FINAL SAMPLE DESIGN FOR THE CENSUS 2000 ACCURACY AND COVERAGE EVALUATION

Randal ZuWallack, Matthew Salganik, Ryan Cromar and Vincent Thomas Mule, Jr., U.S. Census Bureau
Randal ZuWallack, U.S. Census Bureau, Washington, DC 20233-7600

Keywords: Stratification; Systematic sampling; Multi-phase sampling; Census undercount

Introduction

Every ten years, the Census Bureau attempts to enumerate every person living in the United States. Although a complete count is desired, past experience indicates it's virtually unattainable. According to past census evaluations using demographic analysis, the undercount has ranged from 2.8 million in 1980 to 7.5 million in 1940 (Bureau of the Census, 1997). Beginning with the 1950 census, the Census Bureau has also been conducting post-enumeration evaluations to estimate census coverage. These evaluations take a case by case matching approach to identify people who were missed and those who were counted. More recent evaluations of this type include the 1980 Post-Enumeration Program (PEP) and the 1990 Post-Enumeration Survey (PES). For the PEP, information based primarily on the Current Population Survey was used to estimate people not counted in the census enumeration (Fay, 1988). A second part of the PEP involved selecting a sample of census records to estimate the number of erroneous census enumerations. Improvements were introduced for the 1990 PES. Rather than using information that was not specifically designed for measuring census omissions, a survey was designed with this sole purpose in mind. As was done in 1980, a sample was also selected for estimating erroneous census enumerations.

In the tradition of improving census evaluations, the Census Bureau is conducting the Accuracy and Coverage Evaluation (A.C.E.) following the Census 2000 enumeration. Similar to the PES, the A.C.E. checks the quality of the census in two ways. One is by comparing data from the census to data collected from an independent sample of housing units to estimate the number of people missed. The other is by selecting a sample of census records to estimate the number of erroneous census enumerations. This information is combined to determine dual system estimates for many sub-populations, which are then compared to the census results to estimate coverage rates. This paper, building on the previous paper by ZuWallack, Salganik and Mule (1999), discusses all phases of the A.C.E. sample design, how the design was affected by a Supreme Court decision on sampling for the census (Department of Commerce v. United States House of Representatives, 1997), and changes made to the design based on an evaluation of the Census 2000 Dress Rehearsal design.

This paper reports the results of research and analysis undertaken by Census Bureau Staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

P Sample and E Sample

Because there are two types of coverage errors, missed people and erroneous inclusions, two samples are selected to evaluate census coverage: the population sample (P Sample) and the enumeration sample (E Sample). The P Sample consists of the people living in housing units interviewed for the A.C.E. These units are randomly selected from the Independent List, an address list compiled independently of the census list for a sample of geographic areas. The P-sample people are matched back to the census to determine if they were counted or missed. The E Sample consists of people living in a sample of housing units enumerated in the census. The E-sample people are checked to determine whether they were correctly counted in the census, or whether they were erroneously enumerated. Examples of erroneous enumerations include duplicates, fictitious names, people who were born after census day or people who died prior to census day.

Table 1. P Sample and E Sample Comparison

	P Sample	**E Sample**
Estimates	Omissions	Erroneous Enumerations
Universe	All housing units in US[1]	Census housing units
PSUs	Block Clusters	Block Clusters

Block Cluster

The primary sampling units (PSUs) of both the P Sample and E Sample are block clusters, which are one or more geographically contiguous census blocks grouped together. Census blocks are formed by streets, roads, railroads, streams, etc. Forming block clusters involves a complicated hierarchical algorithm involving many rules and constraints. In general, the goal of block clustering

[1]Housing units in remote areas of Alaska are excluded.

is to produce primary sampling units that average about 30 housing units.

Integrated Coverage Measurement Survey

Until January 25, 1999, when the Supreme Court ruled that statistical sampling could not be used for congressional reapportionment, the Census Bureau had planned to conduct an Integrated Coverage Measurement (ICM) Survey. The primary goal of the ICM was to produce accurate and reliable direct state estimates, which would then be used for the reapportionment. Preliminary calculations indicated that the ICM allocation may result in coefficients of variation for the dual system estimates of approximately 0.5% in all states and standard errors of about 60,000 in the larger states (Schindler, 1998).

