METHODOLOGY FOR THE EVALUATION OF SAMPLING AND ESTIMATION IN THE CENSUS

Mary H. Mulry, and Alfredo Navarro*, U.S. Bureau of the Census
Mary H. Mulry, U. S. Bureau of the Census, Washington, D.C. 20233

Key Words: Nonresponse, Coverage error

1. Introduction

The Bureau of the Census is testing the use of sampling and estimation in two aspects of census taking in the 1995 Census Test. One application of sampling is for the followup of the nonrespondents to the mail questionnaires in the 1995 Census Test,and the other is for integrated coverage measurement. The motivation for testing sampling for nonresponse followup is to reduce cost while the aim of the integrated coverage measurement is to reduce the differential undercount.

The concept of sampling for nonresponse to the census mail questionnaires has been recommended by the National Performance Review (1993) and two panels of the National Academy of Sciences, the Panel to Evaluate Alternative Census Methods (1994), and the Panel on Census Requirements for the Year 2000 and Beyond (1994). The recommendation of the Panel on Census Requirements for the Year 2000 and Beyond is that "the Census Bureau make a good-faith effort to count everyone, but then truncate the physical enumeration after a reasonable effort to reach nonrespondents. The number and characteristics of the remaining nonrespondents should be estimated through sampling."

For the 1995 Census Test, the truncation of the physical enumeration happens with the close of the response period for the mail questionnaires. Then a sample of the nonrespondents is selected for followup. However, this criterion will not necessarily be the one defined for the 2000 Census. The Census Bureau has been reengineering the Census (U. S. Bureau of the Census, 1995a). The leading scenario calls for nonresponse followup to be conducted for every nonresponding housing unit until a county is 90 percent complete. Then the nonresponse followup is supposed to be truncated and continued for a 1-in-10 sample of the remaining nonresponding units. An alternative proposal for reengineering calls for conducting nonresponse followup until 70 percent of the county is completed and then sampling the remaining for continued followup (U. S. Bureau of the Census, 1995b). Also part of the reengineering proposals for the 2000 Census is that data from administrative records may be used in combination with sampling to compensate for some nonresponse. Studies associated with the 1995 Census Test will contribute to the decision on the use of administrative records.

While there are policy considerations in the use of sampling for nonresponse followup(NRFU), the 1995 Census Test will provide technical information about the sample design, the sample size, and the estimation. One goal of the 1995 Census Test is to test two basic sampling designs for nonresponse followup in the Oakland test site. The issue is whether to select a sample of nonresponding housing units without regard to geographic clustering (housing unit sample) or to select a sample of blocks and include all the nonresponding housing units in the blocks in the sample (block sample). Overall the sampling rate for the Oakland test site was two-sevenths of the housing units which did not respond with half of the sample allocated to each design. In the test sites in Paterson, NJ and the six parishes in Northwest Louisiana, only a block sample was conducted, and one-sixth of the nonresponding housing units was selected for followup.

The Census Bureau is testing integrated coverage measurement (ICM) in the 1995 Census Test because it is expected to reduce the differential coverage error observed in previous censuses. One goal of the 1995 Census Test is to test two methodologies for integrated coverage measurement. The primary issue is whether a new methodology known as CensusPlus which uses ratio estimation is effective. Another issue is to test dual system estimation which was used for the 1990 Post Enumeration Survey (PES) as an alternative to CensusPlus with ratio estimation. One way of evaluating the effectiveness of the two methodologies is to determine whether they add persons in the traditionally undercounted groups to the census numbers. The evaluations will be performed using data collected in the Oakland test site.

This paper discusses the major evaluations of sampling and estimation in the 1995 Census Test. A description of the data collection and processing methodology for the enumeration, nonresponse followup and ICM can be found in Mulry and Singh (1994). The evaluation of sampling for nonresponse followup will assess its effectiveness by investigating the coverage properties and other aspects of two basic sampling designs, a block sample and a housing unit sample. The evaluation of the integrated coverage measurement will focus on measuring data collection and processing errors plus determine whether the procedure adds persons in the traditionally undercounted groups to the census numbers. Other evaluations of ICM are directed at developing refinements to the basic methodology or evaluations

of specific operations. However, space does not permit discussing these other evaluations in this paper.

