THE SURVEY OF INCOME AND
PROGRAM PARTICIPATION

# OVERSAMPLING THE LOW-INCOME POPULATION IN THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)

No. 160

G. D. Weller, V. J. Huggins and
R. P. Singh
Bureau of the Census

U. S. Department of Commerce BUREAU OF THE CENSUS

# OVERSAMPLING THE LOW-INCOME POPULATION IN THE SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)*

Glenn D. Weller, Vicki J. Huggins, Rajendra P. Singh
U.S. Bureau of the Census

## I. Introduction

The Survey of Income and Program Participation (SIPP) is a nationally representative survey. It is designed to provide comprehensive information that reflects the financial situation of persons, families, and households in the United States (except persons in institutions). Interviews for the first SIPP sample panel (1984 panel) began in October 1983. Later panels (1985-1991) began in February of each calendar year. The SIPP has an overlapping panel design that allows combining panels for multi-panel estimation covering the same period.

The total sample size in the SIPP has dwindled from 20,000 interviewed households in the 1984 panel to about 12,000 interviewed households in the 1986-1989 panels due to budget constraints. Analysis of this reduced sample is not as useful, especially for subpopulations of interest. Even sample based on two combined panels is not large enough to satisfy data users' needs for the analysis of subpopulations such as Blacks in poverty, female-headed households on food stamps, etc. This has prompted investigation into oversampling of low income and aged persons in the SIPP.

Redesign of the SIPP based on the 1990 Decennial Census of Population and Housing is now underway. As part of 1990 redesign research, we researched oversampling methodologies for the following subgroups in the SIPP (in order of priority) to investigate the ramifications of oversampling in 1995-2005 SIPP panels:

1. Poor
2. Near poor
3. Age 65+ or age 75+

This paper presents research into oversampling the low-income population (i.e. poor and the near poor). Oversampling for persons aged 65+ or 75+ can be done using administrative records and we do not discuss it here.

To give further background on redesign of the SIPP based on the 1990 decennial census, we define the following frames which all Census demographic surveys use:

Unit frame - List of addresses from the decennial census.

Area frame - Addresses for blocks with incomplete addresses or areas where new construction permits are not issued. These blocks are taken from the decennial census and listed in the field.

NC frame - New construction permits to capture construction after the census.

GQ frame - List of group quarters such as boarding houses, hotel rooms, and institutions from the decennial census.

There will be no oversampling in either the new construction or group quarters frame. Oversampling research was done using data from the unit frame. However, at implementation there will also be oversampling in the area frame. Adjustments made to the initial results take this into account.

We also analyzed data from the American Housing Survey (AHS) to estimate the effectiveness of the oversampling over time in the unit and area frames. This paper presents the estimated changes due to these effects over the life of the redesigned SIPP.

The following sections and their content are:

Section II - gives a brief theoretical introduction to the method used in the oversampling research.

Section III - presents the results of the oversampling research at the time of the census for the unit frame.

Sections IV - discusses what effects oversampling over time may have on the oversampling gains.

Section V - discusses what effects inefficiencies in oversampling in the area frame may have on oversampling gains.

Section VI - estimated variance reductions for the 1995-2005 panel samples.

Section VII - assumptions used in the research.

Section VIII - a final discussion of the results.

## II. Methodology

The oversampling methodology used in the research creates two strata using geographic units within primary sampling units (PSUs). Sample is taken from each strata at a different rate. These different sampling rates permit the sample size to remain fixed. The fixed sample size is necessary due to a fixed budget. We also fix sample sizes at the PSU level to use the interviewers' time more efficiently. This will reduce ongoing survey costs by reducing expenses of hiring and training new interviewers. As the number of PSUs stratified at once increase, so do operational difficulties. For this reason we stratified each PSU in the research one at a time.

Joseph Waksberg[1] first proposed the particular methodology discussed in this paper. Initially, research (stratification within PSUs) was to be done at both the block and housing unit levels. Because of time constraints and cutbacks in research funding, we did less research than we had originally planned. Due to the expectation that the initial housing unit (HU) stratification would be better than block-level stratification, research was done at the HU-level only.

