OVERVIEW OF REDESIGN METHODOLOGY FOR THE SURVEY OF INCOME AND PROGRAM PARTICIPATION

Peter H. Siegel and Stephen P. Mack, U.S. Bureau of the Census[1]
Peter H. Siegel, U.S. Bureau of the Census, Washington, DC 20233

Key Words: Stratification, Primary Sampling Unit, Oversampling, Nonresponse

I. INTRODUCTION

The Census Bureau conducts the Survey of Income and Program Participation (SIPP) to provide information for federal policy makers and academic analysts on topics such as poverty, income distributions, government program participation and eligibility, and health insurance coverage. The SIPP is a nationally representative, longitudinal survey. Every 10 years, we redesign the SIPP to update the sampling frame based on the most recent Census and housing units built since the Census and to make other necessary changes. Since the SIPP began in 1984, researchers and analysts have investigated and evaluated SIPP data and suggested ways to improve the SIPP sample design and weighting methodology. We have incorporated many of these ideas into the redesigned SIPP, which we will introduce in February 1996. Each SIPP panel will have a sample size of 50,000 households, be four years long, cover twelve interviews (waves), and be non-overlapping with other panels. There will be a thirteenth interview so that we can obtain a complete four calendar years worth of data and this interview will overlap with the beginning of the next panel.

We considered continuing to use overlapping panels to reduce time-in-sample bias and bias due to nonresponse which increases over the life of a panel. Time-in-sample bias occurs when response patterns change when persons are subject to repeated interviewing over the life of a panel in the SIPP. However, there was not significant evidence of such biases, so we chose a non-overlap design. (McCormick, et al, 1992.)

The SIPP has a recall period of four months, i.e. we collect data from the respondents for the previous four months. We researched the possibility of a six month recall period but determined that the longer recall period would have an adverse effect on data quality as a result of higher response variance and response bias due to the longer recall period. (Huang, et al, 1994.)

This paper gives a general description of the major areas of the redesign methodology. Section II of this paper defines the primary sampling unit (PSU) and gives an overview of the PSU stratification. We summarize the sample selection of PSUs within strata in section III and the sample selection of households within PSUs in section IV. Section V describes the weighting methodology. We conclude the paper in section VI mentioning additional redesign methodology.

II. PSU DEFINITION AND STRATIFICATION

The PSU for the SIPP is a county (or county equivalent) or a group of contiguous counties. A 1980 PSU is one that we formed based on the 1980 Census, and a 1990 PSU is one that we formed based on the 1990 Census. We generally defined the 1990 PSUs the same as they were in 1980. However, we changed the definitions in some of the following situations:

- A difference occurred in definition between 1980 Current Population Survey (CPS) and 1980 SIPP PSUs. However, some SIPP PSUs cross state boundaries and CPS PSUs do not cross state boundaries. We did not change the PSU definition in such a situation.
- A regional office identified a PSU as a problem area.
- A PSU had less than 10,000 persons in 1980.

Some of the ways we changed PSU definitions include adding a county to a PSU, moving a county from one PSU into a new PSU containing only that county, or switching counties between two or more PSUs.

Johnson (1990) gives details of changing PSU definitions.

We stratified the PSUs for the purpose of first stage sample selection. First, we calculated an average stratum size which would give optimum field representative workloads based on the projected 1995 number of housing units, a two PSU per stratum design, and a sampling interval of 4,200. While calculating the average stratum size, we adjusted for group quarters, undercoverage, and ineligible units. In order to maintain workloads and reduce variance,

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau. We wish to thank Jay Kim, Harland Shoemaker, Lloyd Hicks, and David Hubble for their comments on this paper and Sandy Carnegie for her help in preparing the final version.

we allowed the smallest and largest stratum sizes to be 90 percent and 110 percent, respectively of the average stratum size. To achieve valid probabilities for PSU selection, each PSU could contain no more than half of the housing units in the stratum. Based on these constraints, any PSU with a projected 1995 housing unit count greater than 196,513 is self-representing (SR). We did not have to stratify the SR PSUs because they are in sample with certainty. However, we wanted every state to include at least one sample PSU, and ten states contained no PSUs with enough housing units for us to designate a SR PSU. Therefore, we chose the PSU with the most projected 1995 housing units in each of these states to be SR. (McCormick, 1992a.) There are 112 SIPP SR PSUs.

