# ISSUES IN PROTECTING CONFIDENTIALITY IN NATIONAL HEALTH STATISTICS

Robert H. Mugge, National Center for Health Statistics

It has been pointed out that the Hippocratic Oath speaks to people running health statistics programs as well as physicians when it says, "Whatsoever things I see or hear concerning the life of men, in my attendance on the sick or even apart therefrom, which ought not to be noised abroad, I will keep silence thereon, counting such things to be as sacred secrets." (1) But I think that health statisticians should also take to heart the preceding sentence in the Oath: "Whatsoever house I enter, there will I go for the benefit of the sick, refraining from all wrongdoing or corruption, and especially from any act of seduction, of male or female, of bond or free." (Encyclopedia Brittanica) In conducting statistical research among people we must avoid the temptation to seduce our subjects to divulge any information which, in the hands of researchers, can harm them in any way.

Health statistical programs must be administered in such a way as to bring the best scientific results but also to bring no harm--physical, emotional, or social--to any participating subjects. These programs require a number of policy decisions in order to achieve such a result. Each such decision creates a policy issue for the administrator. Administrators must set policies on:

- o providing informed consent to study subjects;
- o physical protection of records;
- o training and supervising staff on confidentiality matters;
- o avoiding statistical disclosures through published tables or data tapes;
- o establishing appropriate limits to the sharing of data with other researchers;
- o determining what else is necessary to meet the letter and spirit of existing laws and regulations; and
- o changes to be requested in laws and regulations.

## FOUR ISSUES

Like every other organization that operates a statistical program, the National Center for Health Statistics (NCHS) has resolved these issues and established its policies to meet them, consistent with the laws and agency regulations. We are very satisfied with most of these policies and have operated quite successfully, we think, with them for a number of years (2,3,4). But they are always subject to reconsideration as the situation changes, new laws are passed, or new concerns are voiced. In NCHS we have a Confidentiality Committee which studies all new confidentiality issues that arise and recommends actions for the Director to take on them. Also, as new and different surveys are developed they often present new problems on policy applications; these are studied and resolved by the Center's Committee on the Protection of Human Subjects or by an Institutional Review Board, if necessary. A recent example was the set of problems which arose when the National Survey of Family Growth was expanded in its Cycle III to include questions of teenage girls on sex practices (5).

There are four particular issues which--although none is really a burning issue with us at the present time--have been the subject of some controversy and which I shall discuss:

(1) How does the Center deal with and apply the three laws relating to confidentiality of our records?
(2) How do we obtain informed consent in telephone surveys?
(3) How much and what kinds of information should we release on public use data tapes? and
(4) What should be the Center's stance on the proposed Confidentiality of Federal Statistical Records Act?

**1. Enforcing the Laws.** NCHS activities respecting confidentiality are governed by Section 308(d) of the Public Health Service Act (42 USC 242 m(d)), by the Privacy Act of 1974 (5 USC 552a), and by the Freedom of Information Act (5 USC 552).

The confidentiality provisions of Section 308(d) of the Public Health Service (PHS) Act state in effect that (1) data obtained by NCHS under its legal mandate may be used only for the purposes for which they were obtained and (2) they may not be disclosed in identifiable form without the consent of the person or establishment providing the information or described in it. Until recently there was a notable loophole in the law, i.e., the first clause was subject to regulations of the Secretary of Health and Human Services, which, according to our legal advisers, meant that the Secretary could at any time redefine those purposes and then release the data to anyone he or she chose. However, this was changed and the Secretary's option removed in one of the technical amendments included in the recently enacted Orphan Drug Act (P.L. 97-414, Sec. 8(c)).

The Privacy Act of 1974 does not provide to NCHS any additional privacy protection for our records, and it contains some big loopholes which the PHS Act keeps closed for us. Of course, the Privacy Act is very helpful to agencies which have no other legal authorization for protecting the privacy of records on individuals; it is also helpful in providing for severe sanctions to violators of privacy, which the PHS Act does not do. But the Privacy Act contains many other requirements affecting agencies, some of which can be onerous. The Act requires agencies to grant persons access to their records, to allow them to request changes in records they believe to be in error, to appeal any agency refusals to change the records, and to insert their statement in the record if the appeal fails. (This is only a potential problem; we have never been asked to revise any of our records.) Agencies maintaining statistical records under a legal mandate may request a (k)(4) exemption: from such requirements, and the Department has granted to NCHS such an exemption with respect to its statistical files. The Center will nevertheless furnish copies of records to persons when they can show real need for them and it is feasible for us to do so.

Another possible burden to agencies is the Act's requirement that agencies must publish in the Federal Record notices describing all their systems of records on individuals. These notices had to be republished each year until this year, when that requirement was dropped in an amendment to the Act (P.L. 97-375); now only changes in system notices and notices of new systems must be published.

