Survey Methodology, June 1996
Vol. 22, No. 1, pp. 43–53
Statistics Canada

A Comparison of Some Weighting Adjustment Methods for Panel Nonresponse

LOU RIZZO, GRAHAM KALTON and J. MICHAEL BRICK[1]

ABSTRACT

In some surveys, many auxiliary variables are available for respondents and nonrespondents for use in nonresponse adjustment. One decision that arises is how to select which of the auxiliary variables should be used for this purpose and another decision involves how the selected variables should be used. Several approaches to forming weighting adjustments for nonresponse are considered in this research. The methods include those based on logistic regression models, categorical search algorithms, and generalized raking. These methods are applied to adjust for panel nonresponse in the Survey of Income and Program Participation (SIPP). The estimates from the alternative adjustments are assessed by comparing them to one another and to benchmark estimates from other sources.

KEY WORDS: Nonresponse bias; Panel surveys; Generalized raking; Benchmark estimates.

1. INTRODUCTION

Weights are commonly used in the analysis of survey data to compensate for unequal selection probabilities of the sampled elements, to compensate for unit nonresponse, and to make the weighted sample distributions for certain variables conform to known population distributions for those variables (thereby aiming to compensate for noncoverage and to improve the precision of the survey estimates) (Kish 1992). Corresponding to these three objectives, the weights are usually developed in three stages. First, a base weight is calculated for each sampled element as the inverse of the element's selection probability. Second, the base weights of responding sampling elements are multiplied by a nonresponse weight to compensate for the nonrespondents. Third, the adjusted weight is modified to make the weighted sample distributions for certain variables conform to external information on these distributions.

This paper deals with the nonresponse adjustment weights that attempt to compensate for unit nonresponse. A commonly used procedure for obtaining these weights is to divide the total sample into a set of weighting classes based on information known for both respondents and nonrespondents, and then to increase the base weights for the respondents in a weighting class to represent the nonrespondents in that class (Oh and Scheuren 1983; Kalton 1983). In many surveys little information is known about the nonrespondents, beyond the primary sampling units and strata from which they come. In this case, the choice of possible weighting classes is limited, and the procedure can be applied fairly straightforwardly.

In some surveys, however, there is an extensive amount of information available for the nonrespondents. This information may be available from the sampling frame (*e.g.*, when sampling employees from personnel files) or by matching sampled elements with administrative records. Also, in panel surveys and other surveys involving more than one stage of data collection, extensive information on nonrespondents at later stages is available from their responses at the early stages.

The major focus of this research is on methods for developing weighting adjustments for nonresponse when a large number of characteristics of the nonrespondents are known. In this situation, decisions about methods of adjusting for nonresponse involve selecting which auxiliary variables will be used and how they will be used to make the adjustments.

The main ideas are presented in this article by applying several different adjustment procedures in a specific panel survey, the Survey of Income and Program Participation (SIPP). The SIPP is an ongoing household panel survey conducted by the U.S. Bureau of the Census. The nonrespondents to a SIPP panel can be separated in two groups: those who fail to respond at the initial wave of data collection (initial wave nonrespondents), and those who respond at the initial wave but fail to respond at one or more of the subsequent waves of the panel for which they are eligible (panel nonrespondents). For the latter group, extensive information from the initial wave of data collection can be utilized in adjusting for panel nonresponse. The weighting adjustments studied here relate to the panel nonrespondents only. These adjustments modify the weights of panel respondents (*i.e.*, those who provide data for all waves for which they are eligible) to compensate for the panel nonrespondents.

In the SIPP, a national probability sample of households is interviewed each year, and all the adults aged 15 and over living in those households at the initial wave become panel members who are followed for the duration

[1] Lou Rizzo, Graham Kalton and J. Michael Brick, Westat Inc., 1650 Research Blvd., Rockville, MD 20850, U.S.A.

of the panel. Until now SIPP panels have had a lifetime of 2 ⅔ years, but this is being increased with the 1996 panel to 4 years. Interviews are conducted with panel members at four-month intervals to collect data about income amounts received, participation in income maintenance programs, and other factors that may affect their income and economic welfare. Data are also collected about children. See Nelson, McMillen and Kasprzyk (1985) and Jabine, King and Petroni (1990) for further information on the SIPP design.

The investigation reported here was conducted with the 1987 SIPP panel, using the panel's public use data file. That panel started with a sample of about 12,300 households and followed panel members for seven waves of data collection. The household nonresponse rate at the initial wave was 6.7 percent (Jabine *et al.* 1990). Including children, 30,841 individuals were living in the responding households at the initial wave. Of these individuals, 20.8 percent failed to provide data for all waves for which they were eligible, *i.e.*, they were panel nonrespondents.

In addition to selecting auxiliary variables and studying alternative methods of using those variables to form weighting adjustments for panel nonresponse, this research includes a comparative evaluation of the procedures. The evaluation is performed by comparing a range of estimates produced with the alternative methodologies with one another and with benchmark estimates. The final section of this article summarizes the results and draws conclusions about the effectiveness of the alternative weighting schemes investigated. Further details are given by Rizzo, Kalton, and Brick (1994).