The PSUs which we did not designate as SR are nonself-representing (NSR). We stratified the SIPP NSR PSUs separately within the four Census regions, but before stratifying, we determined the number of NSR strata for each region. We began with the projected number of 1995 housing units in the NSR PSUs in each region. Then, we adjusted this number to account for group quarters, undercoverage, and ineligible units. We divided this adjusted number by the 1995 average stratum size to get the number of strata per region. (McCormick, 1992a.) There are 105 NSR strata and a total of 210 SIPP sample NSR PSUs.

We used the Friedman-Rubin (F-R) stratification system to place NSR PSUs into strata in such a way as to minimize the between PSU variance over 15 stratification variables, subject to certain size constraints. The F-R system tried harder to minimize the between PSU variance for the variable persons below poverty than for the other 14 variables because poverty is the variable of the most interest in the SIPP. The stratification variables are the same as those used in the 1980 stratification and include persons 65 years old and older, female, Black, or Hispanic headed households, and occupied housing units receiving public assistance. McCormick (1992b) has a complete list of these variables.

We used output from the F-R stratification system and produced between PSU variances for 19 variables to evaluate the stratification. These variables include 1980 and 1990 persons below poverty, 1980 and 1990 female, Black, or Hispanic headed households, and 1990 persons 65 years old and older. McCormick (1992b) has a list of these variables. We evaluated five stratifications for each region and chose the one for each region which had the lowest between PSU variances for the evaluation variables.

III. PSU SAMPLE SELECTION

We selected two PSUs per NSR stratum and all SR PSUs. Our goal in selecting NSR PSUs was to maximize the amount of PSU overlap. PSU overlap occurs when any part of a 1980 sample PSU is in a 1990 NSR PSU. Maximum overlap saves field costs expended for training new interviewers and improves quality by maintaining experienced interviewers from the 1980 SIPP design.

We accomplished this maximum overlap using a linear programming algorithm or more specifically a linear transportation algorithm. This algorithm calculated the joint probability of selection for each pair of 1990 PSUs, conditional on part or all of each PSU being in the 1980 sample, which maximized the expected number of 1980 sample PSUs in the 1990 sample. It also preserved the 1990 PSUs' unconditional probabilities of selection. We had the following two constraints:

- For each 1980 sample PSU, the sum of the joint probabilities of selection across all of the 1990 PSUs must equal its 1980 unconditional probability of selection
- For each 1990 PSU, the sum of the joint probabilities of selection across all of the 1980 sample PSUs must equal its 1990 unconditional probability of selection

Ernst (1989) presents more details of the transportation algorithm.

There were two NSR strata in which we did not maximize overlap because each of these two strata contained more than 60 PSUs, and we did not have the computer resources to maximize overlap when a stratum contained more than 60 PSUs. For these two strata, we calculated joint probabilities of selection for each possible pairing of PSUs using Durbin's method. (Roebuck, 1992.)

Within each NSR stratum, regardless of the number of PSUs, we selected one pair of PSUs proportional to the joint probabilities of selection of the pairs of PSUs. About 76 percent of the 1990 sample NSR PSUs in the country overlap with 1980 sample PSUs. If we had not maximized overlap, we would expect about a 30 percent overlap.

IV. HOUSEHOLD SAMPLE SELECTION

We selected households within five separate, non-overlapping sampling frames: unit, area, new construction, group quarters (GQ), and coverage improvement. Weller, et al (1991) describe 4 of these frames and Hendrick (1994) describes the coverage improvement frame. Within the unit frame, we selected clusters of housing units, and within the area, new construction, and GQ frames, we selected clusters of measures.[2] A cluster contained four housing units or measures in the unit, area, and GQ frames and 8 measures in the new construction frames. In the coverage improvement frame, we selected sample from an address list of housing units which were found in the Post Enumeration Survey (PES) and not found in the 1990 Census. We selected all 795 housing units which were in unit frame blocks and were single units or multi-unit whole structure Census misses. (Hendrick, 1994.) The national sampling interval in the unit, area, new construction, and GQ frames is 4,200,[3] except in five of the ten SR PSUs which have a relatively small number of housing units. These five PSUs have smaller sampling intervals so that the monthly interviewer workload will not be too small. The within PSU sampling interval is the PSU's unconditional probability of selection times the national sampling interval.

