CENSUS 2000 SAMPLE WEIGHTING

Miriam D. Rosenthal, Eric Schindler, and Alfredo Navarro[1]
Miriam D. Rosenthal, Bureau of the Census, Washington, DC 20233

Key Words: Raking, Mean-squared error, Nonresponse follow-up

Abstract The United States Census of Population and Housing collects basic demographic information of every resident enumerated in the census. Additionally, a sample of households receives a detailed questionnaire, which collects information on a wide range of social and economic topics. To produce full population and housing estimates for the sample, weighting areas are formed to calculate weights for the persons and housing units. Sample estimates for the whole population are produced using the person and housing-unit weights. In 1990, the raking-ratio estimation procedure ensured consistency between the sample estimates and census counts of data collected on a 100-percent basis.

A redesign alternative for Census 2000 is to conduct a sample-based nonresponse follow-up (NRFU) operation. A desirable objective of the Census 2000 sample design is to produce estimates with reliability comparable to 1990 with no increase in the overall sample of households receiving the detailed questionnaire. This paper explores the issue of weighting-area formation, specifically the size criterion as it relates to NRFU sampling and accuracy of sample estimates. We assess several weighting-area-formation schemes using exploratory data analysis methods and other efficiency criteria, such as mean-squared errors (mse) and variances of the estimates.

Introduction The goal for Census 2000 is a faster, less costly, and more accurate census. One of the most costly components of the census is nonresponse follow-up. To reduce cost, sampling of nonrespondents is likely for Census 2000. This paper explores possible ways to maintain accurate estimates when nonrespondents are sampled.

In 1990, there was a differential sampling rate for households receiving a detailed questionnaire (long form). Small governmental units with fewer than 2,500 persons were sampled at a 1-in-2 rate. Remaining tract and block numbering areas with less than 2,000 housing units and remaining list/enumerate areas were sampled at a 1-in-6 rate. (List/enumerate areas are sparsely populated areas where enumerators create address lists while collecting a completed questionnaire from each household.) Other more densely populated tracts and block numbering areas were sampled at a 1-in-8 rate.

Long-form households answered the same questions as the non-sampled (short-form) households. In addition, they answered detailed questions on their housing, demographics, education, employment, income, etc. A raking procedure created weights that summed to the total population and attempted to match long-form and short-form estimates for groups defined by race/Hispanic origin, tenure, age, sex, family type, and householder status.

Nonrespondents are those households that do not return the census questionnaire or do not respond in other ways during the mail-return period. In 1990, all nonrespondents were to be contacted, but for Census 2000, sampling nonrespondents is possible. A scenario referred to as 90-percent truncation is currently being proposed. In this scenario, 100-percent follow-up proceeds until 90 percent of the households have responded. Then 1-in-10 sampling of the remaining nonrespondents follows. Sampling nonrespondents has the effect of increasing the long-form estimates'

[1]The authors are mathematical statisticians in the Decennial Statistical Studies Division. 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.

variances. We examine whether increasing the size of the weighting areas helps to control this increase in variance.

A weighting area is a collection of contiguous blocks in the same county, and, when possible, in the same minor civil division (MCD) and tract. Blocks are combined up to the county level to create 400+ sample persons per weighting area. The raking procedure produces sample weights that sum to the 100-percent census counts at the weighting-area level.

Methodology This simulation uses one rural and one urban county from the 1990 Census files for Connecticut. The census files contain person level, housing-unit level, and geographic data for long-form and short-form households. In addition, we create auxiliary variables based on a reformatting of responses. An example of an auxiliary variable is "grfive," which is a true/false variable based on highest grade completed (whether completed fifth grade) for persons 25 years or over. Appendix A lists a profile of the rural and urban counties. Appendix B lists the long-form variables and auxiliary variables that we produce estimates for in this study.

First we produce weighting-area, MCD, and county estimates for the 1990 data based on the 1990 weighting areas. Throughout the study, we limit the MCD-level estimates to MCD's with population greater than 25,000 (only the urban county had MCD's that met this criterion) to distinguish the results from weighting-area-level results. Next we simulate the 90-percent-truncation scheme, which is the most likely nonresponse follow-up sampling scheme and produce weighting-area, MCD, and county estimates. Finally we regroup the weighting areas by pairs and repeat the complete nonresponse follow-up and 90-percent-truncation schemes and produce original weighting-area, county, and MCD estimates again.

Weights Initial weights for the complete nonresponse follow-up scheme, or modified-initial weights for the 90-percent-truncation scheme, fill the interior cells of the raking-ratio estimation (weighting) matrix. Short-form counts, or short-form NRFU weights for the 90-percent-truncation scheme, make up the marginal cells of the raking procedure. The raking procedure produces final weights, which are applied to the long-form persons to produce long-form totals that are consistent with short-form totals. Described here are the rules for assigning initial weights, NRFU weights, and modified-initial weights.