2. PREDICTORS OF RESPONSE PROPENSITY

The first step in developing panel nonresponse adjustments is deciding which of the large number of items available from the first wave of data collection should be selected for use in the adjustment procedures. That selection is the focus of this section. The approach adopted is to choose items with responses that discriminate persons by their likelihood to respond at all later waves. Little (1986) calls this method a response propensity stratification method and shows that the large sample bias of estimates can be reduced by adjusting the base weight by the inverse of the probability that an element responds.

In the 1987 SIPP panel, there were 58 items available from the initial wave of data collection (Wave 1) that could be used as potential explanatory variables for panel nonresponse. All of the items used currently by the Bureau of the Census for the SIPP panel nonresponse adjustment were part of this set of 58, with the exception of the Metropolitan Statistical Area (MSA) status, which was suppressed from the public use data file because of disclosure concerns.

With panel response status (panel respondent *vs.* panel nonrespondent) as the dependent variable, logistic regression analysis was viewed as a natural method for selecting a model for panel nonresponse. However, before attempting this modeling, an initial screening of the variables was performed to reduce the large number of variables to a more manageable set. As a general guideline, items were retained for the logistic regression analysis if the difference in response rates between any two categories for the item was both statistically significant and at least four percentage points. For a variety of reasons, some items were retained even if they did not meet these requirements. For example, the difference in the panel response rates for males and females was less than 2 percent, but gender was nevertheless used in some subsequent analyses.

The screening process reduced the number of items for the logistic regression analysis from 58 to 31. The items retained were: tenure, public housing, household type, Census region, household education, household size, household income, whether householder holds financial instruments (bonds), gender, race, Hispanic origin, relationship to reference person (RRP), age, marital status, family type, education, student status, whether laid off work, personal income, whether holds multiple jobs, working class, whether a recipient of Medicare benefits, Medicaid, Women, Infants, and Children (WIC), Aid to Families with Dependent Children (AFDC), food stamps, general assistance, Social Security, other welfare, Veteran's status, and the number of imputed items at Wave 1.

The last item, the number of imputed items, was included as an index of cooperation at Wave 1. Other studies have found that individuals who are less cooperative at the initial wave of a panel survey are more likely to be nonrespondents at later waves (see, for example, Kalton, Lepkowski, Montanari and Maligalig 1990). As described below, this index turned out to be highly related to panel nonresponse.

2.1 Logistic Regression Analysis

Since all 31 items identified in the screening analysis were at least marginally correlated with panel nonresponse, they are all candidate variables for use in a weighting adjustment scheme to reduce the panel nonresponse bias in the survey estimates. However, the screening analysis was limited because it did not consider the interrelationships between the items and it retained too many variables for practical use in making the panel nonresponse adjustments. For example, two items that are highly associated with response status might also be highly correlated with each other, so that the use of one of the two might be sufficient in making the adjustments. To address this issue, the next step in selecting predictors of panel nonresponse was to investigate which combinations of the items could best predict panel response status.

Table 1
Parameter Estimates for the Logistic Regression Model

Predictors	Parameter Estimate
Intercept	−0.465
Age (χ^2 = 184.9, *p*-value < .0001).	
< 16	−0.179
16-24	0.446
25-50	0.187
51-71	−0.056
> 71	0.0
Race (χ^2 = 214.0, *p*-value < .0001).	
White	−0.351
Black	0.255
Other	0.0
RRP (χ^2 = 69.0, *p*-value < .0001).	
Family member	−0.251
Nonfamily member	0.0
Census region (χ^2 = 327.3, *p*-value < .0001).	
New England	0.009
Mid Atlantic	0.167
South Atlantic	0.027
East South Central	−0.231
North Central	−0.396
Mountain/West South Central	0.425
Pacific	0.0
Tenure (χ^2 = 207.2, *p*-value < .0001).	
Home owner	−0.154
Renter	0.331
Other	0.0
Items imputed (χ^2 = 434.2, *p*-value < .0001).	
0	−0.626
1	−0.244
2 to 3	0.296
> 3	0.0
Bond status (χ^2 = 97.1, *p*-value < .0001).	
No bonds	0.168
Some bonds	0.0
Layoff (χ^2 = 33.4, *p*-value < .0001).	
Not laid off	−0.179
Laid off	0.0
Food stamps (χ^2 = 39.3, *p*-value < .0001).	
Not recipient	−0.191
Recipient	0.0
Class of work (χ^2 = 31.4, *p*-value < .0001).	
Business	0.100
Other	0.103
Government	0.0
Education (χ^2 = 12.8, *p*-value = .0003).	
Last grade tenth or eleventh	−0.075
Other	0.0
Household income (χ^2 = 14.9, *p*-value = .0006).	
Less than $1,200/month	0.117
$1,200-$8,000/month	−0.088
Greater than $8,000/month	0.0
Gender (χ^2 = 10.3, *p*-value = .0013).	
Male	0.047
Female	0.0
RRP-Age < 16 Interaction (χ^2 = 10.1, *p*-value = .0015).	
Family member, child	0.096
Other	0.0

A logistic regression approach was used to the examine the joint relationships of several items with panel response status. The regression models were fitted using the Wave 1 survey weights that accounted for unequal selection probabilities and initial wave nonresponse. After examining a number of possible models, a model with thirteen main-effect variables and one interaction term was selected as a reasonable representation of the data.

