

Robert W. Pearson, Social Science Research Council and Columbia University

## I. INTRODUCTION

In comments on a paper by George T. Duncan and Diane Lambert in the March 1986 issue of the *Journal of the American Statistical Association*, Joseph L. Gastwirth seized an opportunity to raise issues concerning the confidentiality of federal statistics and the principle of informed consent on which it can be argued the use of these statistics should rest. These remarks provoked considerable reaction by Thomas B. Jabine, Jerry Gates, and Fritz Scheuren.<sup>1</sup> Indeed, Gastwirth's published comments and drafts of the reactions by the others provide in part an agenda for the discussion of this panel.

Although I was not asked by the organizers of this panel to respond to Gastwirth's comments, the temptation to do so is alluring, and I will devote some attention to the *JASA* commentary. I will also suggest how to better resolve one set of existing tensions between access and confidentiality with which I understand this panel to be concerned -- researchers' access to publicly collected data. In this latter regard, I will focus attention on the set of laws and regulations that currently affect researchers' access to federal statistics.

My interest in this discussion stems from two sources: (1) as staff to a conference sponsored by the Social Science Research Council (SSRC) in November 1985 on Researcher's Access to Publicly Collected Data and (2) as my role as Staff Associate at SSRC, an organization that has an interest in the development of fields of inquiry within the social sciences. Clearly, fields such as labor economics, criminology, evaluation research, and public policy studies, to name but a few such fields, have benefited from their access to microdata files sponsored and collected by U.S. statistical agencies. In turn, this research has helped to enrich and improve these statistics and our understanding of people, institutions, and their relationships.<sup>2</sup> I speak only for myself, however, and my views do not necessarily reflect those of SSRC.

## II. DEFINITIONAL DETOUR

A brief definitional detour is required before turning to the discussion of the extended uses of federal statistics. Attention to definitions is required in part because several concepts -- closely related but substantially different -- tend to be confused in discussions of this kind. These discussions often fail to distinguish among issues of privacy, confidentiality, anonymity, and informed consent.

*Privacy* involves control over information about the self. Indeed, privacy is usually not at issue in the release of publicly collected data because the act of collecting information (sometimes voluntarily, sometimes under the compulsion of law) is in some ways already an invasion of an individual's privacy. Violation of the anonymity of

the data is more frequently at issue.<sup>3</sup>

*Anonymity*. The concern with anonymity is linked directly to the protection of an individual's identity. Anonymity is not always linked to the content of information about the person (although identity and information are clearly related). For example, some states permit reporters to protect the anonymity of the source of their information while compelling them to disclose the information that these sources reveal, thus abridging the confidentiality of that information even though it may not be possible to associate this information with a particular individual and thus assure its anonymity. Reducing the risk of disclosing the identity of individual records are more accurately defined as a concern with protecting their anonymity; not their confidentiality, as if frequently asserted.

*Confidentiality* is a property inhering in the information and the manner in which it is collected. The content is private, at times even secret; it is given with an understanding that the information will be cared for by those to whom it is entrusted. Confidentiality also embraces the principle of a special or privileged relationship so important that the state should not force its disclosure. It includes, for example, information exchanged between a doctor and patient, and between a clergyman and parishioner. (The privileged status of the researcher/subject relation continues to be adjudicated in U.S. courts on a case-by-case basis.) The overriding concern with confidentiality is with the *content* of the information.

*Informed consent* directly concerns knowledge about the use of information which is shared between those who collect information and those about whom the information is collected.

This brief definitional detour is unlikely to alter current usage or eliminate our current confusion of such concepts as privacy, confidentiality, and anonymity. Each has multiple and overlapping meanings that are likely to remain entangled in our discussion. My purpose is to anticipate an argument I will subsequently make that many of our concerns about informed consent and anonymity drop away when we disassociate responses from the individuals who provide this information because the identity or self is removed from the record.

## III. THE CASE AGAINST THE EXTENDED USE OF FEDERAL STATISTICS

Gastwirth argues:<sup>4</sup>

- U.S. federal agencies are collecting data voluntarily from people and are using these data in ways not described to nor agreed upon by respondents. (Although the context of his comments concern the literature on the preservation of

confidentiality of data, it does not necessarily follow that using data for purposes other than those revealed to respondents is a breach of confidentiality.)

- This practice is wrong on practical grounds because it threatens to reduce respondents' willingness to answer questions in future surveys.
- This practice is also wrong on ethical grounds because it does not rest on the principle of informed consent, which is embodied in norms of fair play and codified in such statutes as the Privacy Act of 1974.

