered the bureau to make the address register available, reasoning that the confidentiality limitation is "solely to require that census material be used in furtherance of the bureau's statistical mission and to ensure against disclosure of any particular individual's response." The Supreme Court reversed, interpreting the standards of Title 13 to suggest that the release of any microdata, even microdata not identifiable to an individual, is inconsistent with the standards of Title 13. The Court cited the constitutional purpose of the census in apportioning representation among the states and importance of public cooperation in obtaining an accurate census. According to the Court, the confidentiality protections of Title 13 are intended to encourage public cooperation by explicitly providing for the nondisclosure of certain census data, and "[n]o discretion is provided to the Census Bureau on whether or not to disclose the information referred to in §§8(b) and 9(a)" (*Shapiro v. Baldridge* 1981, p. 355).

The cities that sought the master address lists had argued that the confidentiality protections were intended to prohibit disclosure of the identities of individuals who provide census data. Furthermore, the Court rejected the contention that the confidentiality provisions protect raw data only if the individual respondent can be identified, raising a question regarding the authority of the bureau to release individual census data even when the identification of individuals is not possible:

"[Various parties] vigorously argue that

Sections 8(b) and 9(a) of the Census Act are designed to prohibit disclosure of the identities of individuals who provide raw census data; for this reason, they argue, the confidentiality provisions protect raw data only if the individual respondent can be identified. The unambiguous language of the confidentiality provisions, as well as the legislative history of the Act, however, indicates that Congress plainly contemplated that raw data reported by or on behalf of individuals was to be held confidential and not available for disclosure" (Shapiro v. Baldridge 1981, p. 355).

The bureau has not interpreted these standards as broadly as this language would permit and has continued to release unidentifiable microdata for statistical purposes. This opinion, while speaking of "data" and "statistical uses," is in fact about the authority of states and municipalities to audit the findings of the census, a purpose that was specifically precluded when the statute was passed. Furthermore, access to address lists would imply access to any individuals living at the addresses, so characterization of the research data as "unidentified" seems misplaced. Nevertheless, the language of the Supreme Court suggests that the Bureau of the Census has limited discretion to release data to persons who are not sworn to uphold the confidentiality provisions of Title 13.

### 3.2. *Release of public-use microdata*

Since 1963 the Bureau of the Census has released public-use microdata as part of its distribution of decennial census products. Prior to that time researchers outside the bureau were required to rely on published tabulations or to contract with the bureau for special studies (Gates 1988). In 1981, the bureau responded to increasing

requests for information by establishing a Microdata Review Panel to establish standards for release of microdata and to review and approve all such releases. Establishment of this panel recognizes that some risk of disclosure of individual information exists with the release of microdata, and seeks to ensure that such a release does not present an "unusual risk of individual disclosure" (Revised Criteria for Disclosing Public-Use Microdata 1981).

Under the standards established by the panel, microdata records: (1) cannot contain names, addresses, or other unique respondent identifiers; and, (2) cannot include geographic or related demographic information that would identify an area of less than 100,000 population when considered separately or in combination with other records. In addition, each microdata file is reviewed to determine whether the contents of the file would present an "unusual risk" of individual disclosure (Revised Criteria for Disclosing Public-Use Microdata 1981). If such a risk is found, the Panel has broad authority to require additional masking to reduce the disclosure risk, such as data grouping or aggregation, addition of random noise, rounding responses, and suppression of individual items.

### 3.3. *Alternatives to release of microdata*

In recent years the Panel's seemingly conservative standards for release have been criticized by users who seek more detailed geographical and administrative data. Gates (1988) offers of research requests for demographic microdata submitted to the bureau that could not be satisfied because of confidentiality concerns. In order to accommodate such studies, the bureau has developed a number of alternatives that will permit access to information within



the constraints of Title 13. For example, the bureau has expanded the opportunity for individuals to become a Special Sworn Employee as a means of obtaining direct access to census data for limited periods of time. Persons who are not bureau employees may examine confidential data if they are temporarily assisting with Title 13 work, and are sworn to observe the confidentiality requirements of section 9. The regulations implementing this authority suggested that such appointments are justified where a person outside the bureau must have access to protected information in order to assist with a bureau project (Special Sworn Census Employees 1983). The bureau has also expanded fellowship programs to enable researchers outside the bureau to use individually identifiable census data as long as they are sworn to observe the same confidentiality requirements as bureau employees. Recently, the bureau has been experimenting with locating some information in regional offices, thereby permitting more convenient access to researchers in other areas who are willing to serve as sworn employees.