Table 1 presents the parameter estimates for each level of each predictor variable in this model, together with Wald (χ^2) statistics for each predictor variable. The parameter value of the last level of each predictor variable (the benchmark level) is set to zero. The parameter estimates for the remaining levels of each predictor variable represent differences in response propensity from the benchmark level. As can be seen from the Wald statistics, all the predictor variables make highly significant contributions to the model.

A notable feature of this model is that it contains only one interaction term, the relationship to reference person/age under 16 interaction. All other interactions investigated had smaller χ^2 values than this one. Even the relationship to reference person/age under 16 interaction has a relatively low predictive power. In fact, this interaction and the last three predictor variables in Table 1 (education, household income, and gender) were not included in most of the weighting procedures discussed below because of their limited predictive power for panel response status. The weighting procedures are mostly based on a reduced main-effects model comprising the first ten predictor variables listed in Table 1.

3. ALTERNATIVE WEIGHT ADJUSTMENTS

The method used in the SIPP to adjust the weights for panel nonresponse is described by Chapman, Bailey, and Kasprzyk (1986). The method basically consists of forming nonresponse adjustment cells and then adjusting the weights by the inverses of the response rates in the cells. The cells are formed by the cross-classification of the responses from a set of Wave 1 variables thought to be correlated with panel response. Small cells are combined so that the resulting sample size in each collapsed cell is 30 or more. The reciprocal of the observed (weighted) response rate in each collapsed cell is the panel nonresponse adjustment for that cell. The panel nonresponse adjustment is then multiplied by the Wave 1 weight to create a nonresponse adjusted weight. The Wave 1 weight includes an adjustment for Wave 1 nonresponse, but it does not include the Wave 1 poststratification adjustment.

This section examines alternative methods for performing the panel nonresponse adjustments. These methods can be categorized into three groups:

- Logistic regression methods.
- CHAID methods.
- Generalized raking methods.

Each of the alternative approaches to nonresponse adjustment is discussed below. The procedures for developing the weighting adjustments are detailed along with important statistical properties of the adjustments.

3.1 Adjustments Based on Logistic Models

The first set of weighting adjustments we discuss is developed directly from the logistic regression model described in the previous section. This panel nonresponse weighting adjustment, called the *predicted logistic adjustment*, was computed by taking the inverses of the response rates predicted from the reduced main-effects logistic regression model for each of the cells in the crossclassification of the ten predictor variables in that model.

Since the parameters for computing the predicted response rates are estimated with a main-effects model from the marginal responses for the variables, the small sample sizes in the cells of the crossclassification of all the variables are not a concern. However, this benefit is gained by relying completely on the validity of the main-effects model, that is, by assuming that there are no interactions between the variables that need to be taken into account.

One approach to placing less reliance on the main-effects model is to base the adjustments on the observed response rates in cells that have sample sizes large enough to ensure the stability of the observed response rates and to base the adjustments on the predicted response rates in other cells. The second member of the class of alternative adjustments based on logistic regression uses this mixed strategy. In cells containing 25 or more sample persons, the nonresponse adjustment is the inverse of the observed cell response rate. In cells containing less than 25 sample persons, the nonresponse adjustment is the inverse of the predicted response rate for the cell. This adjustment is called the *mixed logistic adjustment*.

A third logistic nonresponse adjustment studied is similar to the current SIPP procedures. Initial cells were defined by the crossclassification of the ten independent variables used in the logistic regression. The cells were then collapsed until the sample size in each cell exceeded 30, and the inverse of the observed response rate within a collapsed cell was then used as the nonresponse adjustment. The strategy for collapsing cells was to group together cells with similar predicted response rates. This nonresponse adjustment is called the *collapsed logistic adjustment*. Although this adjustment is similar to the current SIPP panel nonresponse adjustment, there are some differences in the variables used to define the cells and the methods used to combine small cells are different.

For all three alternative weighting adjustments based on the logistic regression model, the observed and predicted response rates were computed from weighted counts of the number of cases rather than using the unweighted numbers, where the weights were the nonresponse adjusted Wave 1 weights. In practice, the weighted and unweighted adjustments were nearly the same.

3.1.1 Adjustments Based on CHAID Models

The second class of methods for adjusting for panel nonresponse involved using the CHAID categorical search algorithm to divide the data set into adjustment cells. The general approach was to define adjustment cells as combinations of responses to the predictor variables that had the greatest discrimination with respect to panel response rates, subject to the restriction that each cell should have a minimum sample size of at least 25 persons. The panel nonresponse adjustment was the inverse of the observed response rate in the cell.

The CHAID algorithm creates cells by splitting the data set progressively in a tree structure. The splitting along each newly created branch is performed by choosing the variable that maximizes a χ^2 criterion. When the split involves a polychotomous variable, the split may involve several branches. The χ^2 tests are modified using Bonferroni type adjustments to prevent variables from being chosen simply because they have more categories. CHAID is one version of the Automatic Interaction Detector (AID) developed for categorical variables. Kass (1980) presents the theory underlying the CHAID technique. Another version of the same methodology was used by Lepkowski, Kalton and Kasprzyk (1989) and Kalton, Lepkowski and Lin (1985) to model nonresponse in SIPP.

