

SELECTED GUIDELINES FOR NOTIFICATION TO SURVEY PARTICIPANTS

Thomas B. Jabine, National Academy of Sciences

I. PROBLEM STATEMENT

There is a tendency in sessions like this one to discuss issues in broad and dogmatic terms. I believe it is more useful to begin by addressing specific, concrete issues: perhaps by doing so we can eventually develop some general principles for dealing with the difficult ethical questions that are the subject of this panel discussion.

The question I will address is: What information should be given to participants in voluntary surveys conducted by the Federal government? To be even more specific, I will focus on what participants are told about two specific subjects: (1) planned linkages of the information they provide with records from other data systems; and (2) the risk that information about them, as identifiable individuals, will be disclosed as a result of their participation in the survey.

These two topics--linkages and disclosure--would be relatively easy to handle if linkages of survey data with other records were prohibited and if no public-use microdata files were released. In my opinion, such a policy would seriously damage our ability to make intelligent judgements about important public questions such as the effect of government programs on the distribution of income and wealth and the risks associated with exposure to environmental health hazards. I proceed under the assumption that record linkages and release of public-use microdata files will (or should) continue to be important features of Federal surveys.

I will use the term notification for the process of informing survey participants about the implications of their inclusion in a survey. Although more formal informed consent procedures are occasionally used in surveys, this is not a common practice, nor, in my view, should it be. I have deliberately used the term "participants" rather than respondents to refer to persons for whom data are collected in surveys. Typically, in household surveys, the notification process is aimed at respondents, i.e., the persons who actually provide the information. However, for household surveys care should be exercised to ensure that all adult household members are included in this process.

Both the content and the method of notification are important. Oral notification by the interviewer at the time of interview is the traditional approach. A written notification statement, mailed in advance, or provided at the time of the interview, may supplement or take the place of oral notification. The standard notification process can also be supplemented by making explicit provisions to answer additional questions that some survey participants may raise.

Another aspect of notification is what to do about uses of the survey data that were not anticipated when the data were collected. Suppose linkage of survey records with another data set or release of a public-use microdata file is decided on subsequent to the collection of data.

Should survey participants be notified? Should they be given an opportunity to exclude their data from the new use?

II. A REVIEW OF STATUTORY REQUIREMENTS AND NON-BINDING RECOMMENDATIONS

One might ask why this particular subject needs to be addressed now. Aren't the notification procedures for Federal surveys pretty well fixed by statutory requirements and by long years of experience? Aren't the current procedures satisfactory?

With respect to statutory requirements, the relevant legislation (for surveys of persons) is the Privacy Act of 1974. Section (e)(3) requires, in effect, that respondents to Federal surveys of persons be notified of:

- the authority for the survey;
- whether it is mandatory or voluntary;
- the principal purposes for which the information is intended to be used;
- "routine" uses which may be made of the information (i.e., uses which are compatible with the purposes for which the information was collected); and
- the effect on the respondent of not providing some or all of the requested information.

None of these provisions explicitly requires notification of all planned linkages or the risk of disclosure. If a planned linkage required disclosure of identifiable data to another agency and that disclosure were to be made under the "routine use" provision of the Privacy Act, respondents would have to be so informed. However, linkages can be made in many ways that do not require exercise of the "routine use" provision.

Non-binding guidelines with some relevance to these two notification issues have been issued by the ASA Ad Hoc Committee on Privacy and Confidentiality, the ASA Ad Hoc Committee on Professional Ethics and the International Statistical Institute. The ASA Ad Hoc Committee (1977) did not recommend explicit notification of planned linkages. It said:

"In informing respondents of the uses of the data, it is sufficient to state that the data will be used for statistical purposes only, if such is the case. It is neither feasible nor necessary to spell out the possibly manifold ways, some of which may not be known in advance, in which the statistics may be employed."

Three members of the committee, including the chair, who is on this panel, disagreed with this

recommendation. They argued, in effect, that all participants in voluntary surveys should have explicit control over any linkages, either by refusing to provide data for the survey or denying a subsequent request for permission to link their survey data with other records.