In addition, the bureau is exploring a number of other options, such as the development of special masking schemes that are tailored to maintain confidentiality while permitting a specific research inquiry, releasing "surrogate public-use files" with transformations of sensitive economic data; and release of public use tapes with summary statistics (e.g., variance-covariance matrices, or correlation matrices of the data).

### *3.4. Non-Title 13 research by the Bureau of the Census*

Finally, it must be noted that the restrictions listed above extend only to census-based studies undertaken under authority of Title 13. The bureau may undertake

research under other authority and avoid the restrictions of Title 13 on release of identifiable information. Growing demand for identifiable information that can be used in conducting follow-up surveys or linked with administrative data has resulted in increasing numbers of reimbursable surveys sponsored by other agencies being undertaken by the bureau under authority other than Title 13. In such a study the sponsoring agency typically provides the sampling frame, since census records may not be used as a sampling frame if identifiable microdata is to be shared with the sponsoring agency. This would be a disadvantage only in those circumstances in which the more complete and accessible census sampling frame is desirable.

When seeking the consent of the respondent to participate in the non-Title 13 survey, the bureau makes clear that it is collecting the information as an agent of the sponsoring agency and that the agency, not the bureau, will be responsible for maintaining the confidentiality of the information. This assurance is intended to avoid an improper inference by the research participant that the information being provided will have the rigorous protection associated with census-related data.

## **4. National Center for Education Statistics**

Protection of research information on individuals (as opposed to organization) collected by the National Center for Education Statistics (NCES) is similar to protection of data gathered by the Bureau of the Census. This pattern of protection has posed unforeseen problems when extended to educational records that may be easily identifiable and that have been long available to the research community.

#### *4.1. Protection of data gathered under the Hawkins-Stafford Amendments of 1988*

The Hawkins-Stafford Amendments of 1988 (20 U.S.C. §1221e-1) set forth a rigorous system of protection of educational records of individuals collected and maintained by the National Center for Education Statistics. In brief, the statute: (1) prohibits the use of individually identifiable information for purposes other than the research and statistical purposes for which it was supplied; (2) prohibits the publication of information that will permit the identification of an individual; (3) permits examination of individually identifiable reports only by persons authorized by the Commissioner; and, (4) limits access to individually identifiable data to those who take an oath not to individually identifiable data to those who take an oath not to disclose such data. The amendments anticipate use of temporary employees as a means of disseminating individually identifiable statistical information to other federal agencies, state and local agencies, and private researchers, but only if such persons are sworn to observe the limitations described above.

Information collected as part of the National Assessment of Educational Progress, one of the ongoing studies of the NCES, is subject to an additional mandate to maintain the confidentiality of information with respect to individual schools (20 U.S.C. §1221e-1(i) (4)(B)(i)). This additional requirement poses particularly difficulty for those wishing use data from the National Assessment of Educational Progress to study the effects of programs at the level of the individual schools.

#### *4.2. Release of public-use microdata*

Although implementation of this authority

is still underway, the National Center for Education Statistics has developed a policy statement for the release of the statistical data in the form of public use files and statistical tabulations that closely tracks the policies of the Bureau of the Census (Policies and Procedures for Public Release of Data 1989). For each data set and tabulation there is a determination of the likelihood of disclosure of information about identifiable individuals based on "applying reasonable analytical procedures to any data set which NCES could expect to exist in the private or nonfederal sector." The review is conducted by a panel similar to the microdata review panel of the Bureau of the Census that reviews its disclosures. However, unlike the Bureau of the Census review panel, the NCES panel includes one member from outside the agency – a representative of the Bureau of the Census.