Consider a population of size N divided into two strata where:

- $N_1$ is the size of the population in stratum 1. This stratum will have a higher concentration of the subgroup of interest. (In the research, the subgroup of interest is low-income persons.)
- $N_2$ is the size of the population in stratum 2. This stratum will have a lower concentration of the subgroup of interest.
- $t_1$ is the proportion of the population in stratum 1 that is in the subgroup of interest.
- $t_2$ is the proportion of the population in stratum 2 that is in the subgroup of interest.
- $r_1$ is the sampling rate in stratum 1.
- $r_2$ is the sampling rate in stratum 2.
- $\sigma_1^2(Y)$ is the population variance for a characteristic within the subgroup of interest in stratum 1.
- $\sigma 2^2(Y)$ is the population variance for a characteristic within the subgroup of interest in stratum 2.
- $\sigma_1^2(Z)$ is the population variance for a total population characteristic in stratum 1.
- $\sigma_2^2(Z)$ is the population variance for a total population characteristic in stratum 2.

Define

$N_2 = v\,N_1,\ v \geq 1$
$t_1 = u\,t_2,\ u \geq 1$
$r_1 = k\,r_2,\ k \geq 1$
$\sigma_1^2(Y) = w\,\sigma_2^2(Y),\ w > 0$
$\sigma_1^2(Z) = C\,\sigma_2^2(Z),\ C > 0$

Now consider two sampling plans:

A. Select a simple random sample from each stratum with sampling rates $r_1$ and $r_2$ ($r_1 \geq r_2$) such that $r(N_1 + N_2) = r_1N_1 + r_2N_2$

B. Select a simple random sample using rate r.

Then the ratio of the variances for plan A over plan B (i.e. the design effect for oversampling) for a characteristic within the subgroup is

$$R_5 = \frac{\sigma_1^2(Y)\ (wu+kv)\ (k+v)}{\sigma^2(Y)\ kw(u+v)\ (1+v)}$$

where $\sigma^2(Y)$ is the population variance for a characteristic within the subgroup without regard to strata.

It is minimized for $k = \sqrt{uw}$ . The corresponding ratio for an attribute of the total population is given by

$$R_6 = \frac{\sigma_1^2(Z)\ (C+kv)\ (k+v)}{\sigma^2(Z)\ kC(1+v)^2}$$

where $\sigma^2(Z)$ is the population variance for a total population characteristic.

If we wish to hold the variance increase for a total population characteristic to a fixed amount we can set $R_6$=m (where m = 1.05 for a 5% increase) and solve for k. Solving for k we get

$$k = \frac{-b_m + \sqrt{b_m^2 - 4C}}{2}$$

and

$$b_m = \frac{C+v^2}{v} - \frac{m\sigma^2(Z)\ C(1+v)^2}{\sigma_1^2(Z)\ v}$$

Using the k parameter found in this way fixes the total increase in variance for a total population characteristic.

## III. 1990 Redesign Oversampling Research in the Unit Frame

In our research, we estimated the reduction in variance due to oversampling in the 1990 SIPP redesigned panels. Only research into oversampling the low-income population was done. Variance increases for persons aged 55+ were set to 5%, 10%, and 15% using the sampling formulas.[2] The overall goal is to improve estimates for selected subgroups, without significant adverse effects to other important estimates. Recall that with this methodology we assume a fixed budget so sample size must remain fixed.

The Waksberg methodology focuses on the importance of subgroups. Calculating optimal sampling rates using subgroups of interest produces the minimum variance for a given stratification. The subgroups of interest were:

- number of Blacks in or near poverty
- number of Hispanics in or near poverty
- number of female-headed householders in or near poverty.

Ideally, we should use these variables to form within PSU stratifications. Unfortunately, these variables were not available in all cases from the 1990 Census, only from a sample of the Census. When these variables were not available we used a set of auxiliary variables. In general, census sample cases made up about 1/6 of the total U.S. population and more information is available for census sample cases for use in stratification.

