EVALUATION OF MISSING WAVE DATA FROM THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)

Vicki J. Huggins, U.S. Bureau of the Census[1]

I. INTRODUCTION

In longitudinal surveys where survey respondents miss one or more interviews, retrospective data for the missed interviews can be gathered at later interviews. Many people, however, have questioned the value of collecting retrospective data for missed interviews because the longer reference period can introduce response bias arising from recall lapse. A missing interview section was introduced into the SIPP questionnaire as a potential tool for improving longitudinal imputation and noninterview adjustment. Based on the examination of data from the missing interview section for one wave of interviewing, the missing wave form was dropped from the SIPP questionnaire. This paper provides results of the analysis which indicate that imputation and noninterview adjustment procedures using missing wave data would not likely provide significant improvement over other imputation and noninterview adjustment procedures. In particular, imputation and noninterview adjustment procedures that utilize prior and post waves of collected data should be comparable.

Section II of this paper offers a brief description of the SIPP 1984 panel sample design and its terminology. Background discussion of weighting and imputation alternatives for missing interviews and the use of missing wave data in weighting and imputation is given in section III. Section IV describes the methodology the missing wave analysis employed, section V describes the results of data analysis, and section VI contains conclusions as to the usefulness of missing wave data to the SIPP.

II. SAMPLE DESIGN OF THE 1984 SIPP PANEL

The SIPP is a national household survey designed to provide improved information on the income and program participation in government programs of the noninstitutional United States population. Person and family characteristics that may influence income and program participation are also available from the SIPP. The survey will explain the socioeconomic process in the U.S. and aid federal agencies in formulating and evaluating their policies and programs in the areas of income and social welfare.

The 1984 SIPP panel is divided into four groups of approximately equal size called rotation groups. Each rotation group is interviewed once every four months for approximately two and one-half years. In general, one four-month cycle of interviews is called a wave. This design provides a smooth and steady workload for data collection and processing. The reference period for the interview questions is the four months preceding the interview month. For example, the reference period for the December 1985 interview month is August through November 1985. These sample persons are interviewed again in April 1986 for the December 1985 through March 1986 period.

Persons 15 years of age and older present as household members at the time of the first interview were eligible to be a part of the survey for the entire two and one-half year period. With certain restrictions, these sample persons were followed if they moved to a new address. "New" persons who lived with sample persons were considered to be a part of the sample only while residing with these sample persons.

At _each_ interview, eligible persons were asked a set of core questions on labor force status, income and program participation in government programs. Questions on a variety of topics not covered in the core section were assigned to particular interviewing waves of the survey. If more than one observation on these topics was needed, these questions were repeated at a later wave.

III. WEIGHTING AND IMPUTATION ALTERNATIVES FOR MISSING INTERVIEWS

In the SIPP, we can classify sample persons into the following three types:
1. Persons who respond to zero interviews;
2. Persons who respond to some but not all interviews; and
3. Persons who respond to every interview.

In longitudinal weighting, a noninterview adjustment is used to correct for the first category of persons. It was suggested that the availability of retrospective data for missing interviews would influence the decision between weighting and imputation to correct for persons in the second category. (Singh, 1983)

A weighting adjustment to correct for persons who miss some but not all interviews requires discarding a rather large amount of data. For example, in the 1984 SIPP panel, 28% of wave 1 interviewed persons 15 years and older in interviewed households missed one or more interviews. (McMillen, 1987) Thus, 28% of the 1984 panel SIPP sample would be discarded in

longitudinal weighting and estimation despite the presence of some good data. [Note, however, that these data can be used for cross-sectional estimation.]

Alternatively, imputation for wave nonresponse allows full use of all conducted interviews. However, it could be very complicated depending on the pattern of interviewing for which imputation would be employed and the imputation methodology selected. Retrospective data collected on important transitions provided by the SIPP (entering or leaving a program, changing household type, changing labor force status, changing income sources, etc.) can be used to impute income amounts and other related characteristics normally collected in the SIPP. Thus, a missing interview form need not repeat the entire SIPP core questionnaire.

In the 1984 SIPP panel, 72% of wave 1 interviewed persons missed no interviews over the life of the panel. Assuming imputation based on missing wave data were to be implemented for persons who never miss consecutive interviews (7%), the data discard rate of 28% would be reduced to 21%. (McMillen, 1987)

As a result of concern over the potential loss of good data for persons who never miss consecutive interviews, the missing wave section of the SIPP questionnaire was introduced in the fourth wave of the 1984 panel. The missing wave form was designed to collect retrospective data from previously missed waves of SIPP. More specifically, information was gathered for interviews missing in wave(n) at wave(n+1) given that an interview from wave(n-1) was obtained. A short form of retrospective questions was developed to reduce respondent burden and was administered to respondents after completion of the regular wave(n+1) questionnaire. The missing wave section contained a skeleton set of SIPP core questions relating to labor force status, receipt of income and assets and program participation.