The Supreme Court ruling changed these requirements. Direct state estimates were no longer needed for the reapportionment process, and consequently neither was a 750,000 housing unit sample. In contrast to the ICM, which incorporates the survey results into the population estimates, the A.C.E. produces a second set of estimates used to evaluate the census and potentially for other purposes.

Because the Supreme Court ruling came too late to entirely redesign the sample, we selected an initial sample of block clusters using the ICM design. Then, to make the transition from the 750,000 housing unit design to the 300,000 housing unit design, we developed a subsampling operation called the A.C.E. Block Cluster Reduction. There are some inefficiencies to selecting the ICM sample and modifying it with this reduction. The ICM was designed for efficient direct estimates for state total population, while the primary goal of the A.C.E. is to generate reliable demographic group estimates for the purpose of measuring differential coverage. The ICM sample was selected using proportional allocation within a state. While this might be efficient for total population estimates, it is not necessarily efficient for estimating the population of smaller demographic groups. However, regardless of the inefficiencies, we expect the reliability to be better for most of the A.C.E. poststrata estimates than the 1990 PES, due to an increased sample size. We also expect the total state population estimates to be more reliable than in 1990.

Stratification and Sort Variables

Historically, coverage rates in the census have varied for many different groups in the population. In 1990, coverage rates were calculated for 357 poststrata identified by region, geographic area, race, Hispanic origin, age, sex, and tenure (own/rent). Although the estimated undercount for the total population was 1.6%, the estimated undercounts for the 357 groups ranged from -8.29% to 21.27% (Thompson, 1992). The poststrata definitions for Census 2000 (Schindler and Haines, 2000) were still being researched at the time the sample was being designed. However, we assumed they would be similar to those used for 1990.

To estimate the coverage rates for several poststrata with acceptable precision, there must be an adequate amount of sample selected for each of these poststrata. Since the characteristics of people within a block cluster vary, exact sample sizes are unattainable. However, the variation in the sample sizes can be improved by grouping similar block clusters together and selecting a systematic sample across these groups. To better control the sample sizes from these different groups, block clusters were classified into categories based on their estimated size, demographic composition, and level of urbanization.

Block clusters initially were stratified into four mutually exclusive groups within each state: small block clusters (0-2 housing units), medium block clusters (3-79 housing units), large block clusters (80 or more housing units), and American Indian Reservation or trustlands (AIR) block clusters. These strata were sampled at different rates during the selection of the A.C.E. listing sample.

Although there was no differential sampling within these four strata, the clusters were sorted by several variables to select a diverse set of block clusters. The first sort variable is the American Indian indicator, which has three categories: 1) AIR, 2) tribal jurisdiction statistical area, Alaska Native Village statistical area or tribal designated statistical area, 3) all other areas. The second sort variable is demographic group. Block clusters were grouped with other block clusters containing similar demographic proportions based on 1990 census data. This variable is described in more detail in the following paragraph. A third variable used for sorting the clusters is the level of urbanization. Each block cluster was categorized as an urbanized area with 250,000 or more people, an urbanized area with less than 250,000 people, or a non-urban area. Finally, the clusters were sorted geographically.

To aid in selecting a sample that is well represented by the six major race/origin groups (Hawaiian or Pacific Islander; American Indian or Alaska Native; Asian; Hispanic; Black; White or Other) as well as owners and renters, block clusters were classified into 12 demographic groups. Although many block clusters tend to have a large proportion of one demographic group, they are rarely composed of only one, thus many clusters may fit well in two or more categories. To ensure that each cluster is assigned to only one group, a hierarchical assignment rule was developed so that when a cluster exceeds the group threshold, it is assigned to that group. These thresholds were developed by grouping similar

1990 blocks together using a multivariate clustering method.[2] The resulting thresholds range from 10% for some of the smaller populations to 30%. The order of the hierarchy gave the smaller demographic groups priority over the larger ones and renters priority over owners.