Census non-sample cases use auxiliary variables for stratification. The auxiliary variables for non-sample cases were identified in discussions with analysts within the Bureau. These analysts have extensive experience in analyzing poverty and other related statistics. The following is a list of variables used for within-PSU stratification:

For Census sample cases: Poverty status (< 150% of the poverty threshold[3]). The auxiliary variables for Census non-sample cases are:

1. Female householder, no spouse present with own children under age 18

2. Living in a central city of a metropolitan statistical area (MSA)

and

Renter with rent < $150

3. Black householder

and

living in a central city of an MSA

4. Hispanic householder

and

living in a central city of an MSA

5. Black householder

and

householder < age 18 or greater than age 64

6. Hispanic householder

and

householder < age 18 or greater than age 64

Research conducted in 27 PSU equivalents from 1980 census data showed average reductions in variance for persons < 150% of poverty, total Blacks < 150% of poverty, total Hispanics < 150% of poverty, and Female-headed householders < 150% of poverty of 24%, 38%, 22% and 16% respectively. Table 1 reports these results. The stratification used to get the results in table 1 fixed the variance increase for persons aged 55+ to 5%. By doing this we avoided any significant loss for the aged 55+ group. [We also looked at 10% and 15% constraints but there were no significant gains for the poverty subgroups overall for the additional loss to variances for the aged 55+]. These results were very similar from PSU to PSU in the research.

We examined the effects of oversampling on thirty-five other evaluative variables. We know for gains in low-income we will lose in other groups, since we are re-allocating, not increasing, overall sample. Table 1 also presents the auxiliary variables. Oversampling helped variables related to poverty, such as Number of Renter Occupied Units with rent < $150 which received a 27 % decrease in variance. Conversely, those variables related to being affluent, such as incomes greater than $75,000 per year, were hurt by the oversampling of low income, receiving a 13% increase in variance. In general, any increases observed are not alarming considering the variance reductions for poverty related estimates and that CVs for many of the middle to high income related items are reasonably good in the current SIPP design. CVs calculated from the research, before and after oversampling, are given in table 2.

Small sample sizes are of great concern in the SIPP data user community. They wanted a fifty percent increase in sample for low income groups out of oversampling. With this oversampling methodology, an increase of 47% was seen for Blacks in or near poverty, 36% for Hispanics in or near poverty, 29% for female-headed householders in or near poverty, and 22% for all persons in or near poverty. Table 3 presents estimated sample size increases, by PSU, for these characteristics.

## IV. Stratification Over Time

The Within-PSU stratification into high and low poverty strata will lose some effectiveness over time, but how fast this will occur and how much of a loss there will be in the years 1990-2005 is unknown. Therefore, research into how effective the stratification will remain over time was done to assess how effective oversampling will be over time.

The oversampling results cited in the section above are as of the time of the Census. When fielded, the survey will already be five years old. To estimate the effects of changes over time on the oversampling methodology, we studied the American Housing Survey (AHS) data for the years 1974, 1977, 1981, and 1985. Data for certain characteristics is missing for 1974, so analysis for only a few characteristics was possible using 1974 AHS data. Therefore, analysis continued using only 1977-1985 data for all desired characteristics.

The study shows that most of the loss in effectiveness occurred in the first 4 years after redesign and leveled off and often improved after 8 and 11 years (see table 4). For instance, number of persons in or near poverty showed changes in variance over time of +5%, +5%, -1% for 4 years, 8 years, and 11 years respectively. If one assumes a similar economic situation, population movement, growth, etc. will exist in 1995-2005 as 1974-1985, the study provides a fair indication of how much of a loss will occur before phase out of the new SIPP design in the year 2005. Table 4 summarizes estimated deterioration for a selected set of key characteristics over time.

Table 4 shows that the loss of effectiveness over time is small relative to the initial gains made in the oversampling. However, any major changes in the national or regional economy could significantly affect the results and the effectiveness of the oversampling. Overall, the variances for the studied characteristics showed increases of no more than 8% during 11 years. If 1990-2005 exhibits the same increase as 1977-1985 then the effects of time on the stratification are not large enough to warrant concern over future effectiveness of the oversampling.

