

PROCEDURES TO PROTECT PRIVACY AND MAINTAIN CONFIDENTIALITY¹

Joan E. Sieber

California State University, East Bay
jsieber@bay.csu Hayward.edu

The Common Rule emphasizes privacy and confidentiality but is neither specific nor detailed in its recommendations. There is a superb literature on approaches to respecting privacy and confidentiality with which researchers should be familiar. See annotated bibliography.

Standard Issues and Problems

The Common Rule leaves much to the judgment of IRBs with respect to privacy and confidentiality, so that the same degree of caution need not be imposed on all research. 49 CFR 46.111 (a)(1) states that "Risks to subjects are minimized: (i) By using procedures which are consistent with sound research design, and which do not unnecessarily expose subject to risk," and 45 CFR 46.111(a)(7) states that "When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data."

How do IRBs identify the relevant issues? Invasion of personal privacy is a subjective matter, and IRB members judge invasion of privacy based on their own sense of propriety and on the particular circumstances of the study. This can be an inadequate basis for judgment.

Confidentiality is an objective but complex matter and involves many possible judgments depending on the research. Confidentiality pertains to data on identifiable persons. In recent times, IRBs have become increasingly concerned about what constitutes an identifiable respondent of survey research. When is a survey truly anonymous? Even when the names of respondents are never attached to their data, there is increasing concern about deductive identification of otherwise anonymous respondents on the basis of such elements of their data as birth date, occupation, zip code, race, and gender. The issue of an "identifiable subject" also arises if the researcher wants access to existing records to identify persons suitable for the proposed study; if the data being sought are sensitive, the IRB may judge that consent of subjects should be obtained for accessing those existing data.

IRBs do not consider sample survey research anonymous (and hence exempt from IRB review) if identifiers that accompany the data are later stripped from the data. If a unique identifier was attached to it at some point in the process — for purposes of respondent selection and interviewing or for re-contacting selected subjects by a supervisor checking on the work of the interviewer — it is not an anonymous survey. A survey that involves any identifiable data at any point in the research process is subject to IRB review.

If researchers are collecting sensitive survey data about identified individuals, the IRB will inquire whether there are provisions for protecting the confidentiality of the data. Such provisions typically include substituting codes for identifiers and storing the code key elsewhere, removing face sheets (typically containing such information as names, phone numbers, or addresses), destruction of identifying information such as computer sheets, keeping data in locked files, impressing on research assistants the importance of confidentiality, and limiting access to the data by various means. Data from large surveys are normally stored electronically for easy access and analysis. Whenever identifiers accompany these data or when deductive identification would be easy, there is major concern about the security of the computer system on which the data are stored, and researchers must satisfy these concerns.

A particularly difficult issue has to do with the training and supervision of interviewers and research assistants. Where possible, the research staff should not be persons who might know some of the respondents, though this is difficult to ensure. Researchers who are concerned about the cultural and linguistic matching of interviewers with subject populations should consider hiring and training local people. However, there is the risk that the interviewer will know the respondent.

The Common Rule defines and discusses privacy and confidentiality in ways more appropriate to biomedical research than to survey research. It fails to recognize both the aspects of personal privacy that the effective interviewer must respect and the individual subjectivity and diversity of people's sense of privacy. With respect to confidentiality, it gives no hint of the vast technical literature on methods of ensuring confidentiality in survey research.

Privacy

Privacy is about persons and their sense of being in control of the access of others to themselves. In research, privacy typically refers to whether the subject considers it the researcher's business to delve into the

¹ An initial version of this paper appeared as Chapter 3 of a document prepared for, and supported by, the Office of Justice Statistics in June 2001. The entire OJS document, entitled "Summary of Human Subjects Protection Issues Related to Large Sample Surveys" may be found on-line at <http://www.ojp.usdoj.gov/bjs>

subject's life concerning whatever matter is the topic of the research. Whether a subject wishes to give a researcher access to such information depends on the background of the subject, who is sponsoring the research, the context in which the data are to be gathered, and such factors as whether the subject finds the researcher likeable.

Defining Privacy

The difficulty of defining invasion of one's own privacy is evocatively expressed by Melton (1992, p. 66):

'I know it when I feel it.' A gut sense of personal violation may be the tie that binds such disparate events as being subjected to a body search, being the subject of gossip, having one's mail read, being asked one's income, or having one's house entered without permission. It should come as no surprise that such an intensely personal construct is difficult to define.

It is difficult to define, understand, and respect the privacy of other persons situated differently from ourselves. Without a useful definition or theory of privacy to guide them, researchers and IRBs must depend on their own culture-bound notions of privacy. They invoke their personal and idiosyncratic definitions, resulting in a capricious standard of protection.

The meaning of privacy in survey research inheres in the culture and personal circumstances of the particular subject, the context and nature of the research, and the social and political environment in which the research occurs. A useful definition of privacy that recognizes these manifold elements is borrowed from, and based on, the elegant theory of personal privacy developed by Laufer and Wolfe (1977):

Privacy refers to persons and to their interest in controlling the access of others to themselves.

This theory of personal privacy recognizes the manifold cultural, developmental, and situational elements by which individuals orchestrate their privacy. It recognizes that people have an interest in (a) controlling the time, place, and nature of the information they give to others, and (b) controlling the information or experiences that are proffered to them. Thus informed consent serves as a control mechanism, provided the prospective subjects of survey research are adequately informed of what it is they will be asked and what they may experience.