In order to support these points, Gastwirth lists several examples of past, current, and planned activities of U.S. federal agencies.<sup>5</sup>

#### *Securing the informed consent of respondents.*

There exists a disagreement between Gastwirth and others concerning whether respondents were informed of and consented to the uses of records to which he specifically refers in his comments. Gates, in his draft commentary, argues that informed consent is present, at least with respect to the planned match of currently collected survey data from the Survey of Income and Program Participation (SIPP) and administrative data. Gates notes that the U.S. Bureau of the Census notifies SIPP respondents that "we will combine data from SIPP with data from other Government agencies to provide a comprehensive set of summary information about employment, income, and participation in various Government programs."

One could object that this notification is not sufficient to constitute informed consent, because it fails to specify, for example, which government agencies are to provide data. Different agencies collect information which are more sensitive, even secretive, than others. Moreover, different agencies have different interests and powers of enforcement about which a respondent might be reasonably concerned. The appearance or potential of improper use of linked records cannot meaningfully be assessed by a respondent who is told only that his or her responses may be combined with data from "other Government agencies. . . to provide a comprehensive summary, etc." An agency like the IRS, because of the sensitivity of its data and its enforcement capacities, should have a greater need to establish that its data will be used in ways that are agreeable to the respondent than would an agency without such police power.

Several practical issues arise, however, with respect to the uses a respondent can be meaningfully and reasonably informed about, uses which may at the time of the interview may be unforeseen.<sup>6</sup> The issue reduces to the messy question of what is *reasonably* informed consent.

*A general principle for the use of federal statistics.* Devising standards, definitions, or guidelines for reasonableness is obviously difficult. We may gain some leverage with this question by reference to the literature on "deception" in social science research. Terry Pinkard, for example, usefully argues

Our judgments in those cases should be set in the context of a model of moral reasoning that focuses on principles that are *shared* between people and to which we can imagine people *contractually* agreeing. It is not the consequences (in the utilitarian sense) of adopting a principle that justifies it, but its being (at least hypothetically) *agreed upon* [his emphases].<sup>7</sup>

Pinkard, for example, argues that it is justifiable to deceive someone in a game of poker, because of the context of the game. The shared understanding among the players concerning rules and purposes of the game make it unnecessary for each player to explicitly acknowledge prior to playing that they may seek to deceive the others.

The analogy is undoubtedly stretched between deceit in a game of poker and the use of data in ways that respondents do not specifically agree. Deceit implies an intention, for example, that does not usually exist in using data in ways such as those described in Gastwirth's examples. Instead, the purpose of the analogy and of Pinkard's judgment rule is to suggest a general criterion for extending the use of federal statistics beyond those to which are explicitly and specifically agreed to: *Permit the use of information as long as it seems reasonable to imagine that people would agree to this use in the context of prevailing social norms and values.*

The use to which Gastwirth's examples are being put falls within this criterion. No direct action concerning any individual is at stake. Neither the content of anyone's individual record nor the anonymity of that record is revealed by improving the accuracy and reliability of the statistics upon which major national policies may be decided, purposes for which a more explicit contract is unnecessary.

*Implementing principles.* Unfortunately, implementing a general principle may be more troublesome than agreeing to it in the first place. This difficulty applies no less to the application of "golden rules" that Gastwirth would have guide our conduct.<sup>8</sup> The principle suggested above need only be evoked, however, in situations for which rules and standards such as those suggested by Jabine in this panel are ambiguous or uncertain. Other -- perhaps even less well specified -- principles currently govern decisions regarding the extended use of federal statistics. Moreover, the locus of decisions about the release or restraint of data resides with program officers of the data collecting or sponsoring agency. Proposals to release or deny access are not vetted, for example, by commissions whose membership include broader social interests, e.g., public officials outside the agency, scholars, and other members of the public.

Considerable precedent exists for enlarging the constellation of interests represented in such decisions insofar as these decisions affect society, as well as the administration of a particular agency. The commissions that several European countries have established to make decisions about access and the Institutional Review Boards throughout universities in the United States provide clear examples. My purpose is not to outline how to implement any principle; rather to suggest that there exist widely agreed upon standards (see especially Jabine's remarks) and at least one as yet

not as widely agreed general principle suggested above for governing the extended use of information for which prior consent is not present or possible.