For the current analysis, two CHAID models were examined by including different sets of predictor variables. The first model included the seven most important predictors in the logistic regression model (age, relationship to reference person, race of householder, tenure, Census region, imputation flags, and bond-holding status), plus gender. This model resulted in 99 nonresponse adjustment cells. The nonresponse adjustment based on this model is called CHAID 1. The second CHAID model included the 13 predictor variables from the logistic regression model presented in Table 1. This model resulted in 142 nonresponse adjustment cells. The nonresponse adjustment for this model is called CHAID 2.

3.1.2 Adjustments Based on Generalized Raking

The third class of methods examined for adjusting for panel nonresponse was generalized raking. Unlike the other approaches, nonresponse adjustment cells were not developed by crossclassifying the predictor variables. Rather, raking was directly applied to force the panel respondents' marginal distributions for each of the predictor variables (computed using the adjusted weights) to equal the corresponding distributions for respondents and nonrespondents combined (computed using the original Wave 1 weights). Kalton and Kasprzyk (1986) refer to this method as sample based raking. The ten predictor variables from the reduced logistic regression model were used to define the marginal distributions. Hence, the raking problem was ten dimensional, with one dimension for each predictor variable.

Raking involves modifying the original weights in order to satisfy certain marginal constraints while minimizing the distance between the original and adjusted weights. Deville and Särndal (1992) describe some distance functions that may be used and derive the corresponding raking methodologies. The raking algorithm of Deming and Stephan (1942), which implicitly employs a distance function that leads to a multiplicative solution, is one form of generalized raking.

The CALMAR software described by Deville, Särndal and Sautory (1993) was used to compute the adjustments. Three different distance functions were examined: the multiplicative method, the linear method, and the truncated multiplicative method. The adjustments for all three distance functions were found to be nearly identical. This empirical result is consistent with results given by Deville and Särndal (1992) that show that the estimators using weights generated with different distance functions are asymptotically equivalent if the distance functions satisfy certain smoothness conditions. The three distance functions employed in this research satisfy those conditions. Since the adjustments were nearly identical for all three methods, only the weighting adjustment from the multiplicative method was retained for further evaluation. The resulting adjustment is called the *raking* adjustment.

3.1.3 Distributions of Nonresponse Adjustments

The adjustments for each of the six schemes described above were computed for the 1987 SIPP panel file. Table 2 summarizes the distributions of the resulting nonresponse adjustments. The summary is for the adjustments only, not the weights that are the products of the adjustments and the Wave 1 weights. Table 2 is divided into two parts: the upper part shows the mean, median, and extreme values for each adjustment distribution, as well as $(1 + CV^2)$, where CV is the coefficient of variation for each adjustment. The statistic $(1 + CV^2)$ serves as an indicator of the increase in variance of the estimates introduced by having variable nonresponse adjustment factors (see Kish 1992). The second part of Table 2 shows the correlations among the alternative forms of adjustment.

Since the overall weighted panel response rate is 0.794, the mean overall nonresponse adjustment would be $1/(0.794) = 1.26$ if the same adjustment were used for all persons. The mean weighting adjustments for the three weighting adjustments that use the inverses of cell response rates (collapsed logistic, CHAID 1 and CHAID 2) are necessarily equal to the overall nonresponse adjustment of 1.26. The mean weighting adjustments for the other schemes differ only minimally from the mean overall nonresponse adjustment.

For all six schemes, the distributions are positively skewed, with a few cases with large weights. By their nature, the various logistic and CHAID schemes cannot have adjustments less than 1.00, whereas the raking algorithm can, and does, do so. The median weights are similar among all schemes, but the maximum weights are not. The CHAID 2 scheme has a cell with a response rate of only 7 percent, leading to the largest maximum weight of 13.93. The raking scheme has the smallest maximum weight of 2.51.

Table 2

Distribution of Panel Nonresponse Adjustments

	Mean	Minimum	Median	Maximum	$1 + CV^2$
Predicted logistic	1.26	1.04	1.20	4.28	1.02
Mixed logistic	1.26	1.00	1.20	4.28	1.03
Collapsed logistic	1.26	1.00	1.20	3.43	1.02
CHAID 1	1.26	1.02	1.22	3.49	1.03
CHAID 2	1.26	1.01	1.19	13.93	1.04
Raking	1.26	0.91	1.23	2.51	1.02

Correlations

	Predicted Logistic	Mixed Logistic	Collapsed Logistic	CHAID 1	CHAID 2	Raking
Predicted logistic	1.00	0.96	0.73	0.73	0.63	0.95
Mixed logistic		1.00	0.73	0.72	0.63	0.90
Collapsed logistic			1.00	0.69	0.58	0.75
CHAID 1				1.00	0.81	0.73
CHAID 2					1.00	0.63
Raking						1.00

The values of $(1 + CV^2)$ are fairly consistent across the various adjustments. The CHAID 2 adjustment has the greatest value of $(1 + CV^2)$, primarily because of the presence of more outlying adjustments (such as the maximum value of 13.93). However, even for this method, the approximate increase in the variance of the survey estimates is only four percent. The raking adjustment has the smallest increase in variance (two percent), but this increase is not very different from that of the other methods.