Laufer and Wolfe's theory would be highly useful to efforts to educate IRBs; to design ethical elements of recruitment, consent, location, circumstances, and content of surveys; and to train interviewers. Laufer and Wolfe's theory embodies four dimensions of privacy:

The *self-ego dimension* refers to the development of autonomy and personal dignity. Young children have an aversion to being alone. By middle childhood, children seek time alone to establish a sense of self and to nurture new ideas, creating a basis for self-esteem, personal strength, and dignity. By age 6 or 7, children have a need and right to privacy not found in infants and younger children. Teenagers are intensely private, as they seek to forge an identity separate from that of their parents. Teenagers would be embarrassed to be interviewed about personal matters in the presence of their parents or others, and in the presence of their parents would most likely refuse to be interviewed or lie in their answers. Adults continue to need time alone and develop many means of protecting that privacy.

The *environmental dimension* includes socio-physical, cultural, and life-cycle dimensions. Socio-physical elements are physical or technological elements that offer privacy; more affluent individuals tend to have more of such barriers to unwanted intrusion. Cultural elements include norms for achieving privacy; for example some cultures permit lying while others permit persons to have private rooms and telephones. Life-cycle elements vary with age, occupation, available technology, and changing socio-cultural patterns. The ways one establishes privacy at one age, under one set of circumstances or level of maturity and sophistication, may be unsatisfactory or unavailable at a later stage. The ways of protecting privacy that are available to a prison staff member typically differ from those available to a prisoner.

The *interpersonal dimension* refers to how interaction and information are managed. One's social setting and its physical characteristics provide options for managing social interaction; physical and social boundaries can be used to control people's access to one another. Prisoners have less control over such boundaries than do non-institutionalized persons.

The *control/choice dimension* grows out of the other dimensions. Young children have no control over their privacy except through hiding. They learn to use personal, cultural, and physical resources to control their privacy. Events that would threaten one's privacy early in the development of control/choice are later so easy to control that they are no longer considered a threat to privacy. Prisoners may lack significant control over their physical environment, but may have developed verbal skills, including lying, that serve to protect their privacy.

Understanding and respecting the privacy of others

How do the researcher and IRB learn about the privacy interests of persons situated differently from themselves? Networks of local researchers, prison staff and other helping professionals can share valuable information about the most appropriate ways to approach members of various cultures. Interviewers who are of the

same culture and backgrounds are vital to some sensitive research in some kinds of populations. Focus groups and other forms of community consultation are useful ways to learn about a culture, how the individuals within that community perceive the research, and how the research that would be objectionable to them can be made acceptable. The community meetings held by Fisher and Wallace (2000) are a good example of learning the views and suspicions of members of inner-city African Americans about studies of adolescent risk behavior. The community consultation discussed in Melton et al. (1988) is a dramatic example of the explosive acquaintance process of AIDS activists with AZT researchers. While only the Fisher and Wallace account focuses on survey research, there is much that survey researchers could learn from both accounts about the importance of understanding the perspective of one's subjects in applied research.

Confidentiality

The following definition of confidentiality is adapted from that developed by Boruch and Cecil (1979):

Confidentiality is an extension of the concept of privacy; it refers to (a) identifiable data (some information about a person that would permit others to identify the specific person, such as a non-anonymous survey, notes or a videotape of the person) and (b) agreements about how those data are to be handled in keeping with respondents' interest in controlling the access of others to information about themselves.

The two critical elements of this definition — identifiable data and agreement about the handling of the data — indicate the critical role of informed consent, which states how the researcher will control access to the data and secures the respondent's agreement to participate under these conditions. This definition further underlines the importance of planning before gathering sensitive data. It is important that researchers make early plans regarding techniques to ensure confidentiality. They should incorporate these plans into the methodology and into any consent agreements with respondents or contractual agreements with subsequent users of the data, including funders who may wish to audit the data. Investigators should include all of these details in the IRB protocol.

This definition of confidentiality leads naturally to the literature on procedural, methodological, statistical, and legal approaches to ensuring the confidentiality of survey research data.

Methods and procedures of ensuring confidentiality

Approaches to ensuring confidentiality of survey research fall into six categories:

- Procedures that eliminate linkage of data to unique identifiers
- Intersystem linkage
- Statistical strategies
- Agreements: Data sharing, secondary analysis, or audit of data
- Legal protections
- Descriptive statistics and raw data releases

Procedures that eliminate linkage of data to unique identifiers

Anonymity offers the best insurance that disclosure of subjects' responses will not occur. Researchers have developed dozens of techniques that are responsive both to the need for anonymity and to other research needs. Different kinds of data — cross-sectional, longitudinal, and data from multiple sources — bring with them different research requirements and different ways of meeting those without using unique identifiers of subjects. The following brief summary is illustrative, not comprehensive. See Boruch and Cecil (1979) for a more comprehensive review.

Cross-sectional surveys, in their simplest form, require just one data collection session. Anonymity in which even the researcher is at all times ignorant of the identity of subjects protects the respondent from legal prosecution, social embarrassment, and concern that the data may fall into the wrong hands. However it may be desirable to have some form of follow-up to test for sampling validity, response validity, or to do further research on some or all subjects. These refinements are impossible with complete anonymity, but can be achieved through temporarily identified responses with subsequent destruction of identifiers, or through use of brokers to provide anonymous data to the researcher after completing one or more of these refinements.