For surveys from which public-use microdata files are to be released, the ASA Ad Hoc Committee suggested, by way of illustration, a statement of purpose such as:

"The data will be used only for statistical purposes, in which individual reports will not be identifiable." (underscoring added)

Recent statements by persons who have carefully examined statistical disclosure risks suggest that such absolute statements may not be justified. Cox *et al.* (1985) state that zero risk of disclosure is unattainable. Paass (1985) has suggested (but not proven) that some Census Bureau microdata files may be vulnerable to efforts by "attackers" to identify individuals.

The Ethical Guidelines for Statistical Practice drafted by the ASA Ad Hoc Committee on Professional Ethics (1983) include elements that are relevant to these two issues. With respect to notification of planned linkages, the guidelines state that statisticians should

"inform each potential respondent about the general nature and sponsorship of the inquiry and the intended uses of the data."

and

"ensure that, whenever data are transferred to other persons or organizations, this transfer is in conformity with the confidentiality pledges established."

These provisions are essentially the same as those recommended by the Ad Hoc Committee on Privacy and Confidentiality.

A provision of the guidelines that is relevant to the release of microdata files is that statisticians should:

"ensure that the means are adequate to protect the confidentiality to the extent pledged or intended . . . [and] that appropriate techniques are applied to control statistical disclosure."

The use of the word "control" in this provision, rather than "eliminate" or "avoid" may have been deliberate, in recognition of the residual disclosure risks referred to above.

The Declaration on Professional Ethics of the International Statistical Institute (1986) deals at length with notification issues, using the term "informed consent." The ISI Declaration is intended "to be informative and descriptive rather than authoritarian and prescriptive." The declaration states that "No universal rules [for informed consent] can be framed," but recommends that:

"Statistical inquiries involving the active participation of human subjects should be based as far as practicable on their freely given informed consent."

and

"Information that would be likely to affect a subject's willingness to participate should not be deliberately withheld."

The relevant section of the declaration includes a list of 12 items deemed "likely to be material to a subject's willingness to participate;" this list makes no mention of planned record linkages.

A provision of the declaration that is relevant to the release of microdata states that:

"Statisticians should take appropriate measures to prevent their data from being released in a form that would allow any subject's identity to be disclosed or inferred."

The discussion following this recommendation makes it clear that the framers of the declaration are aware of the disclosure risks associated with the release of anonymous records. They call for the use of technical disclosure avoidance procedures "to counteract the opportunities for others to infer identities from their data."

In summary, existing U.S. statutory requirements are not very explicit concerning the two notification issues that I have raised. Non-binding recommendations by statistical organizations provide somewhat more guidance. In general, they do not call for explicit notification of planned record linkages. None of these sets of recommendations explicitly recognizes that zero disclosure risk in the release of microdata is unattainable; hence they do not directly address the question of what to tell survey participants about such risks.

III. THE CURRENT SITUATION

The treatment of planned linkages in notification statements has had to be reconsidered in recent years primarily because of the increased use of social security numbers (SSNs) for linkage purposes. Clearly, some proportion of survey respondents who are asked to provide SSNs are going to want to know why. The survey organization must be prepared to give satisfactory answers. It may not be easy to reach agreement on just what constitutes a satisfactory answer.

An even more difficult problem is to decide what ought to be said to survey participants about non-zero disclosure risks associated with the release of public-use microdata files. When microdata files are released, we cannot guarantee that recipients of these files will be unable, should they so desire, to identify some persons whose records are included (see, for example, Cox, 1985 and Paass, 1985). The same technical tools that have greatly facilitated computerized record linkages are available to "attackers" who might want, for whatever reason, to identify individuals in a file of records from which primary identifiers, such as name, address and SSN, have been removed.

Some of the Federal agencies that conduct voluntary surveys have been carefully reviewing and modifying their notification practices over the

past 3 or 4 years. We have heard about recent developments of this kind from William Butz, the Census Bureau representative on this panel. The environment in which these reviews and alternatives are being undertaken includes the following features:

- As mentioned above, questions raised about the need for collection of SSNs require answers that deal explicitly with planned record linkages.
- Recent increased matching of IRS records to administrative records for other Federal and state programs, for compliance purposes, appears to have raised the level of public concern about large Federal data files and their actual and potential uses.
- The recent furor in Sweden about Project Metropolitan, a longitudinal social science research study involving extensive linkage of sensitive administrative records, has apparently resulted in a substantial increase in non-response to Sweden's monthly labor force survey (Dalenius, 1986).
- In comments on the Duncan-Lambert (1985) article in JASA, Gastwirth accused Federal agencies of "deceptive" practices with respect to record linkages performed for statistical purposes. Responses to Gastwirth's allegations were prepared by Federal agencies and concerned individuals. However, after considerable negotiation, they were denied, as the result of a decision by the ASA Publications Committee, of any opportunity to publish these responses in JASA.