In this paper, Section 2 discusses the sample design for the nonresponse followup and ICM. Section 3 describes the methodology for the evaluation of coverage differences in the two sample designs. Section 4 discusses the evaluation of integrated coverage measurement. Section 5 is a summary.

2. Sample Design

Each sample design being tested for nonresponse followup appears to have its advantages and disadvantages. From simulations with 1990 Census data, the sample based on housing units appears to have less bias and variance for small areas(Isaki, Tsay, Fuller, 1994). However, the block sample may be easier to implement in conjunction with integrated coverage measurement(ICM). This is important because ICM also is being tested in the 1995 Census Test for use in the 2000 Census, and ICM requires a block sample. Research to improve the small area estimation with the block sample is in progress. The difference in field costs for the two designs is unclear, and the test is being designed to collect information on cost for a comparison. The methodology for the cost comparisons is not be discussed in this paper.

The technical evaluation of sampling for nonresponse in the 1995 Census Test will focus on whether there is any difference in coverage between the housing unit sample design and the block sample design. Also, the evaluation will investigate whether there are other differences in the quality of estimates produced from a sample based on blocks or housing units. Identifying refinements and enhancements to the methodology for sampling for nonresponse follow-up is an additional goal.

The sample selection requires four steps. The first step was to combine blocks to form clusters with a minimum of 30 housing units. The block clusters were used as the sampling units. Blocks with more than 40 housing units were not combined with other blocks, and neither will they be split or subsampled. Next the block clusters were stratified and divided into two panels. Half of each stratum will go to each panel. The panels are for the sampling for nonresponse, one for the unit sample and one for the block sample. Each panel fully reflects the heterogeneity of the population in the sites. Interviewers in each panel were of comparable skill levels. Interviewers were assigned to panels at random, or they were assigned work in each panel. Then a new stratification for sampling was defined within each panel. The definition of the strata used size of the blocks and the proportion in categories defined by race and Hispanic ethnicity. One half of the ICM sample was selected from each panel. The last step is to select the samples for nonresponse in each panel from the clusters not selected for the ICM. The unit sample was a systematic sample which selected two-sevenths of the nonresponding units. The block sample was a systematic sample of blocks within strata which selected two-sevenths of the blocks.

Using these samples, estimates for entire sites and for poststrata defined by age/sex/race/Hispanic ethnicity/tenure within sites will be made after nonresponse followup and after ICM. The poststrata are defined by the following groups:

Race/Origin (4)
- Non-Hispanic White and Other
- Black
- Asian and Pacific Islander (API)
- Non-Black Non-API Hispanic

Tenure (2)
- Owner
- Non-owner

Age/Sex (7)

0 to 17	Male and Female
8 to 29	Male
18 to 29	Female
30 to 49	Male
30 to 49	Female
50 & over	Male
50 & over	Female

The estimation after nonresponse followup for the comparison of the two sample designs will be the sum of the weighted responses after imputation for item nonresponse. The estimation after nonresponse followup for use in the ICM estimates will be a ratio estimator. This estimator may include data from both designs in Oakland if there are no coverage differences observed.

3. Evaluations for Nonresponse Followup

An assessment of differences in coverage between the housing unit sample design and the block sample design will examine the coverage of whole households, within households, and the two combined by examining population estimates for the site and for demographic groups. A comparison of the average size of occupied housing units for each design also will illustrate whether there is a difference in coverage within households for the two designs. Comparisons operational data related to coverage of housing units will indicate whether the two designs exhibit differences in various aspects of coverage. For example, the percentage of added units and vacant units will indicate whether there is a difference in the coverage of whole households. Sex ratios for the whole site and for the age/sex/race/Hispanic origin groups for each design will be compared with each other to provide another indication of whether there is a difference in coverage.