Longitudinal surveys track individual subjects over time. There are many ways in which aliases or arbitrary identifiers can be used as a basis for linking observations over time while preserving the confidentiality of individual responses. For example, subjects may choose an easily remembered alias and use it on repeated occasions. Some approaches are quite complex. For example, research by the American Council on Education (Austin and Boruch, 1970) on political activism among American college students used the following three-file linkage system:

1. Initial data collection:
 - File A contains each subjects' data and arbitrary account number (X).
 - File B pairs each subject's name with a second arbitrary account number (Y). File C matches the two sets of account numbers, X and Y.
 - File C is shipped to a researcher in a foreign country.
2. Second data collection:
 - Second set of identifiable longitudinal data are gathered.
 - Names are replaced by their Y account number; this file is shipped to the foreign researcher.
3. Data analysis:
 - Foreign researcher substitutes X account numbers with corresponding Y numbers.
 - Each set of data files is returned to the data analysts.
 - Data are organized in longitudinal sequences; the identity of each subject is unknown.
 - The longitudinal data are analyzed.
 - Foreign researcher destroys File C; the three files can never be merged to learn subject identities.

Thus the data were safe. Conceivably, foreign discovery procedures could be used to obtain some of the identifiable data before File C is destroyed. A Certificate of Confidentiality (see below) could be obtained to preclude that unlikely event.

Intersystem linkage

Intersystem linkage is sometimes necessary to link research records on subjects with other, independently stored records on the same individuals. In the case of highly sensitive data such as psychiatric or police records, a linkage strategy may be needed so that the researcher does not have access to any identified records. One such method is as follows:

1. Researcher wishes to link data on 50 subjects with information from their police records.
2. Subject provides data and an alias (no name) to the researcher.
3. Subject provides to the archive (police) his name and alias.
4. Archive provides the requested police information with the aliases (not the names) attached
5. Researcher analyzes relationship between his research data and the police record data.

This brief summary is merely illustrative of some of the many specific procedures for preserving anonymity or confidentiality. The actual literature on this topic is immense. (See Boruch and Cecil (1979), and Campbell, Boruch, Schwartz, and Steinberg (1977). See also current U.S. Census Bureau web-site papers on this topic.)

Statistical strategies

Researchers have developed various statistical strategies to eliminate any direct link between the respondent's identity and his true answer. Many of these methods involve the injection of a specified amount of random error into the dataset so that researchers cannot ascertain an individual's true identity but can still perform a useful statistical analysis of the data. Since these strategies are many and continually being developed, the best way to learn what is currently available is to visit the website of the Committee on Privacy and Confidentiality of the American Statistical Association. That website may be found at www.amstat.org/comm/cmtepc .

Legal protections of confidentiality

Statutory protection of research data enables researchers to protect the confidentiality of research records on identifiable individuals from subpoena. Subpoena of social research data is rare. However if vulnerable data could not be protected from subpoena, there would be a chilling effect, especially on criminological and delinquency research. There is growing use of Certificates of Confidentiality. However researchers and IRBs are often unclear about the protections these provide and their limitations.

Certificates of Confidentiality. The Public Health Service Act (PHSA) was amended (1970) to authorize researchers to withhold information concerning the identity of participants in research on use and effect of drugs. The Secretary of the Department of Health and Human Services grants this authority by issuing Certificates of Confidentiality. A 1988 amendment broadened its scope to include mental health, biomedical, clinical, behavioral, and

social research. Under this amendment, the Secretary of DHHS may

authorize persons engaged in biomedical, behavioral, clinical, or other research (including research on mental health, including research on the use and effect of alcohol and other psychoactive drugs), to protect the privacy of individuals who are the subject of such research by withholding from all persons not connected with the conduct of such research the names or other identifying characteristics of such individuals. Persons so authorized to protect the privacy of such individuals may not be compelled in any federal, state, or local civil, criminal, administrative, legislative, or other proceedings to identify such individuals. (42 U.S.C. 242a(b)(1989))

Various institutes within DHHS are authorized to issue certificates. Since 1993 DHHS can grant certificates for research that is not federally funded. DHHS regards a certificate's protection to supercede State law; this position has been challenged and upheld in the New York Court of Appeals (*People v. Newman*, 32 N.Y.2d 379, 298 N.E.2d 651, 345 N.Y.S.2d 502, 1973) (Boikess, 2000).

A certificate does not protect identifiable data about other people that the subject may disclose, a point which researchers may fail to clarify in the informed consent. The certificate only protects researchers against compelled disclosure of subjects' names or other identifiers, coupled with their data. It does not protect subjects who voluntarily consent to disclose their research records, nor preclude a researcher from reporting the identity of subjects who disclose intentions to harm themselves or others. Moreover the language of PHSA is imprecise, which gives rise to uncertainty. It offers protection to "names and other identifying characteristics," but the data of a known subject may not necessarily be protected. Melton (1992, p. 81) provides an example of this possible loophole:

[I]n one of my own studies, all of the children in a particular county who are involved in criminal child abuse prosecutions are invited to participate. Knowing that fact, a defense attorney might seek the data of a particular child (not the names of participants) as a fishing expedition for information intended to impeach the child's testimony. A literal interpretation of the statute would suggest that the subpoena might be enforceable if the data could be shown in some way to be relevant to the proceeding. Although it is also possible — perhaps even probable — that a court would interpret the statute more broadly in keeping with Congressional intent, the uncertainty prevents unequivocal offers of confidentiality to participants and, therefore, should be eliminated by a technical amendment.