A.C.E. Listing Sample Selection

For each state, a systematic sample of block clusters was selected for each of the four strata listed in the previous section. A list of housing units was created in the selected clusters, which is called the independent list. As stated earlier, the Census Bureau was preparing to conduct an ICM during the early stages of the sample design. Thus, 25,000 block clusters were allocated to the states to approximately meet the ICM sample requirements, while maintaining a minimum of 300 block clusters per state. The sampling was done in two steps to guard against a listing workload that would be too formidable to complete in time. If the first systematic sample of block clusters resulted in a housing unit workload that was 10% more than the predetermined maximum, a second systematic sample was drawn from the first to approximately meet the listing constraint. Large block clusters were selected at a higher rate than medium clusters during the A.C.E. listing sample selection. These higher rates coupled with large block subsampling result in more clusters in sample while keeping the total number of interviews within budget.

Small block clusters were generally sampled at a lower rate than both medium and large clusters. This is due to the cost considerations explained in a later section. These lower sampling rates cause some small clusters to have high weights, which may disproportionately affect the dual system estimates. In an attempt to avoid the problems associated with the high weights we will initially sample 5,000 small block clusters. Using information about these clusters, we target clusters where there is the potential for problems due to high weights in the small block cluster subsampling operation. These initial 5,000 small clusters were allocated to states proportionally to their projected number of housing units in small blocks. This allocation was bounded by two constraints for each state: a 20 block cluster minimum and a minimum expected sampling rate of 1 in 1000.

To ensure sufficient sample for calculating accurate undercount rates for American Indians on reservations, 355 block clusters were selected from the block clusters on AIR nationwide. Small block clusters on AIR were not included in these 355 block clusters, but were eligible for selection in the small cluster stratum. The 355 clusters were allocated to 26 states proportionally to the 1990 population of American Indians on reservations. Ten states contained AIR clusters with little or no American Indian population. These clusters were not included in an AIR stratum, but instead were eligible for selection in one of the other strata. The remaining 14 states and the District of Columbia contain no block clusters on AIR.

The selection of the A.C.E. Listing Sample for the 50 states, the District of Columbia, and Puerto Rico resulted in 29,695 block clusters containing nearly 2 million housing units.

A.C.E. Block Cluster Reduction

As previously stated, the ICM sample is converted to the A.C.E. sample via the A.C.E. Block Cluster Reduction. This process is the first of three operations that together will reduce the housing units to approximately 300,000, which is nearly twice the sample size of the 1990 PES. The A.C.E. reduction is a subsample of medium and large block clusters in the 50 states and the District of Columbia. Small block clusters, AIR clusters and Puerto Rico are not subsampled during this operation. The housing unit sample was allocated to the states and the District of Columbia proportionally to state population, with a minimum of 1,800 housing units per state.

While reducing the number of block clusters, the A.C.E. reduction was designed with two goals in mind: to provide sufficient sample to support reliable estimates for several sub-populations and to reduce the variance contribution due to clusters potentially with high omission or erroneous enumeration rates. These clusters were identified by comparing the consistency between the independent list and the census list for each cluster. To achieve these goals, medium and large clusters in each state were divided into minority and non-minority as well as low inconsistent, high inconsistent and consistent. Minority clusters are those that have a high concentration of minorities, determined by the demographic group explained previously. Low inconsistent clusters are those where the census list is at least 25% larger than the independent list and high inconsistent clusters are those where the independent list is at least 25% larger than the census list. Based on these groups, we formed four strata per state: Minority, Non-minority Low Inconsistent, Non-minority High Inconsistent, and Non-minority Consistent.

In order to achieve the two goals outlined above, clusters in the first three were sampled at higher rates relative to the fourth. We called these relative rates differential sampling factors. For example, a differential sampling factor of 2 for minorities means that the

[2]PROC FASTCLUS in SAS® uses a multivariate clustering technique called nearest centroid sorting . Refer to pages 824-850 of the SAS/STAT® User's Guide, Volume 1, Version 6, Fourth Edition.

probability of selection for minority clusters is twice that of non-minority consistent clusters.