In longitudinal estimation for the SIPP, missing wave data may fill gaps in interviewing and improve SIPP estimates. The study described in this paper was designed to determine the usefulness of missing wave data to longitudinal imputation and/or weighting procedures and whether the increase in the amount and quality of data collected justified the respondent burden imposed by the additional set of missing wave questions.

IV. METHODOLOGY

In this study, data from the missing wave section of the SIPP 1984 wave 9 questionnaire were examined. Results are based on data analysis; statistical tests were not performed since the number of responses in most instances is small. The following types of data analyses were performed to assess the usefulness of the missing wave data.

A. Item Nonresponse Rates

The rate of nonresponse was determined for selected questions in the missing wave section of the final wave questionnaire for each rotation of the 1984 panel. Questions that require responses on both a four-month and monthly basis were included. In particular, rates of nonresponse for four-month based questions and rates of the patterns of response for monthly questions (e.g., months in which a person received a particular income source) were analyzed to determine which, if any, of the questions were answered well enough (by month or four-month period) to be useful for imputation purposes.

B. Completion Rate

The completion rate of the missing wave section was obtained by determining the number of persons who responded to at least one question on the missing wave form of those eligible to respond. Those eligible to respond were persons for whom a wave 7 and wave 9 interview were obtained but a noninterview occurred at wave 8.[2]

C. Reporting of Transitions

A wave transition is defined as a change in the status of reporting receipt of an income or asset type for a particular characteristic between two reference periods, e.g., the respondent reported receiving Social Security in at least one month of the reference period for the missing interview n, but not in any month of the n+1 reference period.

The actual numbers of wave transitions reported by those eligible to respond to the missing wave form for receipt of income and asset types were examined. Data from preceding and following interviews to the missing interview was transcribed onto the missing wave form only for these characteristics. Transitions in labor force were not examined because of the lack of prior and post waves of data to the missing wave for labor force characteristics.

The numbers of wave transitions that occurred for income and asset types for all persons interviewed during the last three waves of the 1984 panel was not available at the time of this research to use as benchmarks for the numbers of wave transitions detected by missing wave data. However, based on work completed by both John Coder (1986) on monthly transitions in SIPP for the first three interviews of the 1984 panel and Edie McArthur and Kathy Short (1986) on attrition in the SIPP, the number of wave transitions expected for selected income and asset

types for those persons with wave 7 and wave 9 interviews and a wave 8 noninterview were estimated. The estimated number of transitions for a particular income or asset type is the sum of wave transitions that we expect to occur between interviews 7 and 8 and between interviews 8 and 9. To compute estimated numbers of transitions from Coder and McArthur and Short's work, for use as rough benchmarks of missing wave data quality, we assumed the following:

o Noninterviewed persons are like interviewed persons;

o Reporting of characteristics is not affected by respondents' time in sample; and

o Reporting of characteristics does not change if different survey instruments and procedures are used to collect data for missing interviews.

The impact on our results of making the above assumptions is discussed in section V.

An estimate of the total number of wave transitions that should have occurred around wave 2 (the sum of wave transitions that occurred between waves 1 and 2 and waves 2 and 3) for particular income and asset types for all eligible respondents in the 1984 panel was multiplied by the rate for which a missing wave pattern occurred at wave 8, multiplied by the completion rate of the missing wave file and adjusted for the SIPP sample reduction that occurred in March 1985.

The degree to which item nonresponse may have occurred for selected income and asset types in the missing wave data was roughly approximated by dividing the difference in the estimated and the reported number of transitions for an income or asset characteristic around the missing wave by the estimated number of transitions. This ratio actually includes item nonresponse that may occur in missing wave data (in addition to the item nonresponse that occurs normally in wave data), differences in data quality resulting, for example, from the longer recall period of the missing wave, and differences attributable to the method employed to estimate transitions, i.e., the validity of our assumptions. What percent of the estimated item response rate pertains to data quality or differences attributable to the estimation of transitions is unknown.