The policy statement sets forth standards used for review and recognizes that many analyses will pose no problem. For example, where the unit of analysis is at the level of the institution (excluding studies under the National Assessment of Educational Progress) and there exist no similar surveys of individuals (e.g., students, faculty, parents, or administrators), there is no possibility of disclosure and no assessment of risk is required. However, where the unit of analysis is the individual and similar surveys exist that would permit a linkage of information, the policy statement sets forth a procedure that examines common variables and distributions of responses to minimize the risk of disclosure. A review of nonfederal data files is also anticipated to ensure that such files will not present an opportunity for a match with NCES data that would yield individually identifiable information.

Files in which data on individuals are linked with related surveys of institutions

and organizations (e.g., schools, districts) require additional scrutiny. If data on individual students or teachers are linked with school files, the school file cannot include schools which can be uniquely identified. According to the policy statement:

“the assumption is made that school and school district administrators will know which students or teachers were interviewed in the survey, regardless of any procedures used to disguise the identify of these individuals or attempts to keep this information from the administrators. Therefore, if a school or district can be identified in a file, that file cannot be linked to student or teacher records” (Policies and Procedures for Public Release of Data 1989).

#### *4.3. Alternatives to release of microdata*

Where educational research requires individually identifiable information that cannot be released under these standards, the National Center for Education Statistics has developed a second set of policies that may enable the research to continue (Policies and Procedures for Nonpublic Use of NCES Data 1989). On occasion the NCES may undertake special analyses, although the opportunity for such services is presently quite limited. Through contractors, NCES also provides a service whereby researchers can obtain special tabulations which may include items not included on the public use files. The output of these tabulations are reviewed to ensure that individual data are not disclosed. NCES also makes data available at several university installations around the country and permits access only by individuals who are approved by the Commissioner and have sworn to uphold the standards discussed above. For example, analysis of data from the National Assessment of Educational Progress by non-agency researchers is per-

mitted only with such a license because of the restriction on disclosure of information regarding individual schools.

As indicated above, the National Center for Education Statistics is developing an access policy for public use files that closely resembles the policies adopted by the Bureau of the Census, and has gone beyond the practices of the Bureau of the Census in establishing licenses and other innovative means of disseminating identifiable data for research purposes. In developing such practices the National Center for Education Statistics is free of the close scrutiny that is accorded the Bureau of the Census due to its role in determining federal representation and participation in grants. But in other ways the task of the center is more difficult than that faced by the Bureau of the Census. With a limited number of school organizations, years of publicly available information on individual schools and districts, and a highly defined network of educational researchers who have become familiar with existing resources, the opportunities for inadvertent disclosure may be even greater than those faced by the Bureau of the Census. Applying a rigorous system of regulation to data which were often widely disseminated in the past has resulted in considerable difficulty in accommodating both the intension of the statutory protection and the research purpose of the agency.

### **5. National Center for Health Statistics**

Protection for health records collected, maintained and disseminated by the National Center for Health Statistics (NCHS), one of the Centers for Disease Control and Prevention, offers another example of the manner in which the generally inadequate protection offered by the Private Act can be supplemented through specific statutory authority. In fact, the

development of this protection and dissemination of information about it through manuals (NCHS Staff Manual on Confidentiality 1984), policy statements (NCHS Policy Statement on Release of Data for Elementary Units and Special Tabulations 1978), and publications (Mugge 1984), may serve as a model to be considered by other agencies.

### *5.1. Statutory protection of NCHS data*

Some of the most convoluted statutory language ever written (42 U.S.C. §242m(d)) states, in effect, that data obtained by NCHS may be used only for the purposes for which they were originally obtained, usually limited to statistical research and reporting. The statute further indicates that such information may not be disclosed outside the agency in identifiable form without advance, explicit consent of the person or establishment to which they relate.

A number of related publications expand on these issues. Most notably, the NCHS Staff Manual on Confidentiality (1984) offers a thorough discussion of these protections and the manner in which they are interpreted. The Staff Manual includes a number of useful definitions, some of which serve to broaden the extent of the protection. For example, the term "confidential information" is defined very broadly as "any information about an identifiable living person or establishment, when the person or establishment providing the data or described in it has not given consent to the center to make that information public." Note that this definition is not restricted to sensitive information.