3.1 Whole Household Coverage

3.1.1 Vacancy Rate

To examine whole household coverage differences in the two panels, we will compare the proportion of housing units identified to be vacant. The unweighted and weighted vacancy rates for the NRFU universe will be calculated by panel. We will test to determine whether the proportion of vacant housing units is the same in both panels. To describe the statistical tests, we need the following definitions for the block sample:

- H = the number of sampling strata in the block sample.
- W_h = the sampling weight in the block for sampling stratum h, h = 1,...,H.
- B_h = the number of blocks in the block sample in stratum h.
- NR_{hj} = the number of nonresponding addresses in the j-th block in the block sample in stratum h, j = 1,...,B_h.
- A_{hj} = the number of units added in the j-th block in stratum h in the block sample.
- D_{hj} = the number of units deleted in block j in stratum h in the block sample.
- V_{hj} = the number of units vacant in block j in stratum h in the block sample.
- AV_{hj} = the number of vacant added units deleted in block j in stratum h in the block sample.

We also need the following definitions for the unit sample.

- K = the number of sampling strata in the unit sample.
- w_k = the sampling weight for stratum k in the unit sampling, k = 1,...,K.
- nr_k = the number of nonresponding addresses in the unit sample in stratum k.
- a_k = the number of housing units added in stratum k for the unit sample.
- v_k = the number of vacant units in stratum k for the unit sample among the original sample addresses.
- av_k = the number of vacant units in stratum k for the unit sample among those added during the operation.
- d_k = the number of addresses in stratum k in the unit sample which are deleted because there is no housing unit.

We will compare the proportion of the housing units that are vacant in the unit and block samples, $\hat{p}_{b,0}$ and $\hat{p}_{u,0}$, respectively. Note that the $\hat{p}_{b,0}$ and $\hat{p}_{u,0}$, as defined below, do not include the blocks selected for ICM.

$$\hat{p}_{b,0} = \frac{\sum_{h=1}^{H}\sum_{j=1}^{B_h} W_h V_{hj} + \sum_{h=1}^{H}\sum_{j=1}^{B_h} W_h AV_{hj}}{\sum_{h=1}^{H}\sum_{j=1}^{B_h} W_h NR_{hj} + \sum_{h=1}^{H}\sum_{j=1}^{B_h} W_h A_{hj} - \sum_{h=1}^{H}\sum_{j=1}^{B_h} W_h D_{hj}}$$

$$\hat{p}_{u,0} = \frac{\sum_{k=1}^{K} w_k v_k + \sum_{k=1}^{K} w_k av_k}{\sum_{k=1}^{K} w_k nr_k + \sum_{k=1}^{K} w_k a_k - \sum_{k=1}^{K} w_k d_k}$$

The variances of $\hat{p}_{b,0}$, $\hat{p}_{u,0}$, and their difference will be calculated using jackknife estimators and VPLX software (Fay, 1990).

The hypothesis which we will test is

$$H0: \hat{p}_{u,0} = \hat{p}_{b,0} \quad \text{vs} \quad H1: \hat{p}_{u,0} \neq \hat{p}_{b,0}$$

with the test statistic

$$T = (\hat{p}_{u,0} - \hat{p}_{b,0}) / Var(\hat{p}_{u,0} - \hat{p}_{b,0}).$$

3.1.2 Add and Delete Rates

The rate at which housing units are added to the address list by panel and their characteristics will be compared if there is sufficient data. The number of additional housing units is expected to be very small because the interviewers were not trained to look for missed housing units. They were instructed to add units if they found them. However, for completeness, the study is including a comparison of the household distribution of added units for the site and for subgroups if there is sufficient data.

Another aspect of whole household coverage is the identification of addresses which do not represent housing units, called deletes. An address could be deleted for reasons such as it is a business with no living quarters; the dwelling has been demolished; or two apartments have been merged into one. The proportion of addresses that are deletes for the block and unit samples will be compared.