Assuming similar economic conditions is a pretty strong assumption that is unlikely to be true. However, looking at the worst case of available data, the period 1977-1985, we still have significant gains with the stratification scheme. We can only extrapolate that losses due to stratification over time for SIPP 1995-2005 panels will not be extremely worse. Hence, oversampling should be a viable resource for improving SIPP statistics for the low income population in the unit frame even with losses over time.

## V. Adjustments for New Construction and Area Frames

Other effects on expected gains that will occur at the time of implementation result from two sources. The first is that oversampling will not be done in new construction, which is approximately 10% of the population. The other source is that stratification of the area is at the block level. We believe that stability at the block level is somewhat higher than at the housing unit level. However, stratification at the block level will be less effective than housing unit level stratification.

The area frame is about 20% of the population. We estimated the effect of these two frames, new construction and area, on expected variance reductions for the poverty groups in table 5 by assuming that the 20% population in the area frame will receive half of

the reduction of the unit frame. The 10% of population in the new construction frame will receive no gain from the oversampling. The second column in table 5 shows estimated changes in variances, due to inefficiencies in the area and new construction frames in the subgroups number of Blacks in or near poverty, number of Hispanics in or near poverty, and number of persons in or near poverty of +6%, +2%, and +4% respectively. Since the oversampling methodology focuses on improvement for the subgroups, our greatest concern is with changes for those groups.

## VI. Estimated Variance Reductions 1995-2005

The estimated reductions in variance for three groups/subgroups during the 1995-2005 implementation are given in table 5. These estimated reductions include the increase in variance discussed in section V as well as stratification-over-time increases (the 1977-1985 period was chosen since it provided a worst case scenario). Variance reductions for number of Blacks in or near poverty, number of Hispanics in or near poverty, and number of persons in or near poverty are 31%, 20%, and 15% respectively. The variance reductions for the two subgroups are large enough to benefit in their analysis.

## VII. Assumptions

The main assumptions used in the research are:

1. Stratification over time for 1995-2005 will be comparable to results from the research period of 1974-1985. This implies that results from the 11-year period from 1974-1985 are indicative of changes that can be expected for the 5 to 15 year period of sample implementation.
2. The size of the average household is two adults with two children. This assumption was used only in the stratification-over-time analysis to define poverty cutoffs.
3. Housing unit and block level stability are assumed to be comparable.
4. There will be gains for practically all PSUs as shown in research.
5. Housing unit and block level stratification will be different with block level stratification being inferior.
6. Housing unit size will vary by stratum. Stratum 1 households (high poverty) are assumed to have a larger size of 3.09 persons per household, while stratum 2 households (low poverty) are assumed to have a household size of 2.57 persons per household. This assumption was only used to calculate SIPP sample sizes.[4]
7. The research included data from 27 metropolitan statistical areas (MSAs). These MSAs were chosen based on several criteria. Each of the MSAs needed block level information so research could be done at the block level if desired. As a group, the MSAs provide a mix of rural and urban areas as well as a mix of characteristics that we want to oversample.

## VIII. Discussion

When discussions began on whether the SIPP should oversample, SIPP data users felt that the SIPP should settle for no less than a 50% increase in sample sizes for total persons with low-income as well as important subgroups of persons with low-income. The gains in sample size for persons and subgroups with low-income was accomplished while limiting the increase in variance of persons aged 55+ to only 5%, since this group was considered second in importance only to persons with low-income. The only low-income subgroup that showed the desired increase in sample size in the research was the number of Blacks in or near poverty, which showed an increase of 47%. Even if oversampling in the 1990 redesign doesn't give the SIPP very large sample size gains for all poverty subgroups, the gains are still significant and it does provide valuable experience in oversampling that could improve methods of oversampling in the future.

During research, we made the assumption that while the area frame would have only half the variance reductions of the unit frame, it would have all of the variance increases due to stratification-over-time. Implementation will help determine the contributions of the area frame much more accurately. Implementation should also help verify other assumptions.