In certain circumstances the names of establishments may be released for example, for purposes of developing a directory of health care institutions. Such a use is consistent with the purpose for which the data were collected and the consent of establish-

ments to such disclosures are sought at the time the data are collected. According to the Staff Manual on confidentiality practices, "it is not expected that any situations will arise in which the center will wish to disclose identifiable data on individuals."

The Staff Manual also recognizes that some degree of risk of deductive disclosure is always present when information is released, and proceeds to implement protection for "identifiable" information in a manner that takes into account the presence of some risk of deductive disclosure and the benefit to the public of pursuing such health care research. This is one of the very few instances of explicit recognition of this fact that appears in official agency policy statements. This issue is addressed in greater detail below in the discussion of the release of microdata and public use data tapes.

Use of the information collected by the National Center for Health Statistics is restricted to those uses that the individual consented to at the time the information was collected, without a catch-all "routine use" exemption as in the Private Act. In general, the research participants are informed of the authority that authorizes the solicitation of the information, whether disclosure is mandatory or voluntary, the principal purposes and uses of information to be made of the information, and the consequences, if any, of not providing the information. If identifiable information is to be released, then consent must be obtained for that specific release as well. The recommended assurance of confidentiality is as follows:

"Information contained on this form which would permit identification of any individual or establishment has been collected with a guarantee that it will be held in strict confidence, will be used only for the purposes stated in this study, and

will not be disclosed or released without the consent of the individual or the establishment in accordance with Section 308(d) of the Public Health Service Act (42 U.S.C. §242m)."

Clearly, the anticipated uses mentioned at the time the information is collected set the limits on the extent to which the data may be used by NCHS and others. Apparently, a general statement of anticipated uses will suffice. When data are collected over the telephone, the Staff Manual suggests that the respondent be informed that the information is being collected, "for statistical research on health problems."

Since the consent solicited by the National Center for Health Statistics rarely mentions disclosure of identified data outside the center, the opportunity for exchange of identifiable information, at least beyond the boundaries of the Department of Health and Human Services, would seem to be quite limited. According to the Staff Manual there is very little inter-departmental disclosure of identifiable information. The NCHS shares some information though interagency agreements with the Bureau of the Census concerning surveys that census conducts as a contractor for the center. NCHS has chosen to forego the census sampling frame with Title 13 restrictions and rely on less efficient area sampling frames for studies conducted on its behalf by the Bureau of the Census in order to have access to detailed microdata. The NCHS also participates in a number of cooperative agreements with state agencies, usually involving the receipt of information from the state agencies. Otherwise, the NCHS seems to have little need for the exchange of identifiable information with other agencies.

## 5.2. *Deductive disclosure of NCHS data*

The most notable aspect of NCHS policy

concerns the explicit recognition of the opportunity for deductive disclosure of individually identifiable information with the release of published tables and public use data tapes and acceptance of such risks in light of the importance of sharing statistical information. The manual notes that inadvertent disclosure of information may be of several types. (1) Disclosures may be "exact" in that a specific characteristic is associated with an individual {e.g., a specific income), or "approximate" in that a respondent is revealed as having a characteristic that falls within a certain range (e.g., income between \$15,000 and \$25,000). The manual suggests approximate disclosures associated with an individual may, in certain situations, be considered harmless because of the indefinite nature of the information. (2) Disclosures may be "probability-based" in that membership in a certain class suggests a specific probability of having a certain characteristic, as opposed to "certain" disclosure. The Staff Manual notes that "[i]n a sense, every published table containing data or estimates of descriptors of a specific population group provides probability-based disclosures on members of that group, and only in unusual circumstances could any such disclosure be considered unacceptable." However, there may be circumstances in which publication of data would reveal that "a highly specific group had an extremely high probability of having a given sensitive characteristic" and that such a table should not be published. (3) Finally, the Staff Manual distinguishes between "internal" disclosures that result completely from data published from one particular study, and "external" disclosures which rely on information outside the study to enable the disclosure.