The differential sampling factors were assigned using guidelines designed to achieve the two goals of the reduction, while also controlling the size of the weights and the amount of differential sampling. Generally, the amount of differential sampling was controlled so that no stratum was sampled at a rate that was more than three times that of the consistent stratum and so that the expected weights for the consistent stratum remained at or below 650. Once the strata sampling factors (k_i) were determined, we calculated the strata sample sizes (n_i) for each state as follows:

$$n_i = n \times \frac{k_i \times HU_i}{\sum_4 k_i \times HU_i}$$

where n = state housing unit sample size
HU_i = weighted number of housing units listed in the i^{th} stratum

The final piece of the reduction design concerns stratum jumpers. Stratum jumpers are clusters that we initially classified as medium, but their independent listing count is 80 or more housing units. These clusters were sampled at the medium rate during the listing sample selection, but will also go through large block cluster subsampling, resulting in a large weight relative to other medium and large clusters. To keep the weights for these clusters down, all are retained during the reduction.

Following the A.C.E. Reduction, the sample for the 50 states, the District of Columbia, and Puerto Rico was down to 15,324 block clusters containing about 900,000 housing units.

Small Block Cluster Subsampling

Small block clusters, those originally estimated to have between 0 and 2 housing units, are not a cost-effective workload for interviewing and follow-up operations. In order to wisely use our resources, we generally sampled small clusters at a lower rate than both medium and large clusters. Thus, people in small clusters tend to have higher weights than people in medium and large clusters. These high weights can disproportionately affect the dual system estimates. In 1990 only about 2.4% of the P sample people and 1.7% of the E sample people lived in small clusters. Yet these clusters contributed almost 10% to the net undercount and 15% to the estimated variance (Fay, 1998). To improve our estimates we have developed the following design component for small clusters.

Initially we selected 5,000 small clusters during the A.C.E. listing sample selection. Then, through the small block cluster subsampling operation, we reduced the number of small clusters in sample to about 1,500 while at the same time attempting to achieve two goals. First, we would like to prevent small clusters from having weights that are extremely high compared to other clusters in the sample. Second, we would like to limit the weights on clusters which might have a disproportionate effect on the variance of the dual system estimates.

Generally, we stratified the small clusters into four groups based on the number of housing units in the cluster. The strata were formed so that clusters with more housing units than we initially expected would be sampled at a higher rate to keep their weights lower. Our measure of size was the larger of the number of housing units on the census address list and the number on the independent address list, which gives us a conservative estimate for the number of housing units that are actually in the cluster. To determine the sampling rates for these strata, we imposed two conditions so that our sample would have the properties we desired. Table 2 contains a summary of the sampling conditions.

One of these conditions was that, if possible, the number of weighted housing units in a cluster did not exceed 2,400 housing units. Through computer simulations we tried a number of different limits until we found that a cap of 2,400 yielded a sample of the appropriate size. The second condition was a minimum sampling rate, which varied among the four strata. Clusters in the first stratum had a minimum sampling rate of 1 in 10. The minimum rates for the remaining strata decrease relative to the increase in the maximum number of housing units. The resulting rates are 1 in 4 and 1 in 2.2 for the second and third strata, respectively. All clusters in the fourth stratum remain in sample.

Using these two conditions, we calculated the sampling interval for each stratum and then took a systematic sample of small clusters for each state. The equations below illustrate the process for calculating the sampling intervals for each stratum (r_i). The first equation imposes the first condition of having a 2,400 weighted housing unit cap if possible, otherwise the interval is 1. The second restricts the intervals to the minimum rates. The two equations are:

$$r_i' = \max\left\{1,\ \frac{2400}{W_i \times M_i}\right\}$$

$$r_i = \min\left\{r_i',\ \frac{M_1}{M_i} \times 10\right\}$$

where W_i = i^{th} stratum cluster weight prior to small cluster subsampling
M_i = Maximum housing units in a cluster in the i^{th} stratum

Additionally, several groups of small clusters were retained with certainty regardless of the number of housing units on either list. To reduce weight variation in the American Indian poststratum, all small clusters on AIR, tribal jurisdiction statistical area, Alaska Native Village statistical area, or tribal designated statistical area were retained. Additionally, small clusters in very remote areas were retained. We wanted the weights on these clusters to be as low as possible due to the potential operational difficulties in enumerating these clusters and since they are outside of the scope of the Targeted Extended Search (Navarro and Olson, 2000).

Following small block cluster subsampling, the sample for the 50 states, the District of Columbia, and Puerto Rico was down to 11,802 block clusters, still containing about 900,000 housing units.