First we assign initial weights for the complete nonresponse follow-up simulation. Long-form households receive weights of 1/[observed probability of selection]. The observed probability of selection is the observed proportion of persons in a block group receiving and answering long-form sample questionnaires out of the total persons in a block group. A block group is a set of blocks with the same first digit in the Census Bureau's three-digit block numbering scheme. The initial weights hover around six, which is approximately the inverse of the long-form sampling rate. Persons who are in the short-form households receive an initial weight of zero.

Next we assign nonresponse follow-up weights to the households included in the 90-percent-truncation scheme. Up to the 90-percent level, long-form and short-form households receive a NRFU weight of one. After the 90-percent-truncation level, we use NRFU sampling rates of 1-in-6 for long-form and 1-in-12 for short-form households to approximate an overall 1-in-10 sample. Thus, long-form households included in the NRFU sample receive a NRFU weight of six, and short-form households included in the NRFU sample receive a NRFU weight of twelve. Households in the last ten percent, not included in the NRFU sample, receive a NRFU weight of zero.

Finally we assign modified-initial weights to the households included in the 90-percent-truncation scheme. Those households responding up to the 90-percent-truncation level have modified-initial weights the same as the complete nonresponse follow-up scheme initial weights. For the NRFU sample, short-form households receive a modified-initial weight of zero as initial weights. For the NRFU sample, short-form households receive a modified-initial weight of zero as before, but long-form households receive a modified-initial weight of the NRFU weight (about six) times the initial weight (also about six). Thus, the modified-initial weights for long-form persons in the NRFU sample are about 35 or 40.

Analysis Evaluating the new nonresponse follow-up and weighting-area-size schemes consists of measuring the changes in the variance and bias. We look at cv ratios and the Wilcoxon rank-sum test to assess variance, relative bias to quantify bias effects, and mean-squared error for an overview of the combined effects of variance and bias. Herein is a description of the sampling/weighting-area-size problem and an explanation of the tests we use.

For variance, we break down the sampling/weighting-area-size problem into parts, whereas for bias, we look at the overall effect of truncation and size. First controlling for larger weighting areas, we compare the variances of the complete versus truncated nonresponse follow-up schemes. This measures the cost in variance of sampling for nonresponse follow-up. Next controlling for truncation, we compare the variances of the original weighting areas with those of the larger weighting areas. This measures the gain in using larger weighting areas. Finally, we examine the combined effect of sampling and weighting-area size. Namely, we compare the variances of the complete nonresponse follow-up and original weighting areas with those of the truncated nonresponse follow-up and larger weighting areas.

Coefficient of variation (cv) ratios are a ratio of the new cv to the original cv. They provide a measure of the magnitude of the variance change. We look at cv ratios at the original weighting-area level and MCD level. A cv ratio of less than .8 is an improvement, .8-1 is not much improvement, and greater than one is a loss. The most extreme cv ratios are greater than two.

Using the Wilcoxon rank-sum test, the two scenarios, e.g., original weighting-area size versus larger weighting-area size, are combined and ranked. Appendix B lists the variables--sixteen long-form characteristics crossed by seven groups--whose variances we rank. Every weighting area does not have estimates for every combination of characteristic and group. At the original weighting-area level, we compare the sum of the ranks of the variances of the two scenarios of interest. Then, the sum of the ranks between the two scenarios are compared, where a difference in rank sums indicates a difference in variances i.e., the hypothesis we are testing is H_0: variances are equal versus H: variances are not equal.

Besides assessing the effect on variance, we monitor bias. We examine bias at the original weighting-area, MCD, and county levels. The larger weighting areas introduce bias because larger weighting areas create more heterogeneous groupings. Assuming the 1990 census is the truth, we take the difference between the truncated follow-up/larger weighting areas' NRFU average jackknife estimate and the complete follow-up/original weighting areas' NRFU estimate to get an estimate of the bias. Then we determine the relative bias: what percent the bias is of the complete follow-up/original weighting areas' NRFU estimate.

$$\text{relative bias} = \frac{\overline{jk}_{2000} - \text{estimate}_{1990}}{\text{estimate}_{1990}} \times 100$$

Furthermore, we look at the mean-squared error of the truncated follow-up/larger weighting area and compare it to the variance of the complete follow-up/original weighting area to assess the combined bias and variance effect. We look at the mean-squared error at the original weighting-area, MCD, and county levels.