The pairwise correlations between the six alternative sets of weights range from 0.58 to 0.96. Not surprisingly, the predicted logistic and mixed logistic weights are highly correlated. Given the similarity of the predicted main-effects logistic regression scheme to raking, it is also not surprising that their two sets of weights are highly correlated. The relatively high correlation between the raking weights and the CHAID 1 weight and the collapsed logistic weight is consistent with the earlier result showing no large interaction terms. The CHAID 2 weights have the lowest correlations with the other sets of weights, except for their correlation with the CHAID 1 weights. This finding is probably explained by the wide variability in the CHAID 2 weights resulting from the use of as many as 142 adjustment cells.

3.2 Final Panel Weights

The panel nonresponse adjustment weights discussed in the previous section represent the adjustments to the Wave 1 weights to compensate for panel nonresponse. The final panel weights that may be used in the analysis of the SIPP panel file are obtained by multiplying the panel nonresponse adjustment weights by the Wave 1 weights, and then applying poststratification to make weighted sample totals conform to totals derived primarily from the Current Population Survey (CPS). This procedure was applied for each of the six alternative panel nonresponse adjustment schemes.

The poststratification procedure used was equivalent to the current SIPP procedure, except that the latter procedure poststratifies by rotation groups whereas for the alternative weighting schemes the poststratification was performed on all rotation groups combined. The difference should not have an appreciable effect. After poststratification, the six alternative sets of final weights and the SIPP panel weights sum to the same control totals.

To compare the final panel weights for the six adjustment schemes with one another and with the current SIPP panel weight, the correlations between the weights were computed, along with the measure of variability used previously, $(1 + CV^2)$. The results are presented in Table 3. The estimates of the variability due to the weighting $(1 + CV^2)$ indicate similar increases of between 8 and 10 percent in the variances of survey estimates for all of the weighting schemes. The correlations between the alternative sets of final panel weights are all 0.85 or higher. Comparing these correlations to those in Table 2, it is clear that the correlations between the final weights are appreciably higher than those between the panel nonresponse adjustment weights. The correlations between the SIPP panel weight and the alternative final weights are consistently lower than any others, probably because the variables used in forming the nonresponse adjustments for this weight differed from those used for the alternative weights. The variables used in the alternative schemes that are not used in the SIPP panel weight are age, relationship to reference person, number of imputed items, class of work, and food stamp recipiency. Household size is the only variable other than MSA status (which was not available due to disclosure concerns) used in the SIPP panel weight but not used for the alternative schemes because it was not found to be significantly associated with response rates.

Table 3

Correlations Between Poststratified Weights with Variance Inflation Measures

	SIPP panel	Predicted Logistic	Mixed Logistic	Collapsed Logistic	CHAID 1	CHAID 2	Raking
SIPP panel	1.00	0.75	0.74	0.75	0.71	0.68	0.77
Predicted logistic		1.00	0.99	0.91	0.90	0.86	0.98
Mixed logistic			1.00	0.91	0.90	0.86	0.97
Collapsed logistic				1.00	0.89	0.85	0.93
CHAID 1					1.00	0.94	0.91
CHAID 2						1.00	0.87
Raking							1.00
$1 + CV^2$	1.08	1.09	1.09	1.08	1.09	1.10	1.08

4. COMPARING ESTIMATES USING ALTERNATIVE WEIGHTS

The previous section described the development of the alternative sets of final weights that may be used for the analysis of the SIPP panel file. All the final weighting schemes incorporate adjustments for unequal selection probabilities, nonresponse at the initial wave, panel nonresponse, and poststratification to external control totals. This section compares survey estimates obtained using the alternative weighting schemes with one another and with the corresponding estimates obtained using the SIPP panel weights. In addition, where possible, the various survey estimates are also compared with external estimates from other sources. Some of the external estimates are benchmark estimates obtained from administrative records or the Current Population Survey. Other external estimates are obtained from Wave 1 of the 1989 SIPP panel. Data collected in Wave 7 of the 1987 SIPP panel relate to the same time period as data collected in Wave 1 of the 1989 SIPP panel, and hence estimates obtained from these two data sources should be comparable.

In making comparisons with benchmark estimates, it needs to be recognized any differences observed may be explained by a variety of factors of which panel nonresponse is only one. For example, response errors and differences in definitions may explain differences between SIPP estimates and benchmark estimates. Thus the benchmark comparisons need to be treated with caution. Since the 1989 SIPP panel estimates are based on Wave 1 data, they are not subject to the panel nonresponse. Thus, differences between estimates obtained from the 1987 and 1989 SIPP panels are perhaps the most likely to be caused by a failure of the panel nonresponse adjustments to fully compensate for panel nonresponse bias. However, even in this case, alternative explanations such as panel conditioning could contribute to the differences (although Pennell and Lepkowski 1992, show that panel conditioning is not a major factor in most SIPP estimates).