These reviews of notification practices are essential and should be carried out periodically for all continuing survey programs. Voluntary surveys and indeed even the mandatory census of population cannot succeed without public support. To earn that support, the agencies that conduct the surveys must not only comply fully with legal notification requirements, but must also make a good faith effort to put themselves in the shoes of survey participants and to treat them fairly. The problem comes in trying to derive specific procedures from this broad dictum.

IV. SUGGESTED GUIDELINES

The ISI declaration is correct when it states that no universal rules can be framed for notification to participants in voluntary surveys. Much depends on the content of the survey, the nature of the survey population, the data collection mode and the planned uses of the data. Choices must be made among different methods of notifying participants and among a wide range of content, ranging from a bare minimum (essentially only what is enough to satisfy the Privacy Act requirements) to an exhaustive discussion of each of the 12 items listed in the ISI declaration (and possibly other topics). I will try, in the remainder of this section, to

suggest some principles for use in developing procedures for a particular survey, with emphasis on questions relating to record linkages and the release of microdata files.

Method of Notification

- (1) Written statements should be the primary means of notification.

Long oral statements are likely to be burdensome to respondents and may contain information they are not very interested in having. It is difficult to phrase such statements in language that all can understand quickly. Interviewers cannot always be counted on to read the full statement to every respondent. In a household survey, oral statements are not available to adult household members not present at the time of the interview. A written statement, on the other hand, can be read by all adult household members and each one can study it with as much or as little care as he or she chooses.

- (2) Whenever possible, written notification statements should be provided in advance of the survey interview or at the time of the survey.

For a self-administered survey (e.g., a mail survey), the notification statement will usually accompany the questionnaire. For an interview survey, an advance mailing is desirable if mailing addresses are available for the sample units. For some kinds of surveys, e.g., telephone surveys using random-digit dialing, advance mailing is not practicable. An alternative for this situation is suggested by some state and local consumer protection laws which provide that contracts for certain kinds of goods and services can be revoked within a specified number of days of signing. A notification statement could be offered to respondents during the interview, with the proviso that anyone who requested, after receiving and reading the statement, that some or all of his or her data not be used, would have that request honored. The same proviso might even be extended to surveys with advance or simultaneous notification.

- (3) Provisions should be established to respond fully to additional questions that survey participants may have.

To the extent feasible, interviewers should be trained to answer respondent questions not covered in the notification statement. It may be more efficient and effective, however, especially in large surveys, to have one or two specialists who can provide clear, complete and accurate responses to questions about privacy and confidentiality. Addresses and toll-free numbers for these specialists could be provided as part of the notification process.

Content of Notification Statements

(1) Make no false statements.

I doubt that false statements are being made deliberately; they have been and perhaps still are being made through failure to think through their implications. In the past, I have seen statements that data would be released only in aggregate form when, in fact, microdata files were being released. I would attribute this sort of thing to simple oversight or, at worst, a desire to economize by using up a supply of old forms.

A more difficult question is illustrated on a notification statement now in use that says:

"No information that can reveal the identity of a person or business can be released."

Given the consensus among experts that zero risk of disclosure is unattainable, especially when microdata files are released, it is questionable whether this kind of statement can be justified. It is not my intention to propose specific alternative language here, but the obvious direction would be to describe the steps that will be taken to minimize the risk that individual participants will be identified, resulting in disclosure of information about them.

A non-zero risk of disclosure, however small, may not be acceptable to some potential survey participants. It may be made somewhat more acceptable if:

- persons who deliberately disclose survey confidential information or who cause disclosure through their negligence are made subject to appropriate penalties; and
- survey participants harmed by disclosure of information about them can receive appropriate compensation.

The first requirement suggests the need for legally binding user agreements with persons who receive survey microdata files. Both conditions may require new legislation. Is it worth the effort? I think so: if these protections are not established, it may not be possible to continue releasing micro-data files. That would be a very unfortunate outcome.

