

THE CREATION AND EVALUATION OF LONGITUDINAL NONRESPONSE CELLS
FOR THE 1996 SURVEY OF INCOME AND PROGRAM PARTICIPATION

Mark R. Hendrick, U.S. Bureau of the Census¹
Washington, DC 20233

KEY WORDS: Nonresponse Bias, Poverty Threshold Ratio, Replicate Weights

I. MOTIVATION FOR NONRESPONSE RESEARCH

Nonresponse is a critical problem for the Survey of Income and Program Participation (SIPP). The SIPP is a nationally representative longitudinal survey conducted by the Census Bureau. The survey collects information about the financial situation of persons, families, and households in the noninstitutionalized population of the United States. Being a longitudinal survey, SIPP requires multiple interviews over a period of years. Nonresponse increases with successive interviews, causing the sample to become less representative of the population it is approximating. This issue creates concerns about the size of person nonresponse bias in the longitudinal estimates of the SIPP.

Most previous SIPP panels have included six to eight interviews for up to a two and a half-year period. Longitudinally, person nonresponse rates have ranged from 25% to 29% for previous SIPP panels. Beginning in April 1996, Field Representatives will attempt to interview each sample household twelve times over a four year period. Under this new design, person nonresponse rates are expected to be higher for future panels, making nonresponse an even more critical issue for the SIPP (Weinberg and Petroni, 1992).

The expected increase in nonresponse over time should lead to an increase in nonresponse related bias. The objective of longitudinal nonresponse adjustment is to reduce this bias in SIPP estimates by accounting for those persons who were noninterviews for at least one month beyond the wave 1 interview. Developing techniques to improve SIPP nonresponse adjustment, in order to reduce nonresponse bias, is considered a major research goal. The modification of nonresponse adjustment cells is currently used as a technique to meet this research goal.

This paper summarizes research involving the determination and evaluation of redesigned longitudinal nonresponse adjustment cells. This research deals specifically with the improvement of the longitudinal nonresponse adjustment for the first calendar year of a given SIPP panel. However, this form of research would

also be beneficial in improving the longitudinal nonresponse adjustment for an entire panel. With SIPP moving to longer longitudinal panels, it will become important to consider longer periods of time for research purposes.

In this paper, we will define the methodology for various components of the research, present results, and then give conclusions. Results from this research will be implemented for longitudinal weighting in the redesigned 1996 SIPP panel.

II. METHODOLOGY

A. Nonresponse Bias

For this study, we selected nonresponse bias as our primary statistic for the determination and evaluation of redesigned longitudinal nonresponse adjustment cells. Previous research has suggested that nonresponse bias is small when the nonresponse rate is approximately 5% or less, but it increases as the nonresponse rate in the survey increases. With the expected high longitudinal person nonresponse rate of at least 30% (a result of longer SIPP panels), the need to minimize nonresponse bias has become increasingly important to SIPP. Nonresponse bias is computed using the following formula (assume that a person is a sample unit):

$$\text{Bias} = \sum_{i=1}^k P_i [1 - R_i][\bar{Y}_{i(R)} - \bar{Y}_{i(NR)}]$$

- where K = Total number of nonresponse adjustment cells, and
- P_i = Proportion of sample units falling in the i^{th} nonresponse adjustment cell, and
- R_i = Response rate of sample units falling in the i^{th} nonresponse adjustment cell, and
- $\bar{Y}_{i(R)}$ = Mean of the characteristic of interest of the population in the i^{th} cell which did respond in the sample, and,
- $\bar{Y}_{i(NR)}$ = Mean of the characteristic of interest of the population in the i^{th} cell which did not respond in the sample.

This equation states that the amount of bias depends on the response rate and the difference in the mean values of the characteristics for respondents and nonrespondents.

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.

Bias increases with a small response rate even if the difference in the means of respondents and nonrespondents is small. For additional background regarding nonresponse bias please refer to Singh and Petroni (1988).

B. Research Data

For this research, we used first calendar year data from three previous SIPP panels (1988, 1990, and 1991) to form redesigned longitudinal nonresponse cells. We used first calendar year data because this research focuses on longitudinal nonresponse adjustment for a calendar year instead of the entire panel (which extends beyond one calendar year). This data was used to create a combined person level file because the longitudinal nonresponse adjustment occurs at the person level. For additional background regarding SIPP longitudinal weighting procedures please refer to King (1990).