It is also unclear whether child abuse reporting laws are abrogated by a Certificate of Confidentiality. Is such reporting a "legal proceeding" that cannot be mandated under a certificate?

Researchers must request the certificate before each research undertaking. Subpoenas typically occur for reasons unrelated to the study itself and therefore are not reasonably foreseeable by either the subjects or the investigator. Hence the protections offered by a certificate may be unavailable when needed. Researchers sometimes send data to a foreign country, although this does not always guarantee protection.

Placing data in a foreign country and laws governing foreign discovery.²

Many survey researchers believe that sending confidential data to a foreign country — such as to a colleague in Canada — protects the data from subpoena. However this is only a deterrent from subpoena, not a guarantee of protection. Federal Rules of Civil Procedure govern the procedures for discovery, including foreign discovery. Rule 26(b) states that parties may obtain discovery of anything that is relevant, not privileged, and admissible or "reasonably calculated to lead to the discovery of admissible evidence." Rule 34 states:

(a) *Scope. Any party may serve on any other party a request (1) to produce and permit the party making the request, (2) to inspect and copy, any designated documents, or (3) to inspect and copy, test or sample any tangible things which constitute or contain matters within the scope of rule 26(b) and which are in the possession, custody, or control of the party upon whom the request is served.*

(c) *Persons Not Parties. A person not a party to the action may be compelled to produce documents and things or to submit to an inspection.*

The courts cannot compel a party to produce data if the party does not have "possession, custody, or control" of the documents, but it is unclear what constitutes "control." If a researcher sends data out of the country for the express purpose of preventing subpoena, does this qualify as loss of control in the eyes of a court? Jason Gilbert (2000), a legal intern at the Federal Judicial Center, offers the following analysis of this question:

While the courts seem to have settled on defining control as when a party has a legal right to obtain something, questions remain for the researcher seeking to give up control of research data to a foreign colleague in an attempt to protect it from being disclosed. Legal rights to possession can come from a variety

²³I am indebted to Dr. Joe Cecil and Jason Gilbert, Federal Judicial Center, for providing me with their detailed summary and analysis of these issues.

of sources, particularly when one is considering intellectual property such as research data. If a researcher were to create a set of data, when exactly would he or she no longer have a legal right to that set of data? What if the researcher gave one part of the data to a colleague? What if the researcher only gave up a small "key" to the data that allowed the individuals who participated in the study to be identified? What if the researcher gave part, or even all, of the data to a colleague but still continued to collaborate with that colleague to perform analysis on the data even though it was not in the researcher's possession? Would that researcher still have a legal right to get back what he or she had surrendered? While the concept of giving away the legal right of possession is relatively straightforward, the mechanics of how exactly a researcher can give away the legal right to possess his own data (particularly if one does not allow for a sale or some type of contract) remains unclear.

Gilbert also reminds us of some other implications of "loss of control" of data. (1) Transfer of all data out of the country would mean loss of all electronic or hard copies in the researcher's possession. (2) A researcher must never transfer data after receiving a subpoena. Even if it is done as a safeguard beforehand, the researcher may still be found to have acted not in good faith and be cited for contempt of court. (3) If the research is done under a contract requiring that the researcher maintain control of the data, relinquishing control to a foreign colleague would constitute a breach of that contract. (4) The researcher's professional code of ethics or a future journal editor may require that the researcher maintain control of the data.

A researcher who loses control of data by sending it to a foreign colleague places that colleague at risk of receiving a subpoena for the data and of having to seek legal means of protecting confidentiality. However the rules and procedures of foreign discovery are complex, expensive, and time-consuming. If the colleague who controls the subpoenaed information is a foreign national residing outside of the United States, the party seeking the data must follow appropriate procedures for foreign production. The United States has ratified various treaties concerning obtaining of evidence from foreign countries, and each country has its own procedures. Discovery in a foreign country involves sending a formal "letter of request" by the court where the action is pending to a court in the foreign country. This letter requests the foreign court to request documents of the person in possession of the desired information. There are various diplomatic and legal approaches to delivering such a request and accomplishing the discovery. These may make discovery of the information too unattractive to pursue.

Descriptive statistics and raw data releases

Statisticians in governmental agencies in the United States, Great Britain, and Sweden have developed practices of adjusting tabular presentations so that deductive identification is not possible. Deductive identification could occur if one knew some facts about an individual, perhaps in conjunction with their zip code. By searching the files for that zip code and locating the individual whose data matched those known facts one could deduce additional information from the other data associated with that individual. For example, if an 84-year-old Hispanic woman from the 01373 zip code area was known to have had quite a few husbands and to have become quite wealthy, it would confirm some suspicions to learn that there was an 84-year-old Hispanic woman at that zip code whose annual interest income was in seven figures and who had had 20 husbands.