In the unit, area, new construction, and GQ frames, we adjusted the within PSU sampling interval to account for the cluster size, number of samples desired (16), and selection of sample by other Census Bureau demographic surveys which selected sample before the SIPP. (Cantwell and Rothhaas, 1993a and 1993b.) Then we used this TE to select the clusters of measures in the new construction and GQ frames.

We oversampled low income households in the unit and area frames. Data users need large enough sample sizes for subpopulations to get more reliable estimates. We wanted to increase the sample size for small subpopulations of persons in poverty, and as an additional benefit we would lower the variance for persons in poverty. We researched oversampling low income households and found potential variance reductions for the number of Black, Hispanic, and total persons in or near poverty of 34 percent, 27 percent, and 5 percent, respectively. This research also showed that if we oversampled, we would get an expected 20 percent of sample persons below 150 percent of poverty, i.e. persons who had income less than 150 percent of the poverty threshold. We found expected sample size increases of 47 percent, 36 percent, and 18 percent for Blacks, Hispanics, and total persons in poverty, respectively. Based on this research, we decided to oversample, but we designed the sample to allow us to switch to a self-weighting design, if desirable. Weller, et al (1991) give more details on the oversampling research.

Within each PSU, we divided unit and area frame households into two strata, the oversample (low income) stratum (stratum one) and the undersample stratum (stratum two). We used 1990 Census data to form the strata. In the unit frame, we had data at the household level, and in the area frame, we had data at the block level. However, we only had income data for households that filled out a Census long form (approximately one in six households), so we used auxiliary variables, such as households with a female householder, no spouse present, and at least one of her own children under the age of 18 present in the household and households living in a central city of a Metropolitan Statistical Area (MSA) with a Black or Hispanic householder, to identify low income households. We constrained the oversampling by limiting the variance increase for persons 55 years old and older to five percent because we were concerned about large variance increases for this group of persons.

In the unit frame, we placed all housing units which were below 150 percent of poverty in stratum one and the remaining housing units in stratum two. We used an iterative procedure to stratify the area frame blocks within each PSU. We ranked these blocks by the proportion of their population in poverty, so that the first block had the highest proportion of poverty. We moved one block at a time into stratum one, which already contained some housing units from the unit frame, and left the remaining blocks in stratum two, which contained the remaining unit frame housing units, until we met at least one of the following three criteria:

- the size of the population in stratum one is greater than or equal to the size of the population in stratum two or
- the proportion of the population in poverty in stratum one is less than or equal to the proportion of the population in poverty in stratum two or
- the weighted number of persons below 150 percent of poverty is zero in stratum two.

We reviewed all iterations for each PSU to select the stratification. We looked at design effects for three subgroup characteristics and 13 variables. We chose the iteration which minimized the design effects,

2 A measure contains four expected housing units in the area and new construction frames and four expected housing unit equivalents in the GQ frame.

3 Originally, we wanted a sample size of 20,000 households, so the sampling interval was 4,200. After we selected 16 samples of 20,000 households, we decided on a sample size 50,000 households. We combined samples, and the sampling interval is 1,683.33.

with the most emphasis on minimizing the design effects for total persons, Blacks, Hispanics, and female householders below 150 percent of poverty. If the design effects were too large for a given PSU, then we had the option of not oversampling in that PSU. However, this situation never arose, so we oversampled in every sample PSU. Weller and Siegel (1993a) discuss the details of the SIPP oversampling.