V. RESULTS

Table 1 presents item nonresponse rates for selected characteristics in the missing wave data. Item nonresponse occurs when a characteristic does not have a valid response in its data field on the file. As a result of the missing wave questionnaire's skip pattern, the population that is eligible to respond to certain questions varies. Nonresponse rates are based only on persons eligible to respond to the question. Column 1 in table 1 provides a numbering of the selected questions in sequence from the missing wave form for which responses are analyzed. The number eligible to respond to the question is given in column 2 and the resulting nonresponse rate is given in column 3. Detailed tables are presented in Singh (1987).

The number of transitions that occurred around wave 8 in the missing wave data and the number of transitions expected to have occurred based on wave 2 for a selected set of income sources identified on the SIPP income source list roster are provided in table 2, columns 2 and 3 respectively. The total number of respondents who reported at least one transition for receipt of income types is 39. Since the total number of transitions reported is 40, only 1 respondent reported a receipt transition for more than one type of income.

Table 3 contains results for receipt of asset types, and is set up similar to table 2. The total number of respondents who reported at least one transition for receipt of asset types is 69 and the total number of asset transitions that was reported for receipt of assets is 70. Therefore, only 1 respondent reported receipt transitions for more than one asset type.

The nonresponse rate (table 1) for the first question is 13% and nonresponse rates for the majority of subsequent questions examined are very small with two exceptions. The response rates for questions 10 and 16 are high, 18% and 11% respectively, relative to other questions examined. It is interesting to note that questions 10 and 16 request the same information for two different characteristics, i.e. receipt of income, receipt of assets. Preceding questions to 10 and 16, questions 9 and 15 request similar information for receipt of income and assets as questions 10 and 16 but the nonresponse rates are lower, 6% and 9% respectively. It is possible that after responding to questions 9 and 15, which can be very tedious, respondents may have chosen not to burden themselves further by responding to a second question, similar in nature to the preceding question.

The completion rate of the missing wave form is 94%. In addition to the 87% of persons that responded to the first question, 7% responded to at least one subsequent question in the questionnaire. From this, it appears that interviewers made a concerted effort to obtain as much information as possible from the respondent, attempting to maneuver around item nonresponse.

The total number of transitions reported for receipt of the selected income types, presented in table 2, is itself small, only 40 transitions.

Breaking the number down by specific income types, only minimal numbers of transitions are reported. For example, one transition in the receipt of Veteran's Compensation or Pensions, 1 transition for receipt of Worker's Compensation, 0 transitions for SSI and State Supplemental Security Incomes were reported. Also, only 4 transitions for Social Security, 2 transitions for WIC and AFDC and 3 transitions for Food Stamps were reported. Significant improvement in estimates of transitions around wave 8 for most individual income types using missing wave data for imputations is not likely.

For receipt of asset types, presented in table 3, all but one asset type have minimal reporting of transitions at wave 8 for missing wave data. Six transitions for Stocks or Mutual Funds, 0 transitions for Money Market Funds, U.S. Government Securities, and Municipal Bonds and 1 transitions for mortgages. There were, however, 57 reports of changes in receipt of savings, Money Markets, CD's and Now Accounts. For all but this asset type, significant improvement to SIPP estimates of transitions using missing wave data in imputation would not be gained.

Although nonresponse rates for most characteristics examined in the missing wave section are not extremely high, the quality of the data for several characteristics examined appears suspect if the assumptions made earlier to obtain estimated numbers of transitions are correct. For example, only 26% of the estimated number of transitions expected to be reported for receipt of asset types were reported by eligible respondents in the missing wave section (table 3). Therefore, the estimated item nonresponse rate for these reports is 74%.

Similar results occur for reporting of income types, although not as marked (table 2). Only 51% of the number of transitions expected to be reported were reported which gives an estimated 49% item nonresponse rate for reporting of income and asset types at the missing wave.

Given the low rate of item nonresponse that occurs for most other variables examined (table 1), it is possible that item nonresponse is indeed reasonable but that the quality of transitional data collected for the missing interview, which is included in the estimate of item nonresponse, declines. The potential decrease in data quality could result from the longer recall period of the missing wave questions and the burden imposed by the missing wave section. Also, it is known in various surveys that there is a higher level for reporting certain characteristics at the first interview than for subsequent interviews due to respondents' time in sample. For longitudinal surveys such as the SIPP, the effect may even be more serious.

(Bailar, 1986.) Differences in the reported number of transitions to the number of transitions detected by the missing wave form could be attributed to factors specific to the missing interview and more generally to the comparability of the estimated to the reported number of transitions at the missing wave. We do feel however, that our overall results will not change if the assumptions stated earlier are incorrect. Primarily, because the reported numbers of transitions for receipt of income and assets at the missing wave are small whether they are benchmarked or not.