The most common way to prevent deductive disclosure is to broaden categories so that data from unique individuals or from groups of data containing some unique individuals such as top income earners or persons who have been involved in high-profile criminal activity are not apparent. Another method is error inoculation so that no individual case could be assumed to be correct (as in the random-response method). For example, when the U.S. Census Bureau's Disclosure Review Board reviews datasets in anticipation of their release to the Inter-University Consortium for Political and Social Research (ICPSR), the Review Board's primary role is to scrutinize data for extreme cases that might uniquely identify the individuals involved or even subject them to criminal investigation. Thus, instead of showing that a given 84-year-old Hispanic woman had been widowed 20 times and had an interest income of over \$1,000,000, it might show that the person had been widowed more than four times and had an annual interest income of over \$100,000.

Confidentiality assurances and their consequences

The link between promises of confidentiality and willingness to participate in surveys is tenuous. Researchers' promises of confidentiality do not always produce trust in research participants. Such promises cannot always be kept due to faulty data management practices and other possible compulsory disclosures. Moreover, the relationship between faith in confidentiality promises and participation in survey research is not what most suppose it to be.

Singer, Mathiowetz, and Couper (1993) investigated the relationship between concerns about confidentiality and mail returns to the 1990 census. Such concerns have only a very slight effect on survey participation, and this relationship holds even when demographic variables known to be related to concerns and survey participation are controlled. Similarly assurances of confidentiality have unexpected effects. Singer, VonThurn, and Miller (1995) conducted meta-analysis of 30 research reports on the relationship between various forms of confidentiality assurances (anonymity, use of the randomized response method, and verbal assurances). They found that **the effect**

of confidentiality assurances on willingness to respond is small, positive, statistically significant, and robust in the presence of various control variables, but only when sensitive questions are asked. The effect is small and negative when the questions asked are not sensitive ones. Elaborate assurances of confidentiality defeat their purpose when the contents of the survey are not sensitive. Apparently such assurances of confidentiality heighten respondents' perceptions of the sensitivity or threat of the survey or arouse their suspicions (Singer, Hippler, and Schwarz (1992)).

Ethical Proofreading in Qualitative Research

It is sometimes possible to deduce personal information about identifiable individuals from qualitative data, such as a cultural anthropologist's "anonymous" account of life in community that is described, but in which persons and places are given fictitious names (Johnson, 1982). The same kinds of problems will probably arise as more studies are conducted in virtual communities of participants in on-line chat rooms (King, 1999). Any clue such as a description of the research site (e.g., a map or a web address) might permit deductive disclosure by anyone familiar with that territory. Johnson (1982) reviewed cases of well-known publications in cultural anthropology in which the identities of specific rural community members could be deduced. In some of these cases, anthropologists had written detailed accounts about the secret, illegal, immoral or reprehensible deeds of these people, with no awareness that the actual identities would be discovered.

Those who do qualitative studies of the lives of others cannot ensure confidentiality; the subjects themselves, the research assistants, or even the sponsor of the research may inadvertently leak the identity of the research site. Since total confidentiality or anonymity cannot be guaranteed, the issue becomes one of on-going communication and agreement with subjects (informed consent), and respectful communication of the findings. There is, by now, a growing literature on this issue (e.g., Cassell, 1982; Gallaher, 1964; Glazer, 1982; Johnson, 1982; Kelman, 1968; King, 1999; Wax, 1982).

For example, Johnson recommends guidelines for "ethical proofreading" of cultural anthropology manuscripts to diminish potential harm to subjects or communities as follows:

- Assume that the identities of the location and individuals studied will be discovered. What would be the consequences within and outside the community? Will its effect on individuals and relationships be positive or negative? Does the importance of the material warrant any risk of harm?
- Look at the language. Is it descriptive or judgmental? For example, "Twenty percent of the adults are functionally literate." is less judgmental than. "Most of the people are backward."
- When describing private or unflattering characteristics, first describe the cultural context then describe the specific characteristic. This is more informative and does not single out individuals as much.
- Negative stereotypes may affect similar other people and communities even if the specific people and communities are not identified. Ask yourself how the information might be used in a positive way? In a negative way? Are revelations worth the possible risk?
- Decide whether the research site will be usable again, or will it have been destroyed if the residents read what has been written about them?
- Have some of the subjects proofread the manuscript for accuracy and invite them to provide any general feedback they are inclined to offer. Have some colleagues (ideally ones who are skeptical or not totally supportive of your research) proofread the manuscript using the above guidelines as criteria for acceptability.

Emerging issues of privacy and confidentiality

Electronic data collection and storage practices are changing rapidly. Interviewers and their subjects need not meet face-to-face and may even reside in different parts of the world. Emerging issues of confidentiality are more varied and dangerous than current policy makers can easily anticipate. Soon issues of confidentiality will be transformed in ways we cannot imagine today. There are already digital communication networks on a global scale, and hackers with a laptop computer and Internet technology could download any electronic data stored on any server anywhere in the world. There are also emerging technologies for protecting communication and personal identity, and there is a whole new cohort of technology-sophisticated privacy activists. Governments are developing and testing laws that protect data, and globalization of culture and policy processes is occurring. The American Association for the Advancement of Science, the National Science Foundation, and various Internet research groups are actively exploring these issues. Scientific societies concerned with the protection of social and behavioral research data — the American Statistical Association, American Psychological Association, and American Sociological Association — will continue discussing these issues at their annual meetings and on their web sites for years to come.