Within this framework the Staff Manual present a number of guidelines for avoiding

inadvertent disclosure. Some of these guidelines are fairly standard. For example, when publishing tables the manual recommends that all cases in a line or column should not be found in a single cell, and that the total figure for a line or column of a cross-tabulation should be at least three cases. What is remarkable is that the document then continues to identify "mitigating circumstances in a given situation which make it acceptable to publish data that, strictly speaking, could result in 'disclosures'," which would justify a "special exception" to the guidelines. For example, if data are based on a sample that is a small fraction of the universe, it might be assumed that disclosure will not occur through published tables. Similarly, the existence of errors in the data or incomplete reporting may reduce the certainty of disclosures taking place to the point that would justify permitting the publication of otherwise revealing tables.

Until recently, an exception to the above standards that did not require a "special exception" involved the publication of vital statistics. For example, tables were published that indicated that within a specific county during a specific period there was one infant death or two deaths from rabies. Such exceptions were permitted because of "a long-standing tradition in the field of vital statistics not to suppress small frequency cells in the tabulation and presentation of data," and such publication "rarely, if ever, reveals any information about individuals that is not known socially." This exception was recently dropped to comply with standards of state agencies reporting these vital statistics.

In discussing the standards for the development of public-use microdata tapes, the Staff Manual, with refreshing candor, recognizes that:

"[t]he only absolutely sure way to avoid

disclosure through microdata tapes is to refrain completely from releasing any microdata tapes, but this would deprive the Nation of a great deal of very important health research. Therefore, the Center must make a determination as to when the public's need is sufficiently great to justify the risk of disclosure. It is the Center's policy to release microdata tapes for purposes of statistical research only when the risk of disclosure is judged to be extremely low."

In assessing the acceptability of the risk, the NCHS considers the extent to which the data involve a sample of the universe of relevant individuals or establishments, the extent and availability of outside information necessary to identify an individual or establishment, the expense of undertaking such an effort, and the sensitivity of the information provided.

Several methods are used to diminish the likelihood of inadvertent disclosure of information: geographical areas of less than 100,000 are not identified; sampling frames may not be revealed; and the recipient of a microdata tape must sign an agreement that the information will be used for research and statistical purposes only. The practice of injecting random error into public-use microdata as a means of reducing the probability of disclosure is discouraged since it lessens the value of the microdata for making sensitive analyses.

The standards of the National Center for Health Statistics stands as a refreshing counterpoint to the consideration by many agencies of the difficult issue of deductive disclosure of information. There is clear recognition that all disclosures involve some degree of risk and an explicit recognition of the role of the agency in balancing the degree of risk against the benefit to the public that is likely to arise from the research. Of course, the National Center

for Health Statistics has a number of advantages that may permit it the latitude to develop such policies. It is an agency that recognizes its primary role as research, thereby avoiding the difficulties that arise when trying to design a system for records that are used for administrative purposes as well. Furthermore, it is a relatively self-contained agency, collecting much of its information and not requiring the cooperation of agencies who may follow more restrictive practices. Finally, the NCHS maintains a skilled staff of researchers, permitting some of the more sensitive analyses to be completed without disclosure outside the agency. Nevertheless, the policies developed by the NCHS offer an opportunity to examine the consequences of policies that recognize the possibility of an inadvertent disclosure of identifiable information and attempt to minimize this risk while releasing information that permits research and statistical goals to be accomplished.

## **6. Conclusion**

This review of statutory protection demonstrates the varying levels of protection extended to research and statistical records maintained by federal agencies in the United States. Information may be easily available for research uses if governed by the Privacy Act, and similar information is essentially unavailable if collected as part of the census. The level of protection is governed not by the nature of the information but by the statutory authority of the agency that maintains the records. Even within some agencies, such as the National Center for Educational Statistics, statutes governing research and statistical records offer a fragmented pattern of protection. Variations in standards of protection across agencies results in a bewildering array of regulations which often

thwart interagency exchange of research records.

Agencies seek specific statutory protection to compensate for the generally inadequate protection of statistical records offered by the Privacy Act. This inadequate protection stems from the failure of the Privacy Act to distinguish between identifiable statistical data and identifiable data used for administrative purposes. In seeking specific statutory protection, each agency then fashions a system of protection with little regard for the patterns of protection in other agencies, making interagency exchange of information especially difficult.