In the research, stratification of PSUs singly rather than in groups was primarily due to PSU interviewer workload constraints. Theoretically, it is better to stratify many PSUs at once to reduce variability of weights. The optimum ratio of the sampling rate in stratum 1 to the sampling rate in stratum 2 had little variation from PSU to PSU. As a result, there would probably be little gain in stratifying several PSUs at once, so the implementation plans are to stratify within PSUs rather than form groups of PSUs and stratify within the groups. Due to the workload constraints at the PSU level, this plan is more advantageous overall for the SIPP at this time.

Oversampling has been, and will probably continue to be, an important methodology in the SIPP for improving reliability of many statistics. With uncertainty about the realization of the gains stated in this paper, the Bureau has defined a fall-back plan. If the oversampling methodology used for the 1995-2005 panels gives smaller gains than expected or if a self-weighting design is just more desirable, all of the 1995-2005 redesign panels have a built in option so a switch back to a self-weighting design can be accomplished at any time. Currently though, oversampling the low income population in the SIPP 1995-2005 sample panels is in the implementation stage at the Census Bureau.

* This paper reports general results of research undertaken by Census Burea staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.

## IX. Footnotes

[1] Waksberg, Joseph, "The Effect of Stratification With Differential Sampling Rates on Attributes of Subsets of the Population", Proceedings of the Social Statistics Section, American Statistical Association, pp. 429-434 (1973).

[2] The variance for persons 55+ was constrained since this group was considered second in importance only to persons < 150% of the poverty threshold. Also, Health Interview Survey (HIS) oversampling research in 1980

found that variances for persons 65+ in poverty increased significantly with a decrease in variance for poverty. For the HIS results see internal Census Burea memo from R. P. Chakrabarty to G. M. Shapiro entitled "HIS Redesign: Differential Sampling to Achieve a Reduction in Demographic Subgroup Variances." May 3, 1982.

TABLE 1
Percent Change in Variance for Selected SIPP Characteristics Based on Data from 27 Research PSUs

| Variable Name | Characteristic | Design Effect (DEFF) | Decrease (+/-) in Variance |
|---|---|---|---|
| y1 | Blacks Below 150% of the Poverty Level | 62% | -38% |
| y2 | Hispanics Below 150% of the Poverty Level | 78% | -22% |
| y3 | Female-headed Householders Below 150% of the Poverty Level | 84% | -16% |
| z1 | Number of Blacks (16+) | 71% | -29% |
| z2 | Number of Persons Residing in Urban Areas | 106% | +6% |
| z3 | Number of Owner-occupied Units | 91% | -9% |
| z4 | Number of Renter-occupied Units | 91% | -9% |
| z5 | Number of Owner-occupied Units with Value < $50,000 | 104% | +4% |
| z6 | Number of Renter-occupied Units with Rent < $150 | 73% | -27% |
| z7 | Number of Persons Age 55+ | 105% | +5% |
| z8 | Number of Persons Age 65% | 104% | +4% |
| z9 | Number of Female-headed Households | 90% | -10% |
| z10 | Number of Female-headed Households, No Spouse Present with one or More Own Children Under Age 18 | 53% | -47% |
| z11 | Number of Black-headed Households | 63% | -37% |
| z12 | Number of Hispanics (16+) | 90% | -10% |
| z13 | Number of Unemployed | 99% | -1% |
| z14 | Number of Black and Spanish Unemployed | 79% | -21% |
| z15 | Number in the Civilian Labor Force (CLF) | 104% | +4% |
| z16 | Number of Blacks in the CLF | 78% | -22% |
| z17 | Number of Persons Below 150% of the Poverty Level | 76% | -24% |
| z18 | Number of Persons Below the Poverty Level, not Receiving Public Assistance | 79% | -21% |
| z19 | Households with Household Income: <$5,000 | 82% | -18% |
| z20 | $5,000 - $9,999 | 97% | -3% |
| z21 | $10,000 - $14,999 | 104% | +4% |
| z22 | $15,000 - $24,999 | 109% | +9% |
| z23 | $25,000 - $34,999 | 108% | +8% |
| z24 | $35,000 - $49,999 | 116% | +16% |
| z25 | $50,000 - $74,999 | 112% | +12% |
| z26 | > $75,000 | 113% | +13% |
| z27 | Black Households With Household Income: < $5,000 | 62% | -38% |
| z28 | $5,000 - $9,999 | 78% | -38% |
| z29 | $10,000 - $14,999 | 78% | -22% |
| z30 | $15,000 - $24,999 | 81% | -19% |
| z31 | $25,000 - $34,999 | 86% | -14% |
| z32 | $35,000 - $49,000 | 88% | -12% |
| z33 | $50,000 - $74,999 | 90% | -10% |
| z34 | >$75,000 | 91% | -9% |
| z35 | Number of Households Below the Poverty Level | 78% | -22% |