Only those persons eligible for first calendar year (from their respective panel) nonresponse adjustment were included on the combined file. All persons, including children, on the file who were first wave interviews but nonrespondents during a subsequent month were classified as nonrespondents. Persons in interviewed sample households, who were classified as type Z nonrespondents at the first wave of the panel, were also considered nonrespondents. Type Z nonresponse occurs when a member of an interviewed household is not interviewed and a proxy interview is not obtained. All other persons on the three panel combined file were considered respondents.

C. Variables

In order to create redesigned longitudinal nonresponse cells for the 1996 panel, seven original SIPP variables and three new SIPP variables were used to classify all persons. The original seven variables were used in the longitudinal nonresponse adjustment portion of previous SIPP panels. The original variables were thought to be correlated to SIPP characteristics of interest. These seven variables and their categories for nonresponse cell determination are:

1. Average monthly household income (<\$1,200, \$1,200-\$3,999, >=\$4,000).
2. Employment status (Self-employed, others).
3. Type of income (welfare etc., unemployment compensation, others).
4. Assets (bonds, other).
5. Education level (<12 years, 12-15 years, >=16 years).
6. Race and origin (white and not Spanish, other).
7. Labor force status (in labor force, not in labor force).

In addition, three redesign variables were selected based on previous research concluding they have the potential to improve SIPP nonresponse adjustment. For additional background regarding the selection of these

redesign variables please refer to Rizzo, Kalton, and Brick (1994). The variables and their categories are:

1. Census sub-region (1, 2, 3, 4, 5, 6, 7, 8, 9).
2. Number of imputations (0, 1, >=2).
3. Within PSU poverty stratum (poverty, nonpoverty).

Census sub-region was determined by categorizing states based on the following geographic locations; Pacific, Mountain, West North Central, West South Central, East North Central, East South Central, Middle Atlantic, South Atlantic, and New England. Number of imputations refers to the number of item imputations for a given person. We approximated within PSU poverty strata by comparing average monthly household income over the 12 month period to 150% of the average poverty threshold over the 12 month period. Poverty threshold depends on the number of persons and the number of children in a household.

D. Nonresponse Cell Creation

The ratio of household income over poverty threshold was selected as the nonresponse bias test criteria for determining the redesigned longitudinal nonresponse cells. This ratio was chosen because poverty is a critical statistic for SIPP and is defined at the household level. More importantly, previous research has shown that poverty appears to be associated with nonresponse.

In order to determine nonresponse cells, a specialized set of computer programs was used. The programs were designed to determine optimal cell breakdowns based on nonresponse bias.

The program formed nonresponse cells as a tree structure. The cells were partitioned by the variable minimizing absolute bias of the percent of household income to poverty threshold ratio. Specifically, nonresponse bias was computed locally within the tree structure. The nonresponse bias of a parent cell was compared to the combined cell nonresponse bias of each of the possible splits under the parent cell in order to determine the optimal partitioning.

Cell partitioning was stopped if the addition of another variable failed to further reduce the nonresponse bias. However, cells were partitioned which contained more than 10% of the data set even if nonresponse bias increased. In addition, categories were collapsed within variables if the unweighted number of nonrespondents or respondents within the cell was less than 30. Wave 1 initial weights (from the respective panel) were used in forming the new nonresponse cells.

III. Results

A. Nonresponse Cell Analysis

149 redesigned longitudinal nonresponse cells were formed as a result of processing the combined person level file through the set of specialized programs. In comparison, longitudinal nonresponse adjustment for SIPP panels prior to the redesigned 1996 panel contained

132 nonresponse cells. For the remainder of this paper, we will refer to the original 132 cells as the “original longitudinal nonresponse cells” and to the 149 cells as the “redesigned longitudinal nonresponse cells”. As previously discussed, the original nonresponse cells were formed using seven classification variables instead of ten.

The total nonresponse bias (for the ratio of household income over poverty threshold) for the redesigned nonresponse cells, when summed across all cells, was 1.44. In order to evaluate the adequacy of the redesigned cells, we used the same combined person file to compute total nonresponse bias for the original nonresponse cells (this required additional programs using the seven original classification variables). The resulting total nonresponse bias for the original nonresponse cells was 1.84.

We expected the redesigned cell total nonresponse bias to be lower due to the nonresponse bias minimization techniques used to form the redesigned cells. The difference in these two total nonresponse biases provided sufficient motivation to conduct a cross validation of the 149 redesigned longitudinal nonresponse cells.

B. Nonresponse Cell Evaluation

In order to conduct a cross validation of the redesigned longitudinal nonresponse cells, a SIPP panel other than those used for cell creation was required. Therefore, we selected the 1992 SIPP panel for cell evaluation.