We initially planned to calculate PSU-level sampling and subsampling rates within both strata one and two. However, we researched the possibility of using national-level sampling and subsampling rates within both strata one and two and found by using test data from 1980 in 27 MSAs that the variances got better for some characteristics but worse for others. We decided to calculate the rates at both the PSU and national levels, compare the variances of several key characteristics computed using these two different rates, and choose the rates that gave the lowest variances. Based on these comparisons, we decided to use the national-level rates. Weller and Siegel (1993a) describe the formulas for the rates. Hendrick (1995) has information on the national-level rates.

Initially, we selected clusters of households and measures in the unit and area frames, respectively using the stratum one national oversampling take-every (TE) in both strata one and two. We used the stratum one rate in both strata so that we left an unbiased universe from which other Census Bureau demographic surveys would sample. Next, we selected sample for a self-weighting design using a subsampling TE equal to the national SIPP optimal oversampling rate to subsample the strata one and two clusters which we initially selected for sample. Then, we selected sample for an oversample design. All stratum one clusters which we initially selected for sample are in sample in the oversample design. We subsampled stratum two clusters which we selected for sample in the self-weighting design using a subsampling TE equal to the original national SIPP subsampling retention TE divided by the subsampling TE for the self-weighting design. (Weller and Siegel, 1993b.)

We assigned a sample designation to each sample unit in the unit frame and to each measure in the area, GQ, and new construction frames to indicate when we should interview it. We spread the sample such that neighboring households would not be in sample close in time and this would help us reduce bias. Also, we systematically divided each sample into quarters to spread the workload of a sample over a period of four months. Finally, we systematically assigned each sample unit or measure to a reduction group. We can use these reduction groups to drop or reinstate sample some time during the decade in the event of a change in the sample size due to budgetary or other reasons. (Allen and Gorsak, 1994.)

V. WEIGHTING

The SIPP provides weights for persons that are appropriate for different types of analysis. Cross-sectional weights are provided for analysis of data from a particular month. Longitudinal weights are provided for analysis of data over time such as for a particular calendar year or for the life of the panel. All persons classified as interviewed for the appropriate period receive a positive weight. Interviewed persons are persons who were self, proxy, or imputed respondents for each month in the weighting period for which they were eligible to be interviewed. Persons who die or move to ineligible addresses are no longer eligible to be interviewed. SIPP weights are comprised of the following components:

Cross-sectional Weights

- Base Weight (BW) - The inverted probability of selection of a person's household
- Duplication Control Factor (DCF) - Adjusts for subsampling done in the field when the number of sample units is much larger than expected
- Wave 1 Noninterview Adjustment Factor (F_{N1}) - Adjusts for noninterviewed households in wave 1 that were eligible for interviews
- Movers' Weight (MW) - Adjusts for persons in the SIPP universe who move into sample households after wave 1
- Household Noninterview Adjustment Factor (F_{N2}) - Adjusts for noninterviewed households in waves 2+ (two and greater) that were eligible for interviews
- Second Stage Adjustment Factor (F_{2S}) - Adjusts estimates to population controls and causes husbands' and wives' weights to be equal

The final cross-sectional weight is $\mathbf{FW_C = BWxDCFxF_{N1}xF_{2S}}$ for wave 1 and is $\mathbf{FW_C = IWxF_{N2}xF_{2S}}$, for waves 2+, where IW is either $\mathbf{BWxDCFxF_{N1}}$ or MW. James (1995) and Siegel (1995a) describe SIPP cross-sectional weighting in greater detail.

Longitudinal Weights (panel and calendar year)

- Initial Weight (IW) - Cross-sectional weight before second stage adjustment from the appropriate panel or calendar year "control" month
- Noninterview Adjustment Factor (F_L) - Adjusts for noninterviewed persons with nonzero initial weights
- Second Stage Adjustment Factor (F_{L2S}) - Adjusts estimates to population controls. Husbands' and wives' weights are not equalized

The final longitudinal weight is $FW_L = IWxF_Lx F_{L2S}$.