[3] The poverty threshold is the amount of household income below which a household is considered in poverty. This threshold is a function of the total number of persons in the household and the number of children.

[4] This assumption results in only minor changes in the results as compared to assuming equal household sizes in the two strata.

TABLE 2

Coefficients of Variation at the National Level for Selected SIPP Characteristics[1]

| Characteristic | Estimate of Population Proportion | Non-Oversample CV % | Oversample CV% |
|---|---|---|---|
| Blacks in or near Poverty | 6.1% | 1.7% | 1.3% |
| Hispanics in or near Poverty | 5.6% | 5.6% | 4.9% |
| Female-headed Householders in or near Poverty | 4.3% | 3.3% | 3.0% |
| Number of Female-headed Households | 29.4% | 1.1% | 1.0% |
| Number of Female-headed Households, No Spouse Present with one or More Own Children Under Age 18 | 7.8% | 2.6% | 1.9% |
| Number of Black-headed Households | 13.4% | 1.8% | 1.4% |
| Number of Persons Below 150% of the Poverty Level | 20.1% | 0.9% | 0.8% |
| Households with Household Income: $35,000 - $49,999 | 8.3% | 2.4% | 2.3% |
| Households with Household Income: $50,000 - $74,999 | 2.7% | 4.2% | 4.3% |
| Households with Household Income: > $75,000 | 1.1% | 6.7% | 7.1% |
| Number of Households Below the Poverty Level | 22.0% | 1.3% | 1.2% |

1 Based on 20,000 Households and 53,200 persons in the sample.

$CV = \sigma_x / \bar{x}$

Table 3

Changes in SIPP Sample Sizes by MSA in the Unit Frame Due to Oversampling [a]