Table 4 presents a variety of estimates from the 1987 SIPP panel file using the SIPP panel weight and the six alternative weighting schemes, and corresponding benchmark estimates and estimates from the 1989 SIPP panel where available. The estimates are percentages, except for the estimates of the mean number of months without health insurance, median household income, and annual wages. The estimates are for the total population, except for the employment estimates (percent employed, unemployed and out of the labor force), which are for persons over the age of 15, and for annual wages, which are for persons over the age of 14. The estimates are for three different time periods: June 1987, January 1989, and the calendar year of 1987. For example, the first three estimates in Table 4 are the estimated percentages of persons participating in the AFDC (Aid for Families with Dependent Children) program in June 1987, in January 1989, and at any time during the 1987 calendar year. A comparable estimate from the 1989 SIPP panel is available only for the January 1989 time period.

The most notable finding from Table 4 is the similarity of the estimates computed with all the weighting schemes from the 1987 panel. The percentage estimates in Table 4 are in fact given to two decimal places because the use of the conventional one decimal place would often show no difference between the alternative estimates. The largest difference occurs for the percentage employed in January 1989, where the estimate using the SIPP panel weight is 62.7 percent and the estimate using the mixed logistic regression weight is 62.3 percent. Even this largest of differences is relatively small, especially when considering that the estimated standard error for this estimate is 0.3 percent.

When the 1987 SIPP panel estimates are compared with the external estimates from the 1989 SIPP panel and from other sources, some of the differences are much larger and of substantive importance. To examine these differences in more detail, standardized differences between the alternative estimates and the benchmark estimates were computed and are shown in Table 5. A standardized difference is defined as the difference between the alternative estimate and the external estimate divided by the standard error of the difference.

The upper part of Table 5 shows the standardized differences when the 1989 SIPP panel is used to produce the external estimate. The standardized differences for most of the estimates are less than 2.0 in absolute value, indicating that the differences may be accounted for by sampling error. However, the standardized differences for the percentage unemployed and for the poverty rate are greater than 2.0 and highly significant. Thus, the alternative weighting adjustments do not succeed in bringing the 1987 survey estimates in line with the 1989 survey estimates for all characteristics.

The lower part of Table 5 shows the standardized differences when other benchmark estimates are used. These standardized differences are generally large and in many cases very large. Only a few are less than 2.0 and many are greater than 10.0. Given the much smaller standardized differences found in the upper part of Table 5 for similar statistics, it seems likely that factors other than panel nonresponse bias are largely responsible for the magnitude of these differences. The standardized differences based on these largely administrative data sources may signal important issues related to the quality of the data (from either the SIPP, the benchmark data source, or both), but they do not provide much help in assessing the effectiveness of alternative nonresponse adjustments in reducing panel nonresponse bias.

Table 4

Estimates for the Total Population from the 1987 SIPP Panel with Alternative Weighting Schemes and Estimates from Other Sources

	SIPP Panel	Predicted Logistic	Mixed Logistic	Collapsed Logistic	CHAID 1	CHAID 2	Raking	1989 SIPP Panel	Bench-mark
AFDC – June 1987	3.73	3.70	3.74	3.72	3.71	3.60	3.69		4.28[1]
AFDC – January 1989	3.10	3.12	3.14	3.12	3.14	3.02	3.10	3.56	4.24[2]
AFDC – Annual 1987	4.85	4.78	4.82	4.81	4.80	4.69	4.78		
Food stamps – June 1987	7.43	7.26	7.30	7.34	7.38	7.20	7.21		7.35[3]
Food stamps – January 1989	6.71	6.63	6.67	6.64	6.70	6.59	6.58	6.30	7.29[4]
Food stamps – Annual 1987	10.30	10.11	10.16	10.18	10.24	10.05	10.06		
Medicaid – January 1989	6.77	6.78	6.81	6.75	6.81	6.68	6.76	6.97	
Medicaid – Annual 1987	9.21	9.21	9.24	9.21	9.25	9.09	9.21		
SSI – June 1987	1.68	1.70	1.69	1.67	1.69	1.65	1.69		1.68[3]
SSI – January 1989	1.65	1.67	1.66	1.64	1.66	1.61	1.66	1.65	1.74[3]
SSI – Annual 1987	1.80	1.82	1.82	1.80	1.82	1.78	1.82		
Social security – January 1989	14.92	14.87	14.87	14.89	14.88	14.89	14.85	15.14	
Poverty rate – June 1987	10.88	10.75	10.79	10.76	10.79	10.69	10.74		
Poverty rate – January 1989	12.91	12.98	13.02	12.97	12.99	12.91	12.93	14.46	
Entering poverty 1987/1988	2.25	2.31	2.32	2.30	2.29	2.32	2.31		
Leaving poverty 1987/1988	2.69	2.63	2.64	2.60	2.62	2.63	2.63		
Mean months without health insurance – 1987	1.66	1.69	1.70	1.67	1.67	1.69	1.69		
Median household income – January 1989	2,601	2,600	2,597	2,607	2,607	2,607	2,602	2,550	
Annual wages 1987 (in trillions)	1.93	1.94	1.93	1.94	1.94	1.94	1.94		2.22[4]
Employed – January 1989	62.74	62.36	62.34	62.43	62.42	62.52	62.42	61.60	
Unemployed – January 1989	3.57	3.64	3.63	3.60	3.58	3.60	3.63	4.52	
Out of labor force – January 1989	33.69	34.01	34.03	33.96	34.01	33.88	33.95	33.88	
Married in 1987	1.39	1.41	1.40	1.39	1.39	1.39	1.41		1.86[5]
Divorced in 1987	0.51	0.50	0.50	0.49	0.50	0.51	0.49		0.90[6]
Changed address in 1987	12.88	13.32	13.32	13.19	13.36	13.37	13.33		17.99[6]