BRIEFLY ANNOTATED BIBLIOGRAPHY OF LITERATURE CITED AND RELATED OTHER MATERIAL

- American Psychological Association, *Ethical Principles of Psychologists and Code of Conduct* (Washington, D.C.: APA, 1992).
Most of this code pertains to clinical practice., Sections 6.06 to 6.26 concern research and are heavily relied upon by psychologists and others concerned with ethics of human research.
- A. W. Austin and R. F. Boruch, "A 'Link File System' for Assuring Confidentiality of Research Data in Longitudinal Studies," *American Educational Research Journal* 7, 615-624 (1970).
This article describes highly sophisticated techniques for linking files without the possibility of discovery of individual identities.
- P. J. Barker and R. F. Cooper, "Do Sexual Health Questions Alter the Public's Response to Lifestyle Questionnaires?" *Journal of Epidemiology and Community Health* 50, 688 (1996).
The finding that just one sensitive question did not affect response rate leaves one to wonder what is the threshold for making a survey too sensitive, and hence for affecting response rate.
- David M. Bersoff and Donald M. Bersoff, "Ethical Issues in the Collection of Self-Report Data," in A. A. Stone, J. S. Turkkan, et al., eds., *The Science of Self-Report: Implications for Research and Practice* (Mahwah, N.J.: Lawrence Erlbaum Associates, 2000), pp. 9-24.
The risks connected with self-report surveys are discussed, pointing out that surveys can ask for far more personal information than is available by observation or experiment. An anonymous survey, requiring no IRB review or informed consent, may deeply invade privacy and cause much emotional upset by probing into painful experiences.
- O. Boikess, *Certificates of Confidentiality*, personal communication, 2000.
This communication sets forth details of a legal case, heard in the Supreme Court of New York, in which the power of the Certificate of Confidentiality was tested and upheld.
- R. F. Boruch and J. S. Cecil, *Assuring the Confidentiality of Social Research Data* (Philadelphia: University of Pennsylvania Press, 1979).
*This book, now out of print, might be regarded as the most important publication on approaches to assuring the confidentiality of social research data. It catalogs and describes the many kinds of approaches to assuring confidentiality, such as procedural, statistical, and legal. Though a few parts of it would be rewritten in light of new information, it remains a seminal work.
research and to improve methods and theory in this area.*
- D. T. Campbell, R. F. Boruch, R. D. Schwartz, & J. Steinberg, "Confidentiality-Preserving Modes of Access to Files and to Interfile Exchange for Useful Statistical Analysis," *Evaluation Quarterly* 1, 269-300 (1977).
The authors present many techniques that enable useful data sharing and secondary analysis without jeopardizing confidentiality of data.
- Cassell, J. (1982). Harms, benefits, wrongs, and rights in fieldwork. In J. Sieber (Ed.), *The ethics of social research: Fieldwork, regulation and publication*. New York: Springer Verlag. Pp. 7-31.
The nature of harms, benefits, etc. in fieldwork differs from that of most social and behavioral research. These differences are explored.
- A. H. Church, "Estimating the Effects of Incentives on Mail Survey Response Rates: A Meta Analysis," *Public Opinion Quarterly* 57, 62-79 (1993).
This is a useful empirical study of the effects of financial incentives on willingness to participate in survey research.
- R. B. Cialdini, *Influence: The Psychology of Persuasion* (New York: Morrow, 1993).
This important book helps us to better understand subtle forces of persuasion that are at work in survey research and other contexts.
- W. W. Darrow, H. W. Jaffe, P. A. Thomas, H. W. Haverkos, M. F. Rogers, M. E. Guinan, D. M. Auerbach, T. J. Spria, and J. W. Curran, "Sex of Interviewer, Place of Interview, and Responses of Homosexual Men to Sensitive Questions," *Archives of Sexual Behavior* 15, 79-88 (1986).

Their examination of a variety of physician-patient interviews revealed virtually no differences as a result of gender of interviewer and location of interview.

J. L. Freedman and S. C. Fraser, "Compliance Without Pressure: The Foot-in-the-door Technique," *Journal of Personality and Social Psychology* 4, 195-202 (1966).

This seminal study of the way in which persons could be gotten to accede to unreasonable requests has triggered a variety of studies that have informed survey researchers about ways to approach sensitive topics.

Gallaher, A. (1964). Plainville: The twice-studied town. In A. Vidich, J. Bensman & N. Stein (Eds.), *Reflections on community studies*. New York: Wiley, 1964.

The errors and harms involved in these studies are reviewed.

Glazer, M. (1982). The threat of the stranger: Vulnerability, reciprocity and fieldwork. In J. Sieber (Ed.), *The ethics of social research*. Vol 2. New York: Springer Verlag. Pp. 49-70.

The reactions and vulnerabilities of naïve respondents in fieldwork are considered in relation to the ethical obligations of the researcher.

J. Gilbert, *Protecting Confidential Research Information from Disclosure in Litigation*, personal communication, Federal Judicial Center Memo, July 7, 2000.

In this memo, legal intern Gilbert sets forth the legal limits of protection to subpoenaed data; of particular interest is his discussion of the methods of legal discovery that can be used to obtain data that have been sent outside of the country.