VI. CONCLUSIONS

The major finding of this study is that, while changes in receipt of income or asset types may have occurred, the missing wave questions detected a relatively small number in the eligible portion of the sample--four reports of change in receipt of social security, two reports of change in AFDC and WIC, for example.

If the last interview's missing wave data were to be used in imputation, approximately 0.2% of SIPP interviews (counting missing wave persons as interviews) at the next to the last interview would gain improvement in the estimation of between and within wave transitions for receipt of asset types. Only 0.1% of SIPP interviews would gain improvement in the estimation of transitions for receipt of income types. Improvement in the estimation of transitions for the majority of individual income and asset types would be negligible. For receipt of Social Security, for example, only 0.01% of SIPP interviews at the next to the last interview would gain improvement and for receipt of assets from Rental Property, only 0.02% would gain improvement in the estimation of transitions.

If, in weighting, the last interview's missing wave data were to be used to assign noninterviews at the missing wave to noninterview cells, a maximum of 0.3% of the SIPP sample eligible for interview at the next to the last interview would gain improvement in their cell assignment.

Since the number of transitions detected by missing wave data is small (40 transitions scattered over 39 income types and 70 transitions for 6 asset types), prior and post waves of data to the missing interview could be used as an alternative in imputation and/or noninterview adjustment procedures. This amounts to copying data for a respondent from a previous or later interview into fields for the missing interview. Results should be comparable to imputation and/or noninterview adjustment based on missing wave data.

We should note here also that imputation procedures which utilize

missing wave data will affect the covariance structure within the interview because all core data are not collected using the missing wave form. The covariance structure is not affected if imputation procedures utilize the closest wave of known data to the missing wave. However, the number of transitions would be slightly underestimated. For example, 3 transitions for Food Stamps and 1 transition for receipt of income from Mortgages would be lost.

And finally, data quality as discussed in sections IV and V for reporting receipt of income and asset types from the missing wave form appears questionable.

The availability of retrospective data for missing interviews was thought to be an important consideration in deciding whether persons who miss one or more interviews should be corrected for through weighting or imputation in longitudinal processing. Results from this study indicate that although missing wave data provide transitions for income and asset characteristics around the missing interview, the number detected is quite small. The respondent burden and cost imposed by the additional set of missing wave questions is not justified since it appears that other methods of imputation (which require less from respondents) are comparable. [The Census Bureau discontinued use of the missing wave form based on results of this study.]

The purpose of this study was to determine the usefulness of missing wave data to imputation and noninterview adjustment. The decision between weighting or imputation in longitudinal processing to correct for persons with one or more missing interviews remains an important issue. Eliminating 28% of eligible persons from a SIPP longitudinal panel file is very costly in terms of the number of income and program transitions lost and in its effect on variances. Subject matter specialists, data analysts and data users need to give further consideration to other imputation methodologies such as imputation based on the closest wave(s) of data.

Missing wave data from only the last interview of all rotations in the 1984 SIPP panel were analyzed. However, we do not feel that similar analysis of other waves of missing wave data will yield results very different from results provided by the last interview. Missing wave data was collected for up to six interviews in the 1984 SIPP panel, up to seven interviews in the 1985 panel, and up to four interviews in the 1986 panel. This data will be included on SIPP data files for further analysis by SIPP data users. Other uses for this data may also be identified.

REFERENCES

[1] Bailar, B., "Information Needs, Surveys, and Measurement Errors," Presented at the National Symposium on Panel Surveys in November, 1986.

[2] Coder, J., "Monthly Transitions from the SIPP Longitudinal Research File," Internal Census Bureau memorandum from Coder to Schneider, May 20, 1986.

[3] McArthur, E. and K. Short, "Measurement of Attrition from the SIPP through the Fifth Wave of 1984 Panel," Internal Census Bureau memorandum from McArthur and Short for the Distribution List, April 10, 1986.

[4] McMillen, D., "Verification of Matching SIPP Wave Files for the 1984 Panel," Internal Census Bureau memorandum from McMillen to Fink, May 5, 1987.

[5] Singh, R., "Retrospective Data for Missing Waves of SIPP," Internal Census Bureau memorandum from Singh to Bowie, November 20, 1983.

[6] Singh, R., "Evaluation of Missing Wave Data," Internal Census Bureau memorandum from Singh to Shapiro, February 19, 1987.

FOOTNOTES

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.

[2] For purposes of this paper, wave 7 refers to the 6th interview for rotations 3 and 4 and the 7th interview for rotations 1 and 2. Wave 8 refers to the 7th interview for rotations 3 and 4 and the 8th interview for rotations 1 and 2. Wave 9 refers to the last interview for all rotations.