To determine if the redesign nonresponse cells were performing well on the 1992 panel, we selected six SIPP estimates for evaluation. We used the estimates to compare the total nonresponse bias for the redesign nonresponse cells versus the original nonresponse cells. The six estimates follow:

1. The ratio of average monthly household income / average monthly poverty threshold.
2. Average monthly household income.
3. Average monthly person income.
4. Percent of persons in poverty.
5. Percent of persons receiving AFDC.
6. Percent of persons receiving food stamps.

Data from the twelve months of calendar year 1992 was used to determine averages. Any monthly occurrence of poverty during calendar year 1992 was sufficient to classify an individual as in poverty. The same criteria held for AFDC and food stamp reciprocity.

We developed additional programs to calculate total nonresponse bias for the original nonresponse cells and the redesigned nonresponse cells. Using calendar year 1992 data, these biases were calculated for each of the six evaluation estimates.

In addition, replicate weights were used to calculate variances in order to determine if observed differences in nonresponse bias were significant. We used a file of 100 replicate factors, developed for the 1992 SIPP panel, to create a set of 100 replicate weights. Using the replicate

weights, we determined a replicate variance for the difference in redesigned nonresponse bias and original nonresponse bias for each of the estimates. The formula for computing the test statistic for each of the six estimates follows:

$$\text{Test statistic} = \frac{[\text{Nonresponse Bias (original cells)} - \text{Nonresponse Bias (redesigned cells)}] / [SE(\text{Nonresponse Bias (original cells)} - \text{Nonresponse Bias (redesigned cells)})]}{1}$$

Table 1 contains original cell nonresponse bias, redesigned cell nonresponse bias, replicate variance for the difference, and the test statistic for each of the six estimates.

Results from this analysis show a significant (at the 95% confidence level) total nonresponse bias difference for percent of persons in poverty, percent of persons receiving AFDC, and percent of persons receiving food stamps. For each of these estimates, redesigned nonresponse cell total bias is significantly lower than original nonresponse cell total bias. No significant difference was observed for the remaining three estimates.

IV. CONCLUSION

From table 1, we can conclude that the redesigned longitudinal nonresponse cells performed significantly better than the original longitudinal nonresponse cells in terms of total nonresponse bias. The results indicated a successful bias reduction. This is shown by the significant differences in percent of persons in poverty, percent of persons receiving AFDC, and percent of persons receiving food stamps.

The three significant differences provide sufficient motivation to use the redesigned longitudinal nonresponse cells beginning with the redesigned 1996 SIPP panel.

Due to the increase in the number of interviews for future SIPP panels, we expect person nonresponse to continue to be a critical longitudinal issue. Therefore, the importance of similar research in order to reduce nonresponse bias is apparent. Specifically, longer time periods can be used for research leading to the determination of nonresponse cells for the entire panel. In addition, the continuation of research to determine new SIPP variables for nonresponse cell classification is necessary because societal characteristics can change over time.

References

- King, K. (1990), “SIPP 1987: Specifications for Panel File Longitudinal Weighting of Persons.” Internal U.S. Bureau of the Census Memorandum from Waite to Courtland, June 1, 1990.
- Rizzo, L., Kalton, G., and Brick, M. (1994). “Weighting Adjustments for Panel Nonresponse in the Survey of Income and Program Participation.”
- Siegel, P. (1995). “SIPP 96+: Creation and Evaluation of Nonresponse Cells for the Second and Subsequent Waves.” Internal Census Bureau

Memorandum for Documentation, June 13, 1995.

Singh, R. and Petroni, R. (1988). "Nonresponse Adjustment Methods for Demographic Surveys at the U.S. Bureau of the Census." SIPP Working Paper No. 8823, U.S. Bureau of the Census.

Weinberg, D. and Petroni, R. (1992). "The Survey of Income and Program Participation in the 1990's." SIPP Working Paper No. 9206, U.S. Bureau of the Census.

Table 1. Replicate Variance Tests for difference between Original Cell Nonresponse Bias and Redesigned Cell Nonresponse Bias

	Original Cell Nonresponse Bias	Redesigned Cell Non-response Bias	Replicate Variance	Test Statistic
Average Monthly Household Income/Average Poverty Threshold	1.4884	1.4829	8.53×10^{-5}	0.5953
Average Monthly Household Income	105.057	104.93	11.3058	0.0378
Average Person Income	45.1724	45.01	2.6814	0.0992
Percent Persons in Poverty	-1.8594	-1.6306	1.63×10^{-5}	-5.6591
Percent Persons Receiving AFDC	-0.1101	-0.0656	3.86×10^{-6}	-2.2678
Percent Persons Receiving Food Stamps	-0.1587	-0.0384	1.17×10^{-5}	-3.5204