Johnson, C. (1982). Risks in the publication of fieldwork. In J. Sieber (Ed.), *The ethics of social research: Fieldwork, regulation and publication*. New York: Springer Verlag. Pp. 71-92.

Rules for "ethical proofreading" of ethnographic reports are set forth.

Kelman, H. (1968). *A time to speak*. San Francisco: Jossey-Bass, 1968.

Humanistic concerns need to be considered in designing social research.

King, S. (1997). Researching internet communities: Proposed ethical guidelines for the reporting of results. *The Information Society*, 12:119-127.

What is public? What is private? What is decent when studying sensitive behavior in public internet venues?

P. Kissinger, J. Rice, T. Farley, S. Trim, K. Jewitt, V. Mangavio, and D. H. Martin, "Application of Computer-Assisted Interviews to Sexual Behavior Research," *American Journal of Epidemiology* 149, 950-954 (1999).

The authors describe video-enhanced computer-assisted, self-administered interviews, which were shown to elicit more socially undesirable responses than face-to-face interviews.

J. A. Landsheer, P. G. van der Heijden, and G. van Gils, "Trust and Understanding, Two Psychological Aspects of Randomized Response," *Quality and Quantity* 33, 1-12 (1999).

This research established the importance of trust and understanding of the method in producing valid responses on socially undesirable topics.

R. S. Laufer and M. Wolfe, "Privacy as a Concept and a Social Issue: A Multidimensional Developmental Theory," *Journal of Social Issues* 33, 44-87 (1977).

Laufer and Wolfe's theory of personal privacy recognizes the manifold cultural, developmental, and situational elements by which individuals orchestrate their privacy.

J. T. Lesser and J. M. O'Reilly, "Mode of Interview and Reporting of Sensitive Issues: Design and Implementation of Audio Computer-assisted Self-interviewing," *NIDA Research Monograph* 167, 366-382 (1997).

The authors review various approaches to interviewing people about sensitive issues. They include the implementation and results of using audio-computer-assisted self-interviewing.

L. E. Linden and D. J. Weiss, "An Empirical Assessment of the Random Response Method of Sensitive Data Collection," *Journal of Social Behavior and Personality* 9, 823-836 (1994).

The authors studied sensitive aspects of personal history in 285 undergraduates using written questionnaires. They found no difference between direct and randomized response methods of questioning.

P. A. Marshall, "Research Ethics in Applied Anthropology," *IRB: A Review of Human Subjects Research* 14, November-December, pp. 1-5 (1992).

Various problems arise when researchers apply culturally insensitive approaches to non- Western or non-mainstream cultures, and when researchers apply federal regulations designed for mainstream American culture elsewhere. This article describes the resulting dilemmas.

T. Makkai and I. McAllister, "Measuring Social Indicators in Opinion Surveys: A Method to Improve Accuracy on Sensitive Questions," *Social Indicators Research* 27, 169-186 (1992).

The authors compare sealed booklet technique with the face-to-face interview in terms of respondent's willingness to answer sensitive questions.

G. B. Melton, "Respecting Boundaries: Minors, Privacy and Behavioral Research," in B. Stanley and J. Sieber, eds., *The Ethics of Research on Children and Adolescents* (Newbury Park: Sage, 1992), pp. 65-87.

The problems of protecting privacy and assuring confidentiality of behavioral research on minors are discussed. The stages of development of minor's sense of privacy are illustrated. Melton discusses uses and limits of Certificates of Confidentiality.

G. B. Melton, R. J. Levine, G. P. Koocher, R. Rosenthal, and W. Thompson, "Community Consultation in Socially Sensitive Research: Lessons from Clinical Trials on Treatments for AIDS," *American Psychologist* 43, 573-581 (1988).

This classic article describes the process and benefits of consultation between members of vulnerable communities and researchers whose intentions are regarded with suspicion, anger, and plans for non-cooperation by those they seek to study.

P. Raghurir and G. Menon, "Asking Sensitive Questions: The Effects of Type of Referent and Frequency Wording in Counter-biasing Methods," *Psychology & Marketing* 13, 633-652 (1996).

The authors conducted two experiments that tested the effects of counter-biasing methods on respondent's willingness to admit to stigmatized behaviors.

J. E. Sieber, ed., *Sharing Social Science Data: Advantages and Challenges* (Newbury Park: Sage, 1991).

This edited volume discusses a range of data sharing arrangements and the attendant challenges of protecting identities and providing adequate documentation. It includes descriptions of a variety of innovative sharing relationships.

J. E. Sieber, "Typically Unexamined Communication Processes in Research," in B. H. Stanley, J. E. Sieber, and G. B. Melton, eds. *Research Ethics: A Psychological Approach* (Lincoln: University of Nebraska Press, 1996).

This article specifies a range of meta-communication processes that occur in research and that may be deserving of research in relation to ethical and methodological questions.

J. E. Sieber and M. Saks, "A Census of Subject Pool Characteristics and Policies," *American Psychologist* 44, 1051-1063 (1989).

This is an illustration of an experiment within a survey that examined the effects of anonymity on response rate and frequency of acknowledging unethical or illegal practices.

E. Singer, "Informed Consent: Consequences for Response Rate and Response Quality in Social Surveys," *American Sociological Review* 43, 144-162 (1978).