3.2 Within Household Coverage

3.2.1 Average Household Size

To determine whether there is any difference in coverage within households between the two sample designs, we will compare the estimates of average household size produced under each design. In addition, the household size distributions will be compared to examine whether there are any underlying patterns of coverage differences. To describe the

statistical tests, we need the following definitions for the block sample:

I_{hji} = 1, if address i in block j in stratum h is an occupied housing unit,
0, otherwise (vacant or not a housing unit).

IA_{hji} = 1, if the i-th added housing unit in block j in stratum h is occupied,
0, otherwise (vacant).

X_{hji} = the number of persons at the i-th nonresponding address in the j-th block in stratum h in the block sample,

AX_{hji} = the number of persons at the i-th unit added during nonresponse followup in block i in stratum h, i = 1,..., A_{hj}.

We also need the following definitions for the unit sample.

x_{ki} = the number of persons in the i-th nonresponding address stratum k in the unit sample, i = 1,...,nr_k.

ax_{ki} = the number of people in the i-th housing unit added in stratum k for the unit sample, i = 1,...,a_k.

We will compare the average household size in the occupied units in the block and unit samples, $\bar{y}_b$ and $\bar{y}_u$, respectively. Note that $\bar{y}_b$ and $\bar{y}_u$, as defined below, are ratio estimates of the average household size which do not include the blocks selected for ICM. Also, the addresses within each panel have equal sampling weights.

$$\bar{y}_b = \frac{\sum_{h=1}^{H}\sum_{j=1}^{Bh}\sum_{i=1}^{NRhj} W_h X_{hji} + \sum_{h=1}^{H}\sum_{j=1}^{Bh}\sum_{i=1}^{Ahj} W_h AX_{hji}}{\sum_{h=1}^{H}\sum_{j=1}^{Bh}\sum_{i=1}^{NRhj} W_h I_{hji} + \sum_{h=1}^{H}\sum_{j=1}^{Bh}\sum_{i=1}^{Ahj} W_h AI_{hji}}$$

$$\bar{y}_u = \frac{\sum_{k=1}^{K}\sum_{j=1}^{nrk} w_k x_{kj} + \sum_{k=1}^{K}\sum_{j=1}^{ak} w_k ax_{kj}}{\sum_{k=1}^{K} w_k (nr_k + a_k - v_k - d_k - av_k)}$$

The variances of $\bar{y}_b$, $\bar{y}_u$, and their difference will be calculated using jackknife estimators and VPLX software (Fay, 1990).

The hypothesis which we will test is

$$H0: \bar{y}_u = \bar{y}_b \quad \text{vs} \quad H1: \bar{y}_u \neq \bar{y}_b$$

with the test statistic

$$T = (\bar{y}_u - \bar{y}_b) / Var(\bar{y}_u - \bar{y}_b) .$$

A similar statistical test will be performed for each of several population subgroups defined by race, Hispanic origin, and tenure. An example is a comparison of the average household size for housing units occupied by renters in the two panels. These tests will indicate whether there is a difference in within household coverage for the whole population and for demographic subgroups of the population.

3.2.2 Household Size Distribution

Another aspect of the investigation of differences in within household coverage between the two panels is a comparison of the distribution of household size. If a difference in coverage is detected, more analysis is required to identify more specifically the factors lending to the difference.

First we will perform a 2x9 chi-square test to determine whether the distributions of the proportions of household size from the occupied housing units in the block and unit samples are the same. If they do not appear to be the same, we will perform other tests in an attempt to pinpoint the source of coverage differences.

3.3 Sex Ratios

The ratios for the number of men to the number of women for the total population and for various population subgroups by panel will also be compared to determine whether there is a difference in coverage. Taylor Series approximations of variances will be calculated to perform statistical hypothesis testing.

3.4. Other Comparisons

Other evaluations examine whether there other differences in the quality of estimates obtained from sampling blocks or units for nonresponse followup and the validity of the experimental design. Comparisons with other sources of data give information about the validity of the results from each sample design. Since the sample design calls for each site to be split into panels and then sampled, estimates from each design may be made for the entire site. For each design, the sex ratios for the whole site and for the groups defined by age/sex/race/Hispanic origin will be compared with updated estimates from demographic analysis and with 1990 Census data. This will provide information about the effectiveness of each of the designs for sampling for nonresponse followup.

