

DISCUSSION

Roger Herriot, Census Bureau

The work discussed in the three papers given at this session represents the foundation of the 1973 Match Study. Taken together, the papers provide a clear and concise blueprint of the construction of that foundation and of its strength. I think that the authors -- and Fritz Scheuren who directed the work -- should be commended for the fine job they did. Although more remains to be done, soon we expect to have a file on which about 90 percent of the persons 14 years or older in the March 1973 Current Population Survey will have social security numbers that have a high probability of being correct. Furthermore, since some of the remaining 10 percent of the CPS sample have not received a number, we estimate that we will have SSN's for about 92 percent or more of the persons in the Study who actually have had a number issued to them.

Foundations have a tendency to get covered over as people begin to build upon them. Much that might be learned, which would be of use to future studies, can, of course, be lost because of this. Every effort has been made in the 1973 Match Study, as the papers at this session attest, to examine and reflect on the methods used with a view towards improving similar studies in the future. A number of suggestions along these lines can be found in the papers. Some of them would require major changes in the CPS that may be too expensive to implement, such as asking for father's name, mother's maiden name, and place of birth, etc. Other changes could be made at almost no cost, such as having the interviewers attempt to get the middle names or initials of all household members. Procedural changes at Social Security might also be made in future matching efforts. For example, as was pointed out in the second paper, the Study would have reaped significant benefits if there had been time to "re-search" for the missing SSN's of individuals for whom no number was found after just one search. Hopefully, in future studies, multiple searches will be possible, perhaps using somewhat more automated techniques.

In the remainder of my remarks, let me re-emphasize and expand on some of the points already made about the general nature and purpose of the 1973 Match Project. I would like to begin by stating categorically that this work is not being undertaken in anticipation of the development of any large-scale data bank. In fact, I believe that matched sample data sets, such as this one, constitute the best defense against the development of such a data bank for all persons. The

reason is that such data, for a relatively small sample of persons, can fulfill virtually all of the statistical needs, without running the risks for abuse inherent in a population data bank.

STUDY GOALS

As has been said, the 1973 Match Study is a joint statistical project between the Social Security Administration and the Bureau of the Census. It involves an exact match for persons in the March 1973 CPS -- of their survey data, their Social Security earnings and benefits data, and a small amount of data from Federal income tax returns. The purposes of this project are as follows:

1. To use the administrative data to determine the correctness of the survey data and to ascertain any biases in the Bureau's statistics. For example, among the particular projects being planned are:
 - (a) evaluation of the CPS reporting of income with respect to both reciprocity and amounts;
 - (b) evaluation of procedures used to allocate missing income amounts; and
 - (c) evaluation of the weighting and control procedures used to adjust for noninterviews.
2. To use the administrative data to correct misreporting of the survey information and to augment the survey with data which the respondents were not asked to provide, such as work experience or income for a number of years. These data can then be used in many ways; for example,
 - (a) as inputs to models which simulate the effect of proposed new tax-transfer programs or changes in existing programs;
 - (b) to develop corrected income size distributions and to prepare estimates of after-tax income distributions; and
 - (c) to test and evaluate various statistical matching procedures. This, we feel, is particularly important considering the increasing amount of statistical matching which is being undertaken.

CONFIDENTIALITY

The three-way data linkage required that special operating procedures be instituted at Social Security and at the Bureau of the Census to insure the confidentiality of the shared information--in particular, to ascertain that the linked data was used only for statistical purposes and not administrative ones. The laws and regulations under which the involved agencies operate impose very definite restrictions on such exchanges.

Census Requirements.--Information derived from the Bureau of the Census' Current Population Survey is governed by policies and procedures established under Title 13 of the U.S. Code. This title requires that information about identifiable individuals remain under the direct control of employees of the Census Bureau at all times. On rare occasions, to better achieve its statistical goals (such as in this linkage project), the Census Bureau swears in, as its own temporary employees, a small group of employees of other agencies. In this instance, those Social Security Administration employees directly involved in the linkage, about 15 or so, were hired and sworn in as Census employees without compensation. These few individuals--technically employees of both agencies at once--have been legally given the means for access to both Census Bureau and SSA data, so that the linkage could be performed. Both "regular" and "special" employees of the Census Bureau are, of course, sworn to uphold the confidentiality of all census information and are subject to criminal penalties should they fail to do so.

Social Security Administration Requirements.--Information derived from Social Security Administration files is governed by Title II of the Social Security Act and the regulations established under that Act (specifically, Regulation No. 1, Sections 401 and 422). To release to the Census Bureau SSA earnings and benefit information for identifiable individuals, a special Commissioner's decision had to be obtained. This decision, dated June 28, 1973, was made subject to the following conditions:

1. No SSA information was to be given to the Census Bureau for any CPS respondent who refused to give his social security number to the Census interviewer.
2. All SSA data given to the Census Bureau were to continue to have the

protected treatment required by the Social Security laws and regulations. Furthermore, the data were to be subjected to Census' own confidentiality restrictions as imposed under Title 13.

3. After linkage, all individual identifications must be removed or scrambled in the resultant file.

Internal Revenue Service Requirements.--The Internal Revenue Service (IRS), under an executive order (promulgated under IR Code Section 6103) provided a magnetic tape file of abstracts of 1972 individual income tax returns to the Census Bureau for statistical purposes. Subsequently, IRS agreed to permit the Census Bureau to match a very limited amount of this data to CPS and SSA information, subject to the following provisions:

1. that individually identifiable IRS data continue to be subject to Title 13 and the various IRS confidentiality restrictions (specifically, IRS Code Section 7213); and
2. that after matching and the removal of individual identifiers, IRS is to have veto power over any data item on subsequent match files to be prepared for SSA, if IRS believes that the inclusion of the data item could possibly result in disclosure.

As with Census data, unauthorized disclosure of Social Security or Internal Revenue Service information is a punishable offense which can result in fines or imprisonment or both.

A number of operational implications flow from the above confidentiality requirements. Some examples of the steps taken were:

1. when competing confidentiality regulations existed, the strictest provisions were followed;
2. project computer tapes, whether they contained linked data or not, were stored in locked facilities when not in use; and
3. all Census confidential information was in the custody of Census regular or special employees at all times.

As an added precaution, just limited abstracts of the basic CPS and SSA data were used in the searching and matching work done at Social Security.