The requirement to publish system notices could have become very onerous to the Center if every separate survey or study had been defined as a system of records. Instead we were permitted to define families of surveys and reporting programs as systems of records, and we published "umbrella notices" on them, describing in general terms the sets of programs included. Thus, we published only four system notices on all our statistical programs--one covering demogra-

phic surveys, one on vital statistics, one on health manpower surveys, and one on health care utilization studies (6). This does not short-change the public on information about NCHS surveys because in our own published reports we describe the individual surveys and studies in great detail, and persons asked to participate in the surveys are first given full explanations of them (7). We thus minimize the staff work in the Center needed to implement the Privacy Act while fully meeting its requirements.

The evidence reaching us indicates that the public is not worried about the protection of privacy in NCHS. Only very rarely do we receive any public comments on our Federal Register notices, and we almost never receive any requests for Privacy Act records. About the only kind we ever get any more are like the recent request from a person who knew that the government was bombarding him personally with a lethal ray, and he wanted the record from us to prove it!

The Freedom of Information Act (FOIA) presents no very difficult problems to us. The Center exists to provide information, and we try to meet all requests as fully and promptly as possible. Problems only arise when someone requests identifiable records which the law precludes our giving out, i.e., records whose release would violate the rights of privacy of an individual or an establishment. We have been able to deny successfully all such requests on the basis of the (b)(3) exemption in the FOIA and our PHS Act, since to accede to the requests would violate the law by using information for a purpose other than that for which it was obtained.

**2. Providing Informed Consent in Telephone Surveys.** The language of the Privacy Act presented a problem for agencies desiring to experiment with telephone interviewing as a method for collecting data more efficiently. The Act states in Section(e) that "Each agency that maintains a system of records shall --...(3) inform each individual whom it asks to supply information, on the form which it uses to collect the information or on a separate form that can be retained by the individual" as to the several kinds of information needed for informed consent. Some construed this to mean that in a telephone survey such information must be supplied to the respondent on a piece of paper before an interview could begin. This of course is not feasible in most telephone surveys, and the effect of such a determination would be that government agencies could not conduct telephone surveys. Fortunately our Department's General Counsel ruled that that was not the intent of Congress, and the law's requirements can be met by having the telephone interviewer provide the required information orally to the respondent and then sign an affidavit on the interview form that the information had been provided. This has allowed NCHS to embark on a successful telephone survey program (8).

**3. Data on Public Use Microdata Tapes.** The Center cannot begin to exploit adequately the data files it develops through its various surveys; it can only perform and publish a fraction of the potentially very valuable scientific analyses these extensive files make possible. Fortunately there are many researchers in the universities, foundations, and research firms, or working independently, who are able and anxious to study various health problems through the use of these files. The Center wants them to have the files for this use but must be careful that no confidential information about identifiable persons or establishments is disclosed through the release of survey files (2,3).

Of course, all direct identifiers of study subjects, such as name, address, and social security number, are deleted from the public use files. Still, there are so many different items of information about any subject individual or establishment in our typical surveys that the set of information could serve as a unique identifier for each subject, if there were some other public source for many of the survey items (9). Fortunately there is not. But to minimize the chance of disclosure we take additional precautions: We make sure there are no rare characteristics shown on any case in the file, such as the exact bed-size of a large nursing home, or the exact date of birth of a subject, or the presence of a rare disease, or the exact number of children in a very large family. We either delete such items or conceal them in a broad coding category. We also remove or encrypt the code identifying smaller geographic areas--places smaller than 100,000 in population--because anyone trying to identify a respondent will have his task greatly simplified if he knows the respondent's local area.

We also require the purchaser of a microdata tape to sign a statement in which he or she agrees to abide by the law which states that the data may be used only for the purpose for which they were obtained, i.e., for statistical analysis. We think this at least helps to sensitize the purchaser to the importance of protecting the data from misuse.

There are also other built-in protections against disclosures. Typically our surveys are based on samples of less than 1 in 1,000 of the population, so the odds of identifying a known individual through the sample are extremely small. Also, some of the information in the survey file was unknown and had to be imputed, and there are reporting errors, so if one thought he had located a case he could still not be sure about the accuracy of the file's information on that case.

We could take further steps to reduce the probability of disclosure, such as by further generalizing the data or introducing random errors, but any such procedure would diminish the accuracy and value of the files for research purposes. We do not wish to do that, and we do not believe we need to.

So there remains a very small risk of disclosure from our public use files. However, we have determined that the value to society of having the file available for statistical use in this form justifies the very small risk involved, especially in view of the low sensitivity of most of our survey data.

We are not aware of any criticism that we are too liberal in our release of public use tapes, but we are critized for being too strict. This especially comes from those who wish to conduct research relating study findings to physical and social characteristics of local areas; they may be stymied by our refusal to release the local area identifiers. There are sometimes measures we can take to enable such research. We can release such data to other agencies of the Department or their contractors under interagency agreements, and we can sometimes attach outside variables for local areas to a public use tape, if the variables do not then constitute new local area identifiers. But otherwise the would-be researchers remain frustrated, and they dispute our policy.