We also will check some operational data to validate the underlying assumption of the experimental design that we have two comparable panels. The operational data that will be used for the comparisons include the percentage of questionnaires that failed the edit check, the item nonresponse rates, and the percentage that are CATI interviews. If the data appear of comparable quality, then the comparisons of the designs will be considered valid.

For example, since Computer Assisted Telephone Interviewing (CATI) will be attempted prior to some personal interviews, a large difference in the percentage of CATI interviews could cause interviewing mode effects to look like treatment effect, or cause other discrepancies in the data.

4. Integrated Coverage Measurement

Each estimation methodology appears to have its advantages and disadvantages. The CensusPlus method appears as though it can be completed more easily by the December 31 deadline. However, CensusPlus requires assuming that the reinterview is "truth." While the dual system methodology may take longer to complete the data collection, it assumes that the reinterview is another list of the population, independent from the census list but not necessarily the truth. The Census Bureau has experience with the dual system methodology, and found that the dual system estimate generally agreed with demographic analysis estimates in 1990. However, the 1995 implementation of the dual system methodology is different in some important ways from the 1990 methodology which may cause different results. On the other hand, since the Census Bureau does not have any direct experience with CensusPlus, it is unknown as to whether these numbers will be of sufficient quality.

The evaluation of the persons added has two objectives:

(1) Determine whether the ICM methodology appears to add people in the traditionally undercounted groups in the 1995 Census Test.

(2) Determine which methodology, CensusPlus or Dual System, appears to add more people in the traditionally undercounted groups.

The characteristics of the persons added by the ICM methodology after nonresponse followup will be analyzed in two ways. One is at the overall site level for each site, and the other is by block clusters. Examining the census numbers at the site level indicates what types of persons are being added on the whole. Investigating whether the block clusters with additions are the ones that are expected to be hard to enumerate provides some insight into the validity of the methodology.

4.1 Site Level Analysis

The site level analysis will use the estimates for each poststratum in a site. Estimates for all 56 poststrata will be made for Oakland. There are not enough Asian and Pacific Islanders in Paterson to support separate poststrata so APIs will be included in the Non-Hispanic White and Other poststrata. In Paterson, there will be 42 poststrata.

To calculate the persons added by ICM, we first have to calculate the estimate of the population after nonresponse followup which was discussed in Section 3. After nonresponse followup an estimate of the population size may be made for a site,

$\hat{C}_j$ = estimate for poststratum j from the initial enumeration and nonresponse followup.

After the ICM, estimates of the population size may be made for a site,

$\hat{C}_j^+$ = CensusPlus estimate for poststratum j

$\hat{D}_j$ = dual system estimate for poststratum j.

The following differences in the poststratum totals for the estimators and their variances will be calculated:

$$\hat{C}_j^+ - \hat{C}_j, \quad \hat{D}_j - \hat{C}_j, \text{ and } \hat{D}_j - \hat{C}_j^+ .$$

These differences $\hat{C}_j^+ - \hat{C}_j$ and $\hat{D}_j - \hat{C}_j$ will be examined to determine if the ICM is adding people in the traditionally undercounted groups. Note that the "additions" could also be deletions. The difference $\hat{D}_j - \hat{C}_j^+$ will indicate which methodology is adding more people in these groups, CensusPlus or dual system. Another comparison will be based on the ratio of

$$(\hat{C}_j^+ - \hat{C}_j) / (\hat{D}_j - \hat{C}_j).$$

This ratio will indicate the success of CensusPlus in adding people relative to the success of dual system.

The second aspect of the analysis compares the differential rates of additions with the differential undercounts estimated in the sites by the 1990 PES. This examination aims to determine whether the same pattern of missing people is present in the 1995 Census Test as was found in the 1990 Census. The addition rates will be calculated for groups defined by race/Hispanic ethnicity/tenure for CensusPlus and dual system. The undercount rates for the same groups will be calculated in each site using 1990 PES data with the 357 poststrata. The standard of comparison for calculating the differential rates for 1995 and 1990 will be the non-Hispanic white owners. Each set of differential rates will be ranked from lowest to highest. Then the ranks will be compared using the rank correlation test, a distribution-free test.