Results The cv ratios for the urban county at the weighting-area level and mcd level are listed in tables 1 and 2. As expected, most truncation-effect cv ratios are greater than one (63% at the weighting-area level, 76% at the MCD level); whereas, most size-effect cv ratios are less than or equal to one (55% at the weighting-area level, 53% at the MCD level). In addition, the truncation effect drives the combined truncation and weighting-and 75% at the MCD level). Plus, for all effects, the average cv ratios are greater than one, indicating increases in variance outweigh any observed decreases in variance. The rural county's results are similar.

Table 1 CV Ratios: Ranges, Averages, Medians for Truncation, Size, and Overall Effects; Urban County at the Weighting-Area Level

CV ratios	Truncation	Size	Overall
< .8	18%	33%	23%
.8 - .9	7%	11%	8%
.9 - 1	8%	11%	8%
1 - 1.2	18%	14%	15%
1.2 - 2	34%	18%	27%
> 2	11%	8%	14%
CV's=0	4%	5%	5%
Average	1.35	1.13	1.37
Median	1.14	.91	1.08

Table 2 CV Ratios: Ranges, Averages, Medians for Truncation, Size, and Overall Effects; Urban County at the MCD Level

CV ratios	Truncation	Size	Overall
< .8	10%	22%	12%
.8 - .9	5%	16%	6%
.9 - 1	8%	15%	6%
1 - 1.2	19%	19%	19%
1.2 - 2	44%	22%	42%
> 2	13%	5%	14%
CV's=0	1%	1%	1%
Average	1.41	1.13	1.44
Median	1.24	.97	1.25

We use the Wilcoxon rank-sum test to test the significance of the effects at the original weighting-area level. The test is done at the weighting-area level because there are more observations to compare at the weighting-area level than at the MCD level. First, we test the effect of the 90-percent-truncation scheme (see table 3a, cell 4). We expect truncation to increase variances. For both the rural and urban counties at the weighting-area level, we reject the null hypothesis at the α=.05 level that the variances are equal (p=.04, rural county;p=.0001, urban county) i.e., the variances increase for the truncated nonresponse follow-up scenario.

Tables 3a and 3b Comparisons of Interest -- Wilcoxon Rank-Sum Test Results
Rural and Urban Counties at the Weighting-Area Level
H_0: Variances are equal
H_1: Variances are not equal

Table 3a		Complete follow-up	
		Original weighting areas	Larger weighting areas
Truncated follow-up	Original weighting areas	Not calculated (1)	Not applicable (2)
	Larger weighting areas	Rural county: p=.1692 Urban county: p=.0002(3)	R. county: p=.0400 U.county: p=.0001 (4)

Table 3b	Truncated follow-up/ Original weighting areas
Truncated follow-up/ Larger weighting areas	Rural county: p=.4783 Urban county: p=.0021

Next, we use the Wilcoxon rank-sum test to test the effect of using larger weighting areas (see table 3b). We expect the larger weighting areas to decrease variances. For the rural county at the weighting-area level, we cannot reject the null hypothesis at the α=.05 level that the variances are equal (p=.4783), but for the urban county the larger weighting area does decrease the variance significantly at the α=.05 level (p=.0021). The urban county has many more observations than the rural county. If the rural county had had more observations, we conjecture it too would have shown significant differences.

Finally, we use the Wilcoxon rank-sum test to test the combined effect of truncation and larger weighting-area size (see table 3a, cell 3). For the rural county at the weighting-area level, we cannot reject the null hypothesis at the α=.05 level that the variances are equal (p=.1692). However for the urban county at the weighting-area level, we reject the null hypothesis at the α=.05 level (p=.0002). When we look at the sum of the ranks, we observe that the ranks are greater for the truncated follow-up/larger weighting areas scheme. For the urban county, the larger weighting-area size does not offset the increase in variance incurred when sampling nonrespondents. Again, while the rural county tends toward larger ranks, the urban county shows significant results.

We also determine the relative bias of the truncated follow-up/larger weighting areas compared to the complete follow-up/original weighting areas. The relative bias at the weighting-area level for the rural county ranges from -50 percent for a variable with one observation to 136 percent for a variable with two observations. The largest relative bias for variables with observations from all 23 weighting areas is 50 percent. Overall, most of the variables have a relative bias of less than ten percent.

At the county level, the largest relative bias for the rural county is 262 percent for American Indians--proportion in poverty, aged 65+. Since variables have only one observation at the county level there is more variability in the bias. There is one other variable with relative bias over 100 percent. Most of the variables have an overall absolute relative bias of less than ten percent.

At the weighting-area level for the urban county, the relative bias ranges from -11 percent to 138 percent. The urban county has many more weighting areas, up to 173 observations per variable. The largest relative bias for variables with observations from all 173 weighting areas is 12 percent. The largest MCD-level relative bias is 74 percent. The county-level relative bias is even less with the greatest absolute relative bias being 27 percent.