Researchers both inside and outside the Census Bureau conducted evaluations of SIPP weighting methodology and researched alternative methodologies. We are making several improvements to SIPP weighting methods beginning with the 1996 panel:

- We dropped the first stage factor (F_{1S}) from cross-sectional weighting. This factor adjusted for differences between the 1980 Census count of population and an estimate of that count based on Census data for sample PSUs. James (1994) found that it did not reduce variance as was previously thought. Jabine, et al (1990) describe the first stage factor.
- We are using additional variables in nonresponse adjustment. We added within PSU stratum code to wave 1 nonresponse adjustment, and we added household income, geographic division, and number of imputations for selected income and asset items to nonresponse adjustment for waves 2+. Research by Rizzo, et al (1994) and by Folsom and Witt (1994) pointed out the potential of the latter three variables to reduce nonresponse bias.
- We redefined nonresponse adjustment cells for waves 2+ weighting. We formed the nonresponse cells by successively partitioning data from five panels by whichever variable most reduced the bias of the household income to poverty threshold ratio. We used data from a sixth panel to evaluate the results. We calculated the nonresponse bias of six variables at waves two and seven for both the new cells and the original cells using initial weights and data from the most recent interview in the calculations. The new cells had lower bias for five of the six variables. (Siegel, 1995b.)
- The movers' weight adjustment will include both children and adults. Prior to the 1996 panel, only adults were considered in the adjustment.

Research was conducted on a number of promising weighting improvements. Allen and Petroni (1994) reported on an adjustment for mover attrition. Folsom and Witt (1994) and Rizzo, et al (1994) studied alternative nonresponse adjustments using response propensity models. Each study computed weights using an alternative methodology. The researchers then compared estimates of various items to benchmarks. The benchmarks came from administrative records and survey data with less nonresponse than the SIPP. The comparisons did not provide strong evidence of lower bias using the alternative weighting methods.

The Census Bureau is investigating other weighting alternatives:

- The use of regression weighting methods as an alternative to the current longitudinal weighting procedure. (An, et al, 1994.)
- The use of Internal Revenue Service (IRS) data to improve income estimates. (Dorinski and Huang, 1994.) The research focuses on using IRS data in addition to population controls in the second stage adjustment procedure for longitudinal weights.

VI. ADDITIONAL METHODOLOGY

In addition to the methodology mentioned above, we are currently developing variance methodology for the SIPP. We will be calculating variances using partially balanced half-sample replication. Also, recently analysts and data users have expressed an interest in state-level estimates from the SIPP. With the possibility of welfare becoming the responsibility of the states, state-level estimates become more important. We are currently researching the feasibility of producing state-level estimates.

REFERENCES

Allen, T. and M. Gorsak (1994), *Post Sampling Code Assignments - SIPP (Document #CA-S-6)*, Internal Census Bureau Memorandum from Within-PSU Sampling Workgroup (WG-4) Code Assignment Specifications Subgroup for Walter, March 28, 1994.

Allen, T. and R. Petroni (1994), Mover Nonresponse Adjustment Research for the Survey of Income and Program Participation, *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

An, A., F. Breidt and W. Fuller (1994), Regression Weighting Methods for SIPP Data, *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Cantwell, P. and R. Rothhaas (1993a), *Determining and Resolving Crowded BPCs (Doc. #: PF-S-2)*, Internal Census Bureau Memorandum from Within-PSU Sampling Workgroup (WG-4) Parameter File Specifications Subgroup for Walter, February 22, 1993.

Cantwell, P. and R. Rothhaas (1993b), *Computing Adjusted Take Everys (Doc. #: PF-S-4)*, Internal Census Bureau Memorandum from Within-PSU Sampling Workgroup (WG-4) Parameter File Specifications Subgroup for Walter, March 16, 1993.

Dorinski, S. and H. Huang (1994), Use of Administrative Data in SIPP Longitudinal Estimation, *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Ernst, L. (1989), *Further Applications of Linear Programming to Sampling Problems*, SRD Research

Report Number: Census/SRD/RR-89/05, October 12, 1989.

Folsom, R. and M. Witt (1994), Testing a New Attrition Nonresponse Adjustment Method for SIPP, *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Friedman, H. P. and J. Rubin (1967), On Some Invariant Criteria for Grouping Data, *Journal of the American Statistical Association.*

Hendrick, M. (1994), *Documentation of the Sampling and Code Assignment for the HUCS (Housing Unit Coverage Study) Sample*, Internal Census Bureau Memorandum from Hendrick for documentation, December 22, 1994.