4.2 Block Cluster Level Analysis

This analysis checks whether ICM is adding persons in the block clusters considered hard to enumerate. If ICM appeared to be adding persons in areas considered easy to enumerate, further investigation would be necessary to assess the

validity of the methodology. The analysis considers only the ICM block clusters and unweighted data. Block level estimates from the weighted data will not be produced until Spring, 1996.

The unweighted data for CensusPlus will be assessed by the difference between the resolved rosters, RE_b and the number from self-response and nonresponse followup, C_b in a block cluster defined by

$$Diff(RE_b , C_b) = RE_b - C_b.$$

The unweighted data for dual system will be assessed by the difference between the sum S_b of the matches, the other correct enumerations, and the nonmatches and the self-responses C_b which reduces to the difference between the nonmatches NM_b and the erroneous enumerations EE_b.

$$Diff(S_b , C_b) = S_b - C_b = NM_b - EE.$$

For the block cluster level analysis, the block clusters will be assigned two ranks. One ranking will be according to $Diff(RE_b , C_b)$. The other ranking will be based on $Diff(S_b , Cb)$. The ranks will be assigned in decreasing order of the values. The rankings based on $Diff(RE_b , C_b)$ and $Diff(S_b , Cb)$ will be compared with two other rankings designed to indicate how hard a block cluster is to enumerate. The comparison of the sets of rankings will use the rank correlation test. One set of rankings are those assigned by the planning data base. These rankings are already available.

The other set of rankings must be calculated for this project by using estimate the probability of being enumerated in the census based on conditional logistic regression models developed as part of the evaluation of the 1990 Post Enumeration Survey (Alho, Mulry, Wurdeman, and Kim, 1993). With these models, probabilities of enumeration will be estimated for every person in a block cluster. The people who have an enumeration probability lower than 0.75 will be considered hard to enumerate. The block clusters will be rank in descending order of the number of people hard to enumerate. Other ways of ranking the blocks based on the probabilities of inclusion also will be considered.

5. Summary

The decisions based on the results of the 1995 Census Test will determine the form of the 2000 Census. The decisions have to be made by the end of 1995 so that the Census Bureau can procure equipment and space for the census. These evaluations will provide the major technical information for the decisions concerning the basic design for sampling for the followup of nonrespondents to the mail questionnaires. Research will continue after these decisions to refine the sample selection and estimation methodology.

References

Fay, Robert E. (1990) "VPLX: Variance Estimates for Complex Samples, " *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 266-290.

Isaki, C., Tsay, J., and Fuller, W. (1994) "Design and Estimation for Samples of Census Nonresponse," *Proceedings of the 1994 Annual Research Conference*, U. S. Bureau of the Census, 289-305.

Mulry, M. H. and Singh, R. P. (1994) "New applications of Sampling and Estimation in the 1995 Census Test," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, 742-747.

National Performance Review (1993) "From Red Tape to Results: Creating a Government that Works Better & Costs Less," U. S. Government Printing Office.

Panel on Census Requirements in the Year 2000 and Beyond (1994) *Modernizing the U. S. Census*, B. Edmonston and C. Schultze, eds., Committee on National Statistics, National Research Council, Washington, D. C.: National Academy Press.

Panel to Evaluate Alternative Census Methodologies (1994) *Counting People in the Information Age*, D. L. Steffey and N.M. Bradburn, eds., Committee on National Statistics, National Research Council, Washington, D. C.: National Academy Press.

U. S. Bureau of the Census (1995a) "The Reengineered 2000 Census," unpublished manuscript, dated May 19, 1995.

U. S. Bureau of the Census (1995b) "Reengineered 2000 Census Alternative," draft, unpublished manuscript, dated June 6, 1995.

* 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.