The author investigated the effects in face-to-face interviews of more versus less information about sensitive subject matter, of varied assurances of confidentiality, and of requiring or not requiring a signature to document consent. Varying information made no difference, but was noticed if it failed to mention pertinent elements of the study. Subjects were inaccurate in their perceptions of how much confidentiality they had been promised. However perception that one had been given absolute assurance of confidentiality was associated with higher estimates of sensitive behavior. Subjects assigned to a condition in which they were asked to sign a consent form significantly more often refused to do so but were willing to participate in the interview if they didn't have to sign the consent form.

E. Singer, "Informed Consent in Surveys: A Review of the Empirical Literature," *Journal of Official Statistics*, 9, 361-375, (1993).

This paper examines the effects of four elements of consent: the content of the interview and purpose of the research, assurances of confidentiality or anonymity; active versus passive consent, and voluntariness of participation. Of particular interest was the finding that confidentiality promises have a small effect on willingness to participate in survey research. Highly detailed promises are counterproductive when the research questions are not sensitive; presumably such detail raises unwarranted concerns and suspicions.

E. Singer, H. Hippler, and N. Schwarz, "Confidentiality Assurances in Surveys: Reassurance or Threat?" *International*

Journal of Public Opinion Research 4, 256-268 (1992).

When non-threatening questions are accompanied by elaborate assurances of confidentiality, response rate is lowered.

E. Singer, N. Mathiowetz, and M. P. Couper, "The Impact of Privacy and Confidentiality Concerns on Survey Participation: The Case of the 1990 U.S. Census," *Public Opinion Quarterly* 57, 465-482 (1993).

The impact of privacy and confidentiality concerns are only slightly related to survey response.

E. Singer, D. vonThurn, and E. Miller, "Confidentiality Assurances and Response: A Quantitative Review of the Experimental Literature," *Public Opinion Quarterly* 59, 66-77 (1995).

Meta-analysis of 30 studies of the effect of confidentiality assurances (verbal assurances, anonymity, or use of randomized response method) showed that assurances of all three kinds are positively associated with response rate and response quality, but only when sensitive questions are asked.

T. W. Smith, "The Impact of the Presence of Others on a Respondent's Answers to Questions," *International Journal of Public Opinion Research* 9, 33-47 (1997).

This study examines the effects of the presence of spouses and children during interviews to determine whether different responses are evoked under these conditions.

A. L. Stanton, E. J. Burkner, and D. Kershaw, "Effects of Researcher Follow-up of Distressed Subjects: Tradeoff Between Validity and Ethical Responsibility?," *Ethics & Behavior* 1, 105- 112 (1991).

This excellent experiment provides strong evidence for the value of providing respondents with useful information and referrals, and for the importance of not seeking to intervene with help that was not requested.

R. Thompson, "Developmental Changes in Research Risk and Benefit: A Changing Calculus of Concerns," in B. Stanley and J. Sieber, eds., *Social Research on Children and Adolescents: Ethical Issues* (Newbury Park: Sage, 1982), pp. 31-64.

In contrast to the usual view that minors become less vulnerable as they grow older, this chapter presents a description of kinds of vulnerabilities that arise through the developmental sequence from early childhood to adulthood, showing that there are forms of vulnerability that increase with age through adolescence.

R. Tourangeau and T. W. Smith, "Asking Sensitive Questions: The Impact of Data Collection, Mode of Question Format, and Question Context," *Public Opinion Quarterly* 80, 275-304 (1996).

This complex study examined the effects of three kinds of computer-assisted interviewing, open- versus closed-ended answer options, and contextual variables.

A. D. Trice, "Informed Consent. VII. Biasing of Sensitive Self-report Data by Both Consent and Information," *Journal of Social Behavior and Personality* 2, 369-374 (1987).

The author found that a significant number of subjects refused to participate if required to sign a consent form but were willing to participate otherwise.

A. G. Turner, "What Subjects of Survey Research Believe about Confidentiality," in J. E. Sieber, ed., *The Ethics of Social Research: Surveys and Experiments* (New York: Springer Verlag, 1982).

Turner summarizes a series of high-level studies of what American populations, especially minority populations, believe about survey research. Many see surveys as pointless since they believe the government already knows everything about them. They do not trust promises of confidentiality. When individuals find themselves in a position where they cannot refuse to participate, they may be very careful to say only those things that they believe will not come back to haunt them, and are careful to tell each survey taker the same story.

U. N. Umesh and R. A. Peterson, "A Critical Evaluation of the Randomized Response Method: Applications, Validation, and Research Agenda," *Sociological Methods & Research* 20, 104-118 (1991).

This article reviews recent applications of the randomized response method, identifies new issues being studied, and suggests future research.

P. G. van der Heijden, G. van Gils, J. Bouts, and J. J. Hox, "A Comparison of Randomized Response, Computer-assisted Self Interview, and Face-to-Face Direct Questioning: Eliciting Sensitive Information in the Context of Welfare and Unemployment Benefit," *Sociological Methods & Research* 28, 505-537 (2000).

In this study, the respondents and interviewers did not know that the researchers knew about their welfare fraud. Respondents in the randomized response condition admitted more acts of fraud than either the face-to-face interviewed respondents or the computer-assisted self-interviewed respondents.