#### *Noncooperation in future surveys.*

Gastwirth's second point may be translated into a question for which an empirical answer is possible, although not as yet adequately answered in the existing research on the subject. It may be stated as follows: What are the opinions and behavior (e.g., refusal rates) of respondents to federal surveys which vary in the levels of information they provide about the use of their responses (perhaps also under varying levels of assurance regarding confidentiality). Previous related research such as Singer and the National Research Council's Panel on Privacy and Confidentiality as Factors in Survey Response provide interesting and useful insights into the role played by pledges of confidentiality, informed consent, and relatively more or less elaborate introductions as to the purposes and anticipated uses of these surveys.

This research suggests that many members of the general public have a poor understanding of what may be implied by confidentiality or even that such pledges can be trusted. This research also suggests that such pledges play a relatively small role in affecting response rates concerning *innocuous* data, and a relatively greater role in affecting response rates to more sensitive questions.

Existing public opinion research on this issue is inadequate (even potentially misleading). It is also frequently beside the point being considered here. Conventional survey research methods on attitudes toward confidentiality, anonymity, and privacy, for example, are likely to be subject to substantial context, wording, and order effects. An incidental question or two about access and confidentiality added to an existing survey may evoke -- because of the prior question context (often unrevealed to those to whom these results are reported) -- either images of government efficiency on the one hand or of Big Brother on the other. Although the conceptual distinctions between anonymity, confidentiality, and privacy can be made, our language and common use of these terms is so overlapping (nay, sloppy) that one should not expect the public to be able to respond meaningfully to questions that casually use these terms or that evoke such substantially different meaning and images as do these issues.

Public opinion research on attitudes toward the extended use of federal statistics may be unrelated to the concern for disruptions to data collection programs. Research on pledges of confidentiality suggests that a small proportionate difference in these pledges may make a big difference to a small number of respondents. These small numbers are important for practical and philosophical reasons. Although a few objecting people may not affect the overall quality of the data, especially if respondents' sensitivity to pledges of confidentiality are unrelated to the information being collected, they can expose a data collection program to debilitating public criticism. It does not require a majority of adult U.S. population to decry and cause to be discontinued a data collection program or for the media to seize upon (whether or not correctly) practices such as those noted in

Gastwirth's comments as a vehicle for raising the spectre of Big Brother.

Difficult philosophical issues also arise in the context of interpreting public opinion data of the sort that have been collected to date. The average or modal opinion or behavior does not necessarily settle matters of moral inquiry which are touched by issues of privacy, confidentiality, and informed consent.<sup>10</sup> To show, for example, that 90 per cent of the public is fully supportive of linking federal data sets or of permitting researchers' access to publicly collected data, or to demonstrate that 90 per cent or more of the public will consent to provide interviews when they are informed that the data are to be used in this way, begs important and longstanding ethical and philosophical questions concerning the rights and obligations of the other 10 per cent.

What behavioral consequences are likely to follow from using data in ways not specifically agreed upon by respondents? Nonresponse rates in some surveys in the United States appear to have increased during the last 20 to 30 years. And refusals to be interviewed comprise an increasing proportion of all reasons for nonresponse.<sup>11</sup> It is difficult to conclude, however, that increasing refusal rates are attributable to declining trust in federal statistical agencies, which itself is due to the absence of informed consent concerning their extended use. Singer, for example, reports that about two-thirds of those who refused to be interviewed in her study did so prior to an interviewer being able to get their foot in the door. This finding suggests a turnabout to the concept of informed consent: *most nonconsent is uninformed* about the specific sources and uses of a survey.

However, there exist known instances of fairly substantial noncooperation (e.g., the West German census) and reduced response rates (e.g., response rates to surveys conducted by Statistics Sweden have fallen following the highly politicized case of Project Metropolitan) associated with public debate concerning the confidentiality of records. Moreover, uninformed nonconsent in surveys in the United States is undoubtedly linked to the objectionable practices to which many Americans have been exposed in answering questions to a "survey" that turns out to be a pretext for selling products or services of one kind or another.

Unfortunately, the type of attitude (i.e., trust) on which rests cooperation in surveys may be a general or diffuse orientation for many of the public and may have little to do with what a specific agency does or does not do with the records it collects. Consider as evidence for this diffuse orientation findings from the National Research Council's study: only about half of the respondents knew that unemployment statistics were gathered by a government agency (only 2 per cent could specifically name the Bureau of the Census); nearly one-fourth of those who professed to have heard about the Census, failed to recognize it as a government undertaking; nearly half failed to identify the Bureau as the Census taker; about one-fourth did not know whether compliance with the Decennial Census was required by law, and one-fourth thought it was not.