Hendrick, M. (1995), *Documentation and Formulas for Finding the National Oversampling Rate and the National Variance Ratios*, Internal Census Bureau Memorandum from Hendrick for documentation.

Huang, H., K. Cooper, G. Shapiro and K. Short (1994), SIPP Recall Length Decision: An Alternative to Experimentation, *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Jabine, T., K. King and R. Petroni (1990), *Survey of Income and Program Participation Quality Profile*, U.S. Bureau of the Census.

James, T. (1995), *SIPP 96+: Cross-Sectional Weighting Specifications for Wave 1*, Internal Census Bureau Memorandum from Waite to Courtland, June 14, 1995.

James, T. (1994), *SIPP 95+: Dropping the First Stage Factor from the SIPP Weights*, Internal Census Bureau Memorandum from Huggins to Cahoon, April 6, 1994.

Johnson, N. (1990), *Recommended Changes in PSU Definitions for the 1990 SIPP and the CPS Redesign (Doc. #2.2-R-05)*, Internal Census Bureau Memorandum from PSU Definition Subgroup (SG-2.2) for 1990 Redesign Budget and Planning Subgroup (SG-1.2) and Research Agenda Subgroup (SG-1.0), October 4, 1990.

McCormick, M. (1992a), *SIPP 1990 Redesign: Parameters for the Stratification of PSUs*, Internal Census Bureau Memorandum from Singh to Shapiro, October 27, 1992.

McCormick, M. (1992b), *1990 Redesign: Documentation of the SIPP Stratification and Evaluation Variables*, Internal Census Bureau Memorandum from McCormick for documentation, October 22, 1992.

McCormick, M. (1992c), *1990 Design Current Surveys Stratification Program Specifications ("Friedman-Rubin" Program) (Doc. #2.1-R-1)*, Internal Census Bureau Memorandum from Friedman-Rubin Subgroup (2.1) of PSU Stratification and Selection Work Group (WG-2) to Walter, April 30, 1992.

McCormick, M., D. Butler and R. Singh (1992), Investigating Time In Sample Effect for the Survey of Income and Program Participation, *American Statistical Association 1992 Proceedings of the Section on Survey Research Methods.*

Rizzo, L., G. Kalton, J. M. Brick and R. Petroni (1994), Weighting for Panel Nonresponse in the Survey of Income and Program Participation, *American Statistical Association 1994 Proceedings of the Section on Survey Research Methods.*

Roebuck, M. (1992), *1990 Redesign: PSU Selection Specifications (Doc. #2.4-R-7)*, Internal Census Bureau Memorandum from Roebuck to Walter, September 25, 1992.

Siegel, P. (1995a), *SIPP 96+: Cross-Sectional Weighting Specifications for the Second and Subsequent Waves*, Internal Census Bureau Memorandum from Waite to Courtland.

Siegel, P. (1995b), *SIPP 96+: Creation and Evaluation of Nonresponse Cells for the Second and Subsequent Waves*, Internal Census Bureau Memorandum from Siegel for Documentation, June 13, 1995.

Weller, G., V. Huggins and R. Singh (1991), Oversampling the Low-Income Population in the Survey of Income and Program Participation, *American Statistical Association 1991 Proceedings of the Section on Survey Research Methods.*

Weller, G. and P. Siegel (1993a), *Obtaining Within-PSU Stratification and Sampling Rates for SIPP Oversampling (DOC.# OS-S-2)*, Internal Census Bureau Memorandum from Within-PSU Sampling Workgroup (WG-4) Oversampling Specifications Subgroup for Walter, October 29, 1993.

Weller, G. and P. Siegel (1993b), *Subsampling Take Everys for SIPP BPCs in the Area and Unit Frames (DOC.# PF-S-6)*, Internal Census Bureau Memorandum from Within-PSU Sampling Workgroup (WG-4) Parameter File Specifications Subgroup for Walter, November 2, 1993.