

ACCESS TO SURVEY DATA: THE CONSUMER EXPENDITURE SURVEY

Eva E. Jacobs, Bureau of Labor Statistics\*

I am the Chief of the Division of Consumer Expenditure Surveys and my talk will be concerned largely with our policy for that survey. I have chosen to address topic 5 of the proposed topics prepared by the chairperson of this session: "Which of the following statements about current disclosure avoidance policies of Federal agencies comes closest to your views?"

- a. They do not provide enough protection to respondents.
- b. They are about right.
- c. They place unnecessarily severe constraints on user access to the data."

The current Consumer Expenditure Survey has been conducted by the Bureau of the Census for the Bureau of Labor Statistics since 1980. The survey is actually two surveys, one a two week diary completed by 5000 families a year and the other a series of 5 quarterly interviews with a rotating group of 5000 families a quarter. Information is collected not only about all the expenditures of the household but also about the characteristics, labor force status, income, taxes, changes in savings and borrowing and assets.<sup>1</sup> However, we do not collect social security numbers.

The data are collected by the Bureau of the Census under the promise of confidentiality in Title 13 of the U.S. Code. After preliminary editing, Census sends BLS a tape of the micro data from which the respondents' names and addresses have been removed.

At BLS, the data are further processed, allocated, imputed etc. and prepared for use. The major BLS use of the data is for updating the weights and item selection for the Consumer Price Index (CPI). The data however, are also of great interest for economic analysis because they provide the opportunity to associate expenditures with demographic characteristics. It is this aspect that makes the survey unique. For example, the data may be used in models to measure the impact of imposition of a sales tax on households at different income levels, or to project expenditures under different assumptions of demographic distribution of the population.

BLS makes the data available to the public in publications of grouped expenditures by household characteristics. We also make available a public use tape of micro data. Since the data are collected under the Census Bureau confidentiality legislative requirements, the data proposed to be included on the public use tape are submitted to the Census Micro Data Review Panel for approval. A great deal of characteristic information is given on the tape. To avoid possible identification

problems, a major piece of geographic information is omitted, the Primary Sampling Unit (PSU). The region and city size are shown but in the Northeast, for example, a tape user cannot distinguish between New York, Boston, or Philadelphia.

In addition, actual incomes, assets, rent, estimated value of home, etc. are not shown above certain levels. For instance, incomes above \$100,000 are coded as \$100,000 with a flag indicating that the reported value exceeds that number. These are the major adjustments made on the tape though there are others.

Now to return to the issues at hand. The major consideration is, of course, the legal requirement. But beyond that is the need to protect the credibility of the confidentiality promise. Here perception is as important as reality.

In conversations with interviewers, we receive indications that there is great concern with the question of the uses to which the data are going to be put. To fortify our statement that the uses are statistical, we provide copies of the monthly Consumer Price Index press release and the publications of BLS showing expenditures by income class, age, etc. Response to the survey has been good but as in most government surveys, has been declining recently.

There is a conflict in user needs. The primary purpose of the data collection is to satisfy direct government need for a major public data series. For example, a great deal of economic activity is dependent on the CPI. The credibility of the CPI market basket depends on the reliability of the Consumer Expenditure Survey data. The use of government data collected by research workers is not necessarily less important to long range government policy but is not as direct. Many users, economists and econometricians, would say that the constraints are unnecessarily severe and we have received many requests from non-BLS research workers for the obscured information. BLS can sometimes offer to prepare special tabulations or analyses for users in order to maintain basic confidentiality requirements and yet meet user needs.

The issue of trust is extremely important and subtle and examples of misuse or perceived misuse can be damaging to all surveys in all countries.

The story of what happened in Sweden this year is appropriate here. It appeared in the New York Times:<sup>2</sup>

"Deep anxieties over privacy issues were stirred here recently with the discovery that for 20 years a team of sociologists had been compiling detailed profiles of nearly 15,000 Swedes by systematically rummaging through computerized official

records."

The Swedish Data Inspection Board provides licenses for individuals and organizations that want to keep computerized files on individuals. Project Metropolitan, which followed individuals from official files by name and number for 20 years, came before the Board and, by agreeing to respect restrictions set on data collection and storage, had its license renewed. Other than the first questionnaire distributed in high school, the subjects of the project were unaware that their lives, including health, and criminal records, were being tracked. After the story broke in the newspapers there was a great deal of public agitation. In the end, the Board voted to require the destruction of the Project Metropolitan tape which, even when "de-identified," showed the correspondence between the official numbers of the subjects and the research study number.

To quote from another summary of the contretemps prepared by Tore Dalenius, "Before the debate ..., there were already signs of a growing public distrust in government agencies, including Statistics Sweden. The debate appears to have enhanced that distrust." The non-response rate to the Swedish monthly employment/unemployment survey rose from 5 to 11 percent<sup>3</sup>.

It is this kind of reaction that we are trying to avoid by maintaining the current procedures. If public confidence is destroyed then neither the government or other researchers will have any reliable data. My answer to the question we started with is therefore, that for us the present system works. We have been able, with a few exceptions, to assist users who are not served well by the current rules.

The Committee on Privacy and Confidentiality of ASA has initiated a project to prepare a brochure for the public to inform people about surveys. The Committee's pro-

posal is to have a pamphlet for survey takers, respondents, the press, etc. The objectives are to answer questions about privacy and confidentiality of response to surveys; support the survey objective of fulfilling data needs; clarify responsibilities of survey takers; specify rights and reasonable expectations of respondents.

Confidentiality, linkage and access to data have been the subject of numerous ASA sessions, journal articles and special committees. The differences of opinion have been wide and strong. In the light of a recent report of the Office of Technology Assessment (OTA), the discussion may even be moot. According to OTA, "Social Security numbers are irrelevant ... the Privacy Act is unenforced and unenforceable" (because of computer matching).<sup>4</sup> However, new technology can be used to protect as well as to increase the danger. Software that permits micro linkage but provides output on an aggregate basis with a minimum cell count criterion is one procedure that may be available. Various proposals made by some of the speakers at this meeting also merit serious consideration. Perhaps it is time for a quiet review by both government and private users of survey data to resolve some of these differences.

\*Views expressed here are my own and not necessarily those of BLS.

#### References

- 1 For a complete description see BLS Handbook of Methods, December 1982, chapter 6.
- 2 New York Times, March 12, 1986.
- 3 Letter and report from Tore Dalenius, July 16, 1986.
- 4 The Washington Post, July 1, 1986.