**4. The NCHS Position on Proposed Statistical Confidentiality Legislation.** I shall not describe in detail the proposed legislation, since Clark and Coffey are doing so in their paper. Briefly, the proposal is to set up a Federal "statistical enclave" made up of "Protected Statistical Centers" all of which would be given strong legal protection for their "Protected

Statistical Files" but could easily share such files between Centers when justified by valid statistical purposes.

NCHS is comfortable in its present legislative position. As noted, we have a strong law governing confidentiality of our statistical records, and there are few files of other agencies that we anticipate needing to use. Therefore some have said that we should oppose the proposed law and not risk losing our present favorable position. But it is my view that even if NCHS got nothing directly out of the new law for itself it should still support it because it is in the best interests of the total Federal statistical program. However, I should qualify this to say that I am only referring to the latest version I have seen of the draft "Confidentiality of Federal Statistical Records Act" (10). This bill is a moving target, and I can't be sure I would also support the next version!

I must qualify further to state that I share Professor Bonnen's concern that the proposed bill may be unworkable because it decentralizes the federal statistical system and places authority for operating it in the hands of the respective departmental secretaries (11). I believe as he does that the "statistical enclave" can only operate successfully if it is run by a strong Chief Statistician backed up by an effective statistical policy coordinating agency located in the Office of the President.

There is, however, one thing we hope very much to get from passage of the Act. In order to develop the most efficient and effective population sample for use in the various surveys conducted by NCHS, we need to use as a sampling frame the list of household addresses developed for, and partly in the process of conducting, the decennial census of population and housing. The nation could obtain much more reliable data in its health surveys, at considerably lower cost, if we could use that sampling frame. The Department of Commerce will not now let us use that list for a sampling frame, asserting that this would be in violation of the confidentiality requirements of the Census law (Title 13). So we are having to use a more expensive and less efficient area probability sample. Now, we can understand and we strongly support the absolute protection of the personal, family, and household information obtained in the decennial census, but we frankly cannot understand that this protection must extend to the address list, especially when that list shows no indication as to whether the dwelling unit was occupied nor, if it were occupied, any information about the occupants.

It is our understanding the the proposed law would clarify the legal question and assure that as a Protected Statistical Center, and with proper safeguards, NCHS could have access to the list for needed sampling purposes. I understand that the law would still preclude our having access to the 1980 Census list, but it should make possible our using the lists developed in the 1990 and subsequent censuses, so that this unfortunate situation will not continue into the future.

## CONCLUSION

Those, then, are four issues that I think should be of general interest, with brief statements of the policy positions that NCHS has taken on the first three points and a recommended position on the fourth. They are debatable, and it would be helpful to the Center to hear others' views on them.

NOTES AND REFERENCES


1. Lichtig, L. K.: "Confidentiality Issues in Health Statistics: An American Perspective." Unpublished. Albany: Blue Cross of Northeastern New York. 1983.

2. National Center for Health Statistics: Policy Statement on Release of Data for Individual Elementary Units and Special Tabulations. DHEW Pub. No. (PHS) 78-1212. Public Helth Service. Hyattsville, Md.: U.S. Department of Health, Education, and Welfare. May 1978.

3. ...: Staff Manual on Confidentiality. DHEW Pub. No. (PHS) 80-1244. Public Health Service. Hyattsville, Md.: U.S. Department of Health, Education, and Welfare. April 1980.

4. ...: Staff Manual on Confidentiality Revision. Unpublished document. Hyattsville, Md. 1982.

5. Tanfer, K., Grady, W., and Bachrach, C.: "Consent and Privacy in the National Survey of Family Growth: A Report on the Pilot Study for Cycle III." Vital and Health Statistics, Series 2, No. 91. DHHS Pub. No. (PHS) 82-1365. Public Health Service. Hyattsville, Md.: U.S. Department of Health and Human Services. March 1982.

6. Federal Register, Vol. 47, No. 198, Wednesday, October 13, 1982. Washington: U.S. Government Printing Office. Pp. 45685-45690.

7. Pearce, N. D.: "Data Systems of the National Center for Health Statistics." Vital and Health Statistics, Series 1, No. 16. DHHS Pub. No. (PHS) 82-1318. Public Health Service. Hyattsville, Md.: U.S. Department of Health and Human Services. December 1981.

8. Thornberry, O. T., and Poe, G. S.: "NCHS Research on the Telephone Interview: Some Observations." Proceedings of the American Statistical Association, Social Statistics Section, 1982. Washington: American Statistical Association. 1982. Pp. 296-301.

9. Subcommittee on Disclosure-Avoidance Techniques: Statistical Policy Working Paper 2: Report on Statistical Disclosure and Disclosure-Avoidance Techniques. Federal Committee on Statistical Methodology. Washington: U.S. Department of Commerce, Office of Federal Statistical Policy and Standards. May 1978.

10. Office of Management and Budget. Draft, proposed Confidentiality of Federal Statistical Records Act. Unpublished. Washington. 1983.

11. Bonnen, J. T., "Federal Statistical Coordination Today: A Disaster or a Disgrace?" The American Statistician, Vol. 37, No. 3, August 1983. Washington: The American Statistical Association.