Concerns with the abuse of information by government agencies are important. Indeed, a great deal of the current literature on ethical and legal issues in social research concerns the protection of researchers' records from judicial, administrative,

and legislative inquiry.<sup>12</sup> There exist widely publicized uses of administrative records whose use we can imagine would not have consented to.<sup>13</sup> We run the danger, however, of throwing the baby out with the proverbial bath water if we indiscriminantly link issues of administrative abuse of information to other legitimate uses to which such data may be put. In this regard, it is important to bear in mind the distinction between the administrative and research purposes of data and to focus on the purposes to which information are put.<sup>14</sup>

*The ethics of using data for purposes not specifically agreed upon.*

*For what purposes are data being used?* In the five cases Gastwirth notes, exact matches or linked files are being created. These linked records are made possible by unique identifying information (e.g., name or social security number of respondents). But the identity of the records is incidental to the use of the data. That is to say, in each example, a newly created linked record is not tied to decisions that these agencies make about an individual respondent. The purposes of linking these files, which Gastwirth acknowledges, are to improve the accuracy and reliability of statistics; not to decide upon rewards or penalties for any individual who provides the information. In the current examples, the "self" serves the single purpose of permitting records to be matched for research on the quality of statistics themselves -- the "self" is otherwise unimportant.

Both informed consent and confidentiality attach considerable importance to the manner in which the data are used. Had Gastwirth's examples of linked records been used for administrative purposes, identification of the records would have been necessary and the anonymity of records violated. This use appears not to be the case in these examples.<sup>15</sup>

*Confidentiality and anonymity.* Interestingly enough, Gastwirth's illustrations of objectional practices do not focus directly on issues of confidentiality and anonymity. These issues enter when he suggests that preserving the confidentiality of individual observations is made more difficult when survey data are "combined with administrative records for a large portion of the population, . . . especially if the person trying to obtain information has a 'list' of targets and has concomitant information about them" (p. 24). I will pass quickly over two problems in this formulation: (1) Rarely do sample surveys include a large portion of the population. This problem is more properly specified as arising with large portions of subpopulations or with individuals who have characteristics that are "rare" in the population (although they may be rare only in the combination of their characteristics as would be, for example, the class of all families living in a particular census tract with a total family income of \$97,899 in 1986 who owned three cars, and had three children who were 7, 8, and 9 years of age) (2) "Lists of targets" implies that one knows that an individual is in the file. This is highly unlikely for most surveys of the kind Gastwirth mentions. Of more interest is Gastwirth's implication that use of these records is made by someone outside of the agency that

collects the data and that their purpose is to (re)identify records.

Although not at issue explicitly in the examples cited by Gastwirth, it follows that he would extend his concern with making such data available to anyone outside of the agency who sponsored or collected the data unless such a purpose was specified when collecting the data. (To my knowledge none of these surveys specify in any detail that the data may be distributed in public use files in which the identities of individual records are removed.)

*Deductive disclosure.* Let us assume that unique, but single, identifiers are removed from microdata files of individuals to which outsiders are given access, as is suggested by Gastwirth.<sup>16</sup> On the surface, it appears that removing or disguising names and other single, identifying characteristics of individual records may make moot the issue of *privacy*, insofar as personal identification is removed from these records. The question remains whether the probabilities of reidentifying data can be made sufficiently small to warrant an extended use of these data as, for example, in the release of microfiles to researchers. My discussion to this point ignores the troublesome possibility that the identity of an individual record may be deduced from the combination of information contained in the full record.

This concern with disclosure risk drives the decisions of federal statistical agencies concerning what data to release -- if any -- in what form to whom (e.g., including other statistical agencies as well as researchers outside of the agency). Current statutes and regulations provide wide discretionary authority to federal agencies in determining whether such disclosure risks are acceptable and what statistical data will be released. These same statutes, however, place the statutory obligation for protecting the anonymity of records nearly exclusively with the agency that collects the data.

Acceptable disclosure risks are neither easily nor precisely calculated, but such agencies as the Bureau of the Census and the Internal Revenue Service often require (or interpret the laws that govern the release of such data as requiring) that these levels equal zero.

I reveal my prejudices here, if not before, in believing that the extended use of federal statistics per se is not inappropriate; but rather that (1) the value of these data are not fully realized and (2) most current statutes under which the release is governed are inadequate because they recognize only the obligations of those who collect the information, not the obligations of those who may subsequently use them.