| MSA[a] | Blacks in or near Poverty (%)[b] | Hispanics in or near Poverty (%) | Female-headed Householders in or near Poverty (%) | Age 55+ (%) | Persons in or near Poverty (%) |
|---|---|---|---|---|---|
| 0120 | 126.294 | 123.389 | 125.096 | 102.539 | 120.880 |
| 0160 | 157.287 | 139.832 | 130.575 | 99.343 | 125.113 |
| 0200 | 136.014 | 120.406 | 122.913 | 100.319 | 115.993 |
| 0520 | 145.970 | 136.156 | 131.984 | 100.907 | 126.156 |
| 0920 | 132.211 | 118.454 | 121.507 | 99.735 | 115.480 |
| 0960 | 141.184 | 147.378 | 133.567 | 99.889 | 125.468 |
| 1280 | 147.649 | 138.491 | 128.437 | 99.291 | 126.042 |
| 1560 | 158.887 | 138.866 | 134.703 | 101.650 | 127.706 |
| 1760 | 121.488 | 135.475 | 122.893 | 101.380 | 117.872 |
| 1960 | 153.847 | 149.838 | 138.972 | 99.349 | 125.323 |
| 2335 | 147.844 | 133.039 | 127.149 | 98.602 | 119.126 |
| 2975 | 125.292 | 128.959 | 128.779 | 99.743 | 118.893 |
| 3560 | 166.783 | 170.835 | 134.473 | 98.539 | 126.642 |
| 4600 | 138.803 | 140.774 | 126.813 | 101.612 | 122.522 |
| 4640 | 129.929 | 113.781 | 126.853 | 102.126 | 122.092 |
| 5170 | 143.492 | 125.456 | 126.158 | 98.854 | 118.389 |
| 5480 | 132.349 | 127.813 | 119.858 | 100.269 | 115.659 |
| 5720 | 135.398 | 125.857 | 123.913 | 100.365 | 120.666 |
| 6025 | 131.687 | 114.522 | 120.604 | 98.956 | 115.767 |
| 6440 | 148.140 | 137.333 | 126.717 | 99.141 | 129.350 |
| 6480 | 149.181 | 141.889 | 127.165 | 99.688 | 122.678 |
| 6760 | 144.630 | 138.488 | 132.578 | 101.853 | 129.674 |
| 6880 | 146.394 | 118.264 | 126.220 | 100.895 | 123.060 |
| 6960 | 152.960 | 148.547 | 129.991 | 97.697 | 126.753 |
| 7520 | 129.787 | 115.639 | 124.506 | 101.195 | 120.283 |
| 8160 | 156.394 | 154.731 | 130.108 | 99.112 | 124.133 |
| 8840 | 153.896 | 144.231 | 128.265 | 98.778 | 121.786 |
| MSA AVG | 146.785 | 135.672 | 128.823 | 99.999 | 122.381 |
| Sample Size Before Oversampling[c] | 1,680 | 1,230 | NA | 14,180 | 6,920 |
| Sample Size After Oversampling | 2,640 | 1,670 | NA | 14,160 | 8,170 |

NA Not available

b Estimate = (sample sized with oversampling/sample size without oversampling)

a Different MSA numbers refer to different metropolitan statistical areas examined in oversampling research. In research these MSAs are considered to be PSUs.

b Poverty in this table refers to persons < 150% of the poverty threshold.

c Based on 20,000 Households and 53,200 persons in the sample.

Table 4
Effect of Time Changes on
Efficiency Due to Oversampling[1]

| Characteristic | Initial DEFF | Avg. DEFF 4 Years Later[2] | Avg. DEFF 8 Years Later[3] | Avg. DEFF 11 Years Later[4] |
|---|---|---|---|---|
| Blacks in or near Poverty | 62% | 60% | 63% | 61% |
| Hispanics in or near Poverty | 78% | 79% | 78% | 78% |
| Number of Persons in or near Poverty | 76% | 81% | 81% | 75% |
| Number of Persons Residing in Urban Areas | 106% | NA | 111% | 114% |
| Number of Renter-occupied Units with Rent < $150 | 73% | 79% | 84% | 79% |
| Number of Owner-occupied Units with Value < $30,000 | 104% | 101% | 106% | 102% |
| Households with Household Income: $35,000 - $49,999 | 110% | NA | 109% | 109% |

NA Not available

1 Average of two MSAs using American Housing Survey (AHS) Data

2 These changes in design effects are based on the 4 year period 1977 - 1981.

3 These changes in design effects are based on the 8 year period 1977 - 1985.

4 These changes in design effects are based on the 11 year period 1974 - 1985.

$$DEFF = \frac{\text{Variance with Oversampling}}{\text{Variance without Oversampling}} \times 100$$

Table 5
Effects on Design Effects (DEFFs) When Oversampling Poverty

| | DEFF before Adjustment[1] | Increase/Decrease (+/-) in DEFF Adjusting for MC/Area Frame | Increase/Decrease (+/-) in DEFF Stratification Over Time | DEFF with Combined Adjustments |
|---|---|---|---|---|
| Number of Blacks in or near poverty | 62% | +6% | +1% | 69% |
| Number of Hispanics in or near poverty | 78% | +2% | 0% | 80% |
| Number of Persons in or near poverty | 76% | +4% | +5% | 85% |

1 Using 1980 Census Data

2 These changes in design effects are based on the 8 year period 1977 - 1985. This period was chosen to be a worst case.

$$DEFF = \frac{\text{Variance with Oversampling}}{\text{Variance without Oversampling}} \times 100$$