To look at the combined effects of bias and variance at the weighting-area level, we compare the truncated follow-up/larger weighting areas' mean-squared error to the complete follow-up/original weighting areas' variance (see table 4). The truncated follow-up/larger weighting areas do better than the complete follow-up/original weighting areas for only 14 to 31 percent of the observations.

Table 4 Percent of Total Observations Where MSE of the Truncated Follow-up/Larger Weighting Areas <= Variance of the Complete Follow-up/Original Weighting Areas

	Weighting-area level	MCD level	County level
Rural county	31 percent	not applicable*	25 percent
Urban county	26 percent	14 percent	16 percent

*The MCD level estimates are limited to MCD's with population greater than 25,000. The rural county's MCD's do not meet this criterion.

Conclusions In this study, we researched whether in Census 2000 increasing the weighting-area size can compensate for the increased variance of long-form estimates caused by sampling nonrespondents. We found increasing weighting-area size as implemented in this study not to be an effective mechanism to reduce variance. Following are the detailed conclusions and other suggestions for improving variances.

From the cv ratios, we learned that increasing the weighting-area size, although somewhat helpful by itself in improving variances, is not a panacea for variance woes caused by sampling. Both the weighting-area level and MCD level have overall cv ratios greater than one. Most discouraging is that the MCD-level cv ratios do worse than the weighting-area-level cv ratios. For practical purposes, we are trying to reduce variances for publication-like estimates, which are similar to MCD-level estimates.

The combined effects of truncation and increasing weighting-area size are more pronounced in the urban county than for the rural county. For the urban county, increasing the weighting-area size for the truncated-sampling scheme does not provide enough decrease in variance to get estimates comparable to the 1990 decennial census. For the rural county, the variances for the complete follow-up/original weighting areas and truncated follow-up/larger weighting areas are not significantly different, but they tend in the same direction as the urban county.

The combined mean-squared error measurements are not very encouraging either. As few as 14 percent of the observations for the truncated follow-up/larger weighting areas have smaller mean-squared errors than the complete follow-up/original weighting area scenarios. Perhaps increasing the weighting-area size even more, calculating the initial weights based on race, or collapsing cells during the raking process will improve the long-form estimates' variances.

References Navarro, Alfredo and Griffin, Richard, 1990, "Sample Design for the 1990 Decennial Census," Proceedings of the Section for Survey Research Methods, 1990 Meeting of the American Statistical Association.

Schindler, Eric and Griffin, Richard, 1992, "Weighting the 1990 Census Sample," Proceedings of the Section for Survey Research Methods, 1992 Meeting of the American Statistical Association.

Schindler, Eric and Navarro, Alfredo, 1996, "Effect of Sampling for Nonresponse Follow-up on Estimates from Sample Data," Proceedings of the Section for Survey Research Methods, 1996 Meeting of the American Statistical Association.

Appendix A A Profile of Two Connecticut Counties

Based on 1990 Census Counts:	Rural	Urban
Proportion White	95.9%	84.6%
Proportion African American	1.1%	9.8%
Proportion American Indian	.3%	.2%
Proportion Asian Pacific Islander	.7%	2.1%
Proportion Other Race	2.0%	3.3%
Proportion Hispanic	4.2%	8.6%
Proportion in Poverty	8.0%	6.1%
Proportion <Grade 5	1.7%	2.0%
Proportion High School Graduate	71.1%	81.0%
Proportion Some College	37.8%	55.2%
Proportion College Graduate	16.8%	34.2%
Proportion Renters	33.4%	31.8%
Proportion 1-in-2 Sampling Rate	8.5%	.2%
Proportion 1-in-6 Sampling Rate	49.8%	74.4%
Proportion 1-in-8 Sampling Rate	41.7%	25.4%
Mail Response Rate for Occupied Housing Units	71.7%	73.7%

Appendix B Variables: (Long-Form Characteristic) x (Group)

Long-Form Characteristic

Proportion Foreign Born
Proportion Non-English Speaking Home
Proportion Moved Last 5 Years
Proportion High School Dropout, Aged 16-19
Proportion <Grade 5, Aged 25+
Total Civilian Labor Force
Total Worked Last Week
Proportion in Labor Force
Proportion Unemployed
Proportion Used Public Transportation
Per Capita Income
Proportion in Poverty
Proportion in Poverty, Aged 0-4
Proportion in Poverty, Aged 5-17
Proportion in Poverty, Aged 65+
Proportion in Poverty, Families

Group

Total Population	African American
American Indian	Asian/Pacific Islander
Hispanic	Aged 0-17
Aged 65+	