#### IV. STATUTORY SUPPLEMENTS TO STATISTICAL AND PROCEDURAL PROTECTIONS

Clear criteria and broadly defensible thresholds of disclosure risk do not currently exist. It is entirely possible that they may never exist. In any case, their absence poses a considerable burden on those who must make such decisions in an environment in which the uncertainties of future disclosure appear increasingly larger than the certainties (convenience and reduced expense) of

limiting or denying access altogether, and where the uncertainty and hostility of their environment may appear all the more troublesome by published comments such as Gastwirth's. Under such circumstances it is likely that decisions concerning the release of information for purposes of research will fail to assign appropriate costs and benefits to the extended use of this information. Increasingly sophisticated computational and analytical technologies make it easier to breach the anonymity of individuals and institutions who are subjects to government-sponsored surveys and administrative records. Federal agencies may, therefore, seek zero disclosure risks by withholding data altogether, thus denying society an opportunity to better understand itself and to check against the inappropriate interpretation and use of statistics to which only the government would then have access.

Increasingly sophisticated statistical techniques are being developed for reducing the risks of reidentifying individuals from microdata or statistical tables. These advances are to be commended and supported. I concur with those who argue that there exists no single strategy to deal with such complex problems and that procedural and statistical approaches to protecting the anonymity of data under extended uses are as necessary as are statutory approaches.

Statistical and procedural techniques alone, however, may be insufficient in protecting individual respondents against someone who, with considerable resources, sets out to identify an individual. And many of these techniques will have the unfortunate consequence of reducing the quality of data for research purposes.<sup>17</sup>

But what reasonable statutory conditions exist for the protection of records? An answer to such a question is in part contained in current -- if not uniformly used-- practices that are embodied in the numerous laws and regulations that now govern access to data collected by various federal agencies. The following conditions would appear reasonable for the extended use of these data, even beyond explicitly specified conditions of informed consent.

Stated as a rule, *research use of microdata files should be permitted as long as:*

- *a researchers' use of such data is protected from administrative, judicial, or legislative inquiry and subpoena;*
- *secondary users explicitly consent to protect the anonymity of records; and*
- *sanctions (civil and/or criminal) are applied against researchers or anyone who violates the anonymity of these records*

and, as argued above,

- *the purposes to which the data are put are ones that we might reasonably argue respondents would find agreeable.*

Society should not place an exclusive burden on federal statistical agencies for assuring the anonymity of respondents any more than society places an exclusive burden on automobile manufacturers for all traffic accidents. Nor should they be required to meet standards of zero disclosure risks. Although the law recognizes the culpability of automakers for designing automobiles without regard for the safety of the passengers and

drivers, it does not make every misuse of these vehicles their exclusive obligation nor does it require a guarantee that no injuries will follow from their use.

Most current statutes concerning the release of federal statistics place an exclusive burden on the collectors of such data. In order to protect society from breaches of anonymity, we should add to the armament of statistical and procedural techniques the force of the law which recognizes the obligations of the research community who benefits from such access.

#### Notes and References

<sup>1</sup>Mr. Jabine is a consultant to the Committee on National Statistics of the National Research Council and current chair of ASA's Committee on Privacy and Confidentiality; Mr. Gates is a program officer at the U.S. Bureau of the Census; and Mr. Scheuren is Director of the Statistics of Income Division of the Internal Revenue Service.

<sup>2</sup>My remarks, however, are limited to microdata records related to individuals; not institutions or groups of individuals.

<sup>3</sup>Whether such disclosure is (potentially) harmful is another area of contention, but one not considered in any detail here.

<sup>4</sup>Gastwirth, J. L. 1986. Comment on G. T. Duncan and D. Lambert, Disclosure-limited data dissemination. *Journal of the American Statistical Association* 81:23-25.

<sup>5</sup>These five cases are: (1) a link of estate tax return data with tax returns of their heirs by the Internal Revenue Service (IRS), (2) matches of dates of birth provided by the National Center for Health Statistics with data from the U.S. Bureau of the Census, (3) a match of data from the March 1973 Current Population Survey (CPS) with data from the Social Security Administration and IRS, (4) a planned link of data from the Survey of Income and Program Participation (SIPP) with data from IRS, and (5) a proposed link of data on firms collected in the Economic Census with data from CPS and SIPP on a firm's employees.

<sup>6</sup>Jabine discusses this issue in his draft comments.