[1] Social Security Bulletin, Volume 52, No. 3.
[2] Social Security Bulletin, Volume 51, No. 7.
[3] USDA Food and Nutrition Service, unpublished data.
[4] U.S. Bureau of the Census, Current Population Reports, Consumer Income, P-60, No. 174.
[5] National Center for Health Statistics: Vital Statistics of the U.S., 1987, Volume III, Marriage and Divorce, DHHS Pub. No. (PHS) 91-1103.
[6] U.S. Bureau of the Census, Current Population Reports, Population Characteristics, P-20, No. 473.

Table 5

Standardized Differences Between 1987 SIPP Panel Estimates and Benchmark Estimates

	Benchmark Estimate	SIPP Panel	Predicted Logistic	Mixed Logistic	Collapsed Logistic	CHAID 1	CHAID 2	Raking
1989 SIPP panel estimates								
AFDC	3.56	-1.58	-1.52	-1.43	-1.52	-1.44	-1.84	-1.57
Food stamps	6.30	1.02	0.82	0.92	0.86	1.01	0.73	0.69
Medicaid	6.97	-0.50	-0.47	-0.40	-0.53	-0.39	-0.70	-0.51
SSI	1.65	0.05	0.11	0.08	-0.03	0.07	-0.15	0.09
Social Security	15.14	-0.38	-0.46	-0.46	-0.42	-0.44	-0.42	-0.50
Poverty rate	14.46	-2.77	-2.64	-2.57	-2.67	-2.63	-2.78	-2.74
Median Income	2,550	2.05	2.01	1.89	2.30	2.30	2.29	2.09
Employed	61.60	2.42	1.60	1.56	1.76	1.72	1.95	1.73
Unemployed	4.52	-4.93	-4.59	-4.59	-4.76	-4.90	-4.78	-4.60
Out of labor force	33.88	-0.42	0.28	0.32	0.18	0.28	-0.01	0.15
Other benchmark estimates								
AFDC - June 1987	4.28	-2.55	-2.66	-2.49	-2.59	-2.65	-3.14	-2.71
AFDC - January 1989	4.24	-5.71	-5.62	-5.49	-5.63	-5.51	-6.10	-5.70
Food stamps - June 1987	7.35	0.27	-0.31	-0.16	-0.04	0.11	-0.50	-0.48
Food stamps - January 1989	7.29	-2.04	-2.32	-2.17	-2.26	-2.06	-2.44	-2.50
SSI - June 1987	1.68	0.00	0.13	0.08	-0.03	0.08	-0.20	0.11
SSI - January 1989	1.74	-0.57	-0.48	-0.53	-0.67	-0.54	-0.84	-0.50
Annual wages 1987	2.22	-16.12	-15.94	-16.38	-15.66	-15.61	-15.60	-15.78
Married in 1987	1.86	-5.11	-4.93	-4.98	-5.11	-5.10	-5.07	-4.95
Divorced in 1987	0.90	-7.15	-7.37	-7.36	-7.40	-7.32	-7.20	-7.40
Changed address in 1987	17.99	-11.49	-10.50	-10.51	-10.80	-10.42	-10.40	-10.49

5. DISCUSSION

Nonresponse weights are widely used to compensate for unit nonresponse in sample surveys. The basic requirement for this form of weighting is the availability of information on one or more auxiliary variables for both respondents and nonrespondents. In many surveys, this information is available for only a small number of auxiliary variables (such as the PSUs and strata from which the units were selected). In such surveys, the nonresponse weights can often be simply developed as weighting class adjustments for a set of classes based on the crosstabulation of the auxiliary variables.

There are, however, surveys in which data are available for a large number of auxiliary variables for possible use in developing nonresponse weights. This situation often applies when an administrative record system is used as the survey's sampling frame, with all the information in the system then being available for use in making nonresponse adjustments. It also applies when the survey data collection is conducted in two or more phases (*e.g.*, an initial screening interview followed by a detailed interview or some other form of data collection at a later time point) and when nonresponse adjustments are needed for later phases; in this case, data from prior phases of data collection may be used in compensating for nonresponse at later phases. A similar situation applies in panel surveys when adjustments are required for nonresponse at later waves of the panel, as discussed in this paper.

When a large number of auxiliary variables is available for all sampled units, two main choices need to be made. First, there is the choice of auxiliary variables to use in the adjustment. Second, there is the choice of the adjustment method to be applied.

The basic approach adopted in this study for choosing the auxiliary variables for use in the nonresponse adjustment was to identify the set of variables that were good predictors of panel nonresponse. With so many auxiliary variables available, the first step was a screening procedure to eliminate variables that were found to have little association with the panel nonresponse rate. Then, logistic regression models using predictor variables remaining from the screening were examined to identify the set of variables to be retained for use in adjusting the weights. Whether the number of auxiliary variables is reduced to a manageable set by this or some other approach (*e.g.*, by using the CHAID algorithm), this reduction is likely to be a necessary first step when there are many potential auxiliary variables available.