(2) Do not make promises of confidentiality for which you do not have clear legal authority

To my knowledge, most Federal agencies have observed this principle in their notification statements for voluntary surveys. Usually, identifiable survey data are exempt from disclosures requested under the Freedom of Information Act. If they are not, however, survey participants should be so informed. A good example of a statement to cover this situation in a non-survey context is the one that the National Science Foundation makes to reviewers of proposals:

"The Foundation considers reviews to be exempt from disclosure under the Freedom of Information Act but cannot guarantee that it will not be forced to release reviews under the Freedom of Information Act or other laws."

(3) Advise survey participants of any planned or potential non-statistical uses of information about them

In the private sector, many "surveys" turn out to be primarily sales pitches or requests for contributions: the main interest of the surveyor is to get the names of persons interested in his or her product, service or cause. Fortunately, most Federally sponsored surveys are conducted purely for statistical purposes and sponsors are forbidden by law from using the data for any other purpose. There are exceptions, however, and in such cases notification statements should include a clear and complete statement of any possible nonstatistical uses of the data collected.

Other Issues

(1) No uses should be made of survey information that are incompatible with the notification statements to survey participants

The purpose of this guideline is to cover the case where the survey agency would like to use the data in some way that was not anticipated during the period of data collection. Suppose, for example, that the survey agency wants to enhance the survey data base through a linkage with administrative records and that, in order to perform the linkage, some identifying information would have to be disclosed to the agency that maintains the administrative records. If the survey agency had told participants that no identifiable information would be released to other agencies, the proposed linkage would, in my view, violate the conditions under which the data were collected. On the other hand, if the notification had been couched in more general terms, e.g., the data will be used only for statistical purposes, then the proposed linkage would be legitimate, provided it could be done under conditions which assured that promises of confidentiality made to survey participants would not be violated.

(2) In surveys where linkages based on SSNs are to be undertaken:

- (a) No linkages should be undertaken in cases for which SSNs are refused, and
- (b) Linkages should not be used to obtain refused data items, even if SSNs have been provided.

Most agencies already follow these guidelines. One implication is that data collection instruments and procedures should be designed to make possible a clear distinction between refusals and item nonresponse occurring for other reasons.

V. CONCLUDING REMARKS

The Privacy Act and the non-binding recommendations of the American Statistical Association committees and the International Statistical Institute provide useful guidelines for notification to participants in voluntary surveys. The surveying agencies are making a good-faith effort to follow these guidelines. I have attempted to analyze and extend the guidelines to deal with two specific issues: record linkages and the release of microdata files. For the most part, I feel that Federal surveys are dealing adequately with the record-linkage issue. Some adjustments in notification statements are necessary to deal with the non-zero disclosure risks that are associated with the release of survey results, especially in the form of microdata files.

Technical means of disclosure limitation or avoidance, although useful, cannot solve the problem of disclosure risk. Statutory and administrative solutions--binding user agreements with penalties for violations and remedies for survey participants harmed by disclosure--are needed. Lacking such solutions, it may prove to be impossible to continue the release of microdata files.

REFERENCES

American Statistical Association (1977), "Report of Ad Hoc Committee on Privacy and Confidentiality," The American Statistician, 31, 59-78.

----- (1983), "Ethical Guidelines for Statistical Practice: Historical Perspective, Report of the ASA Ad Hoc Committee on Professional Ethics, and Discussion," The American Statistician, 37, 1-20.

Cox, L., Johnson, B., McDonald S., Nelson, D., and Vazquez, V., (1985), "Confidentiality Issues at the Census Bureau," First Annual Research Conference Proceedings, U.S. Department of Commerce, Bureau of the Census, 199-218.

Dalenius, T. (1986), "The 1986 Invasion of Privacy Debate in Sweden," unpublished report.

Duncan, G.T. and Lambert, D. (1986), "Disclosure-Limited Data Disseminations" (with comments), Journal of the American Statistical Association, 81, 10-28.

Gastwirth, J.L. (1986) "Comment" (on Duncan and Lambert, 1986). Journal of the American Statistical Association, 81, 23-25.

International Statistical Institute (1986), "Declaration on Professional Ethics" (adopted: August 1985), in International Statistical Review, 54, 227-242.

Paass, G. (1985), "Disclosure Risk and Disclosure Avoidance for Microdata," presented at the May 1985 meeting of the International Association for Social Service Information and Technology.