<sup>7</sup>Pinkard, T. 1982. Invasions of privacy in social science research. In *Ethical Issues in Social Science Research*. T. L. Beauchamp, R. R. Faden, R. J. Wallace, Jr., and L. Walters. eds. Baltimore: The Johns Hopkins University Press, p. 270.

<sup>8</sup>Many of us can agree in principle that federal agencies should "tell the whole truth" to respondents concerning the uses to which their information are to be put. But what is the *whole truth* and who is to decide it? For example, does the whole truth include a description of masking techniques that are to be applied to the data prior to public release (e.g., micro-aggregation, random noise multiplication, linear transformation)? Should the U.S. Bureau of the Census inform respondents to the SIPP that a micro data review panel at the Census Bureau assesses what data in what form are to be released to researchers? Indeed, wouldn't the

whole truth include a list of the members of the micro review panel and a description of each members' philosophy concerning access and confidentiality?

<sup>9</sup>Singer, E. 1978. Informed consent: Consequences for response rate and response quality in social surveys. *American Sociological Review* 43:144-162. National Research Council. 1979. *Privacy and Confidentiality as Factors in Survey Response*. Washington, D. C.: National Academy Press.

<sup>10</sup>Caplan, A. L. 1982. On privacy and confidentiality in social science research. In *Ethical Issues in Social Science Research*. T. L. Beauchamp, R. R. Faden, R. J. Wallace, Jr., and L. Walters. eds. Baltimore: The Johns Hopkins University Press, 315-328.

<sup>11</sup>Madow, W. G., H. Nisselson, and I. Olkin. eds. 1983. *Incomplete Data in Sample Surveys*, Volume 1. New York: Academic Press.

<sup>12</sup>This literature tends not to concern itself less with researchers' identification of respondents, undoubtedly because the analyses of these data by many social scientists over a considerable period of time has not produced a harmful breach of the anonymity of a single respondent in the United States. See, for example, Boruch, R. F. and J. S. Cecil. 1979. *Assuring the Confidentiality of Social Research Data*. Pittsburgh: University of Pennsylvania Press. Dahmann, J. and J. Sasfy. 1982. *Confidentiality of Criminal Justice Research Data*. The Mitre Corporation. Knerr, C. R., Jr. 1976. What to do before and after a subpoena of data arrives. In *The Ethics of Social Research: Surveys and Experiments*. J. E. Sieber. ed. New York: Springer-Verlag. Nelson, R. L. and T. E. Hedrick. 1983. The statutory protection of confidential research data: Synthesis and evaluation. In *Solutions to Ethical and Legal Problems in Social Research*. R. F. Boruch and J. S. Cecil. eds. New York: Academic Press.

<sup>13</sup>An article appearing in the *Washington Post* in May of this year, for example, announced that Big Brother was born November 9, 1977, when Joseph

A. Califano, Jr., Secretary of what then was the Department of Health, Education and Welfare, announced Project Match. This project compared a list of welfare recipients in 21 states and jurisdictions where there were high numbers of either welfare recipients or federal employes with information provided by all federal employes at the time they are hired. In neither case were respondents informed at the time they provided information that it would be linked with other records for the purposes of nabbing welfare cheaters.

<sup>14</sup>See Dunn, E. S., Jr. 1966. Testimony presented to Subcommittee of the House Committee on Government Operations. 89th Cong., 2nd session. *The Computer and Invasion of Privacy*, 92-95. Boruch, R. F. 1967. Maintaining confidentiality in educational research: A systemic analysis. *American Psychologist* 22: 308-11. Boruch, R. F. 1982. Methods for resolving privacy problems in social research. In *Ethical Issues in Social Science Research*. T. L. Beauchamp, R. R. Faden, R. J. Wallace, Jr., and L. Walters. eds. Baltimore: The Johns Hopkins University Press, 292-314.

<sup>15</sup>The examples of linked records are fully within the laws governing these data, according to Gates and Scheuren.

<sup>16</sup>A variety of steps are typically taken prior to the release of data for public use which are designed to protect the anonymity of responses. These practices include releasing data in which the minimal identifiable geographic unit is no fewer than 100,000 residents, combining detailed items into composites in which that detail is obscured, recoding detailed information about such items as age or family income into broad categories, withholding details about sample design, and removing identifying information about individuals (e.g., names, social security numbers) prior to the release of public use files.

<sup>17</sup>For a fuller discussion of the analytical costs of many procedural and statistical devices for protecting the anonymity of records, see Boruch and Cecil, 1979.