Two changes in the current system of statutory regulation will ease this problem. First, the Private Act should be amended to include a separate system of regulation of research and statistical records. This recommendation has been offered before (Privacy Protection Study Commission 1977; Office of Federal Statistical Policy and Standards 1978), and has become more compelling with novel statutory protection for new research agencies such as the National Center for Educational Statistics. A separate uniform system of regulation of research and statistical records will address many of the concerns that lead agencies to seek separate statutory protection.

Second, a debate is required that will work toward developing a consensus regarding the proper degree of access to identifiable records for research and statistical purposes. Much of the concern about exchange of records is driven by concern over administrative abuses that may result. If such administrative abuses can be prevented by the separate system of regulation for research and statistical records described above, then discussions can focus on the risks and benefits of expanded opportunities for research activities and

procedures for protecting sensitive records from inadvertent disclosure.

The current system of regulation of research records has evolved from an excessive concern over controlling abuses of administrative records and with little awareness of research needs that may be common across federal agencies. As a consequence many research activities are thwarted, even though they are consistent with the purposes for which the information was provided. Perhaps the debate concerning research access to identifiable research records will conclude that broader sharing of records is not appropriate. But that conclusion should arise after consideration of potential research opportunities and duties to the research participants, and not abuses of administrative records.

## 7. References

- Beebe, G. (1981). Record Linkage and Needed Improvement in Existing Data Resources. Cancer: Brandbury Report 9, 661–666.
- Commission on Federal Paperwork (1977). Confidentiality and Privacy. 052-033-00458-5. Washington, D.C.: U.S. Government Printing Office.
- Courtland, S. (1985). Census Confidentiality: Then and Now. *Government Information Quarterly*, 2, 407–418.
- Dalenius, T. (1986). Finding a Needle in a Haystack or Identifying Anonymous Census Records. *Journal of Official Statistics*, 2, 329–336.
- Duncan, G.T. and Lambert, D. (1989). The Risk of Disclosure for Microdata. *Journal of Business and Economic Statistics* 7, 207–17.
- Flaherty, D.H. (1989). *Protecting Privacy in Surveillance Societies*. Chapel Hill: The University of North Carolina Press.
- Gates, G. (1988). Census Bureau Microdata: Providing Useful Research Data While Protecting the Anonymity of Respondents. Paper presented at Joint Statistical Meetings of the American Statistical Association.
- Mugge, R.T. (1984). Issues in Protecting Confidentiality in National Health Statistics. *Review of Public Data Use*, 12, 289–294.
- NCHS Staff Manual on Confidentiality (1984). DHHS Pub. No. (PHS) 84-1244, Washington, D.C.: Government Printing Office.
- NCHS Policy Statement on Release of Data for Elementary Units and Special Tabulations (1978). DHEW Pub. No. (PHS) 78-1212, Washington, D.C.: Government Printing Office.
- O'Neill, H. and Fanning, J. (1976). The Challenge of Implementing and Operating Under the Privacy Act in the Largest Public Sector Conglomerate – HEW. *Bureaucrat*, 5, 171–188.
- Office of Federal Statistical Policy and Standards (1978). *A Framework for Planning U.S. Federal Statistics for the 1980's*. Washington, D.C.: U.S. Department of Commerce.
- Office of Management and Budget (1975). *Privacy Act Implementation Guidelines and Responsibilities*. *Federal Register*, 40, 28948.
- Policies and Procedures for Nonpublic Use of NCES Data (1989). Washington, D.C.: National Center for Education Statistics (draft, August 14, 1989).
- Policies and Procedures for Public Release of Data (1989). Washington, D.C.: National Center for Education Statistics (draft, August 14, 1989).
- Privacy Protection Study Commission (1977). *Personal Privacy in an Information Society*. Washington, D.C.: Government Printing Office, 567–604, 052-003-00395-3.



- Revised Criteria for Disclosing Public-Use Microdata (1981). Federal Register, 46, 22017-01.
- Shapiro v. Baldridge (1981). 455 U.S. 345.
- Special Sworn Census Employees (1983). Federal Register, 48, 31894-05.
- Zeisset, P.T. (1985). Making Decennial Census Data Available. Government Information Quarterly, 2, 419-431.

Received September 1992  
Revised January 1993