After selecting the subset of auxiliary variables, a wide variety of methods exists for creating the nonresponse adjustments. We examined panel nonresponse adjustments based on logistic regression models, categorical search models, and sample-based generalized raking. The final panel weights resulting from these adjustment schemes were highly correlated with one another and they yielded estimates that were very similar. None of the schemes produced estimates that were superior in terms of bias reduction.

In part, the high correlation of the final panel weights generated by the different adjustment schemes may be explained by the similarity of many of the adjustment schemes. In part, it may be explained by the final post-stratification weighting which raised the correlations between the weights. It may also be partly explained by the lack of large interaction effects between the auxiliary variables. If there were sizable interaction effects that were not included in the logistic modeling, then one might expect greater differences between the raking and predicted logistic weights on the one hand and the CHAID, mixed logistic, and collapsed logistic weights on the other hand. Thus, the similarity in weights produced by the alternative weighting schemes for the SIPP may not be as great in other circumstances.

A common concern that arises when many auxiliary variables are used to adjust the weights is that the adjusted weights might be highly variable, thus causing a serious loss of precision in the survey estimates. This proved not to be the case in the methods we evaluated. The variability of the weights with all the weighting schemes turned out to be similar, provided reasonable precautions were taken in creating the adjustments.

Although the empirical results do not show any appreciable differences in the estimates produced using the alternative weighting schemes and those produced using the SIPP panel weights, the correlations of the alternative adjusted weights and the current SIPP panel weight were found to be lower than the correlations among the alternative weights. This finding suggests that the choice of auxiliary variables is an important one, and probably more important than the choice of the weighting methodology. Although the more systematic methods used in this research for choosing the auxiliary variables did not result in major improvements over the current SIPP procedures, an analytic based choice of auxiliary variables may be more productive in other studies.

When a sizable number of auxiliary variables that are correlated to response propensity is available, it seems wise to use as many of them as possible in the nonresponse adjustment to serve as a safeguard in attempting to compensate for nonresponse bias. This general strategy should, however, be tempered by a careful assessment of the variation of the resulting weights in order to avoid too great a loss of precision in the survey estimates. In addition, a practical consideration that should be taken into account is the ease of implementation of the weighting methodology. If, as in this study, alternative weighting methodologies yield very similar weights and estimates, a method that is simple to apply may be preferable.

ACKNOWLEDGMENTS

We thank the referees for helpful comments on the earlier version of the paper. The paper reports research undertaken for the U.S. Bureau of the Census. The views expressed are the authors'. They do not necessarily reflect the views of the Bureau of the Census.

REFERENCES

CHAPMAN, D.W., BAILEY, L., and KASPRZYK, D. (1986). Nonresponse adjustment procedures at the U.S. Bureau of the Census. *Survey Methodology*, 12, 161-180.

DEMING, W.E., and STEPHAN, F.F. (1942). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.

DEVILLE, J.-C., and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

DEVILLE, J.-C., SÄRNDAL, C.-E., and SAUTORY, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88, 1013-1020.

JABINE, T.B., KING, K.E., and PETRONI, R.J. (1990). *Survey of Income and Program Participation (SIPP): Quality Profile*. Washington, DC: U.S. Bureau of the Census.

KALTON, G. (1983). *Compensating for Missing Survey Data*. Ann Arbor, MI: Survey Research Center, University of Michigan.

KALTON, G., and KASPRZYK, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.

KALTON, G., LEPKOWSKI, J.M., MONTANARI, G.E., and MALIGALIG, D. (1990). Characteristics of second wave nonrespondents in a panel survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 462-467.

KALTON, G., LEPKOWSKI, J., and LIN, T. (1985). Compensating for wave nonresponse in the 1979 ISDP Research Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 372-377.

KASS, G.V. (1980). An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, 29, 119-127.

KISH, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics*, 8, 183-200.

LEPKOWSKI, J., KALTON, G., and KASPRZYK, D. (1989). Weighting adjustments for partial nonresponse in the 1984 SIPP Panel. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 296-301.

LITTLE, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 2, 137-139.

NELSON, D., McMILLEN, D., and KASPRZYK, D. (1985). *An Overview of the SIPP, Update 1.* SIPP Working Paper No. 8401. Washington, DC: U.S. Bureau of the Census.

OH, H.L., and SCHEUREN, F. (1983). Weighting adjustments for unit nonresponse. In *Incomplete Data in Sample Surveys, Volume 2: Theory and Bibliographies* (Eds. W.G. Madow, I. Olkin, and D. Rubin), 143-184. New York: Academic Press.

PENNELL, S.G., and LEPKOWSKI, J.M. (1992). Panel conditioning effects in the Survey of Income and Program Participation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 556-571.

RIZZO, L., KALTON, G., and BRICK, M. (1994). Weighting adjustments for panel nonresponse in the SIPP. Final Report submitted to U.S. Bureau of the Census for SIPP Panel Nonresponse Project.