Table 2: Small Block Cluster Subsampling Parameters

Stratum (i)	Cluster HU Size	Max. HU (M_i)	Wt. HU Cap	Min. Rate
1	0-2	2	2,400	1 in 10
2	3-5	5	2,400	1 in 4
3	6-9	9	2,400	1 in 2.2
4	10+	---	---	1 in 1

Large Block Cluster Subsampling

Large block cluster subsampling is the final phase in selecting the housing units designated for an A.C.E. interview. The underlying concept of large block cluster subsampling is to increase the number of clusters in sample, while still remaining within the targeted number of housing units for interview. Because housing units in a cluster are often similar, interviewing all of them is not the most efficient use of resources. Instead, interviewing a smaller piece of several different clusters provides a more geographically diverse sample.

This phase involves selecting a portion of each block cluster containing 80 or more housing units. Housing units are selected by dividing each large cluster into segments of adjacent housing units that differ by no more than one housing unit. We determine the number of segments in a cluster to ensure that each cluster has at least one segment selected. The segment size is tied to the number of segments and the number of housing units in a cluster. For example, suppose there are two clusters, one with 90 housing units and the other with 120. If the clusters are sampled at a rate of 1 in 2.5, then three segments are formed in each cluster. This ensures that at least one segment containing 30 units is selected from the first cluster and at least one segment containing 40 units is selected from the second.

Before selecting the sample of segments, block clusters were divided into seven strata within each state. The first four strata are the same strata used for the A.C.E. block cluster reduction. The fifth is the medium to large stratum jumpers, or clusters previously thought to be medium, but based on updated housing unit counts turned out to be large. The sixth is small to large stratum jumpers. The seventh is AIR clusters, whose housing units all remain in sample.

As stated previously, large clusters were selected with higher probability during the listing sample selection. Thus, the large block subsampling rate is determined so that there would be close to equal weighting for housing units in medium and large clusters within each reduction stratum and state, while remaining within the targeted number of units for interview.

A sample of segments is selected by taking one systematic sample across all large clusters in each stratum within a state. All housing units in the selected segments are designated for A.C.E. interview. A complication of this operation is that large clusters are ready for subsampling on a flow basis as previous field work is completed. To remain on a schedule, it is essential that the sampled units get returned to the field for interviewing as quickly as possible. Thus, following the flow of block clusters from the field, the subsampling will be performed daily until all eligible block clusters have been subsampled. Despite the daily processing, the subsampling is equivalent to a one-time sample since the results of the previous day are carried over to the next and continued. The one difference with the daily operation is the inability to control the cluster sort since the order is determined by the flow of clusters from the field.

Following large block cluster subsampling, the sample for the 50 states, the District of Columbia, and Puerto Rico was 11,802 block clusters containing about 315,000 housing units.

E Sample Identification

Once the housing units have been selected for A.C.E. interview the next operation is to select the housing units that are in the E Sample. The information gathered from these housing units is used to estimate the number of erroneous inclusions in the census. Although an overlapping P Sample and E Sample is not necessary, it is more cost efficient. If the E Sample and P Sample include many of the same people, we can use the information from the P-sample interview to determine whether they were correctly enumerated or not, and thus whether or not a follow-up visit is necessary.

In an attempt to create overlapping samples, and thus save money, the block clusters and segments of block clusters that are used to select the P Sample are mapped onto the census address list. If this step yields any cluster requiring more than 80 follow-up visits, the E-sample housing units in these clusters are subsampled.

Changes from Census 2000 Dress Rehearsal

In 1998, the Census Bureau conducted a Dress Rehearsal to refine the Census 2000 operations. The Dress Rehearsal revealed a few areas in the sample design that needed improvement. Many of the changes were minor operational details, but there are a few enhancements worth noting, many involving small blocks.

The first change involves the formation of block clusters. Small blocks were not clustered with their neighbors for the Dress Rehearsal, whereas, they were under certain conditions in 2000. This reduced the total number of small clusters, thus reducing their weights. Overall, this change reduced the number of small clusters by about 65%, from just under 3 million to slightly over 1 million. Under the new clustering procedure the initial weights for housing units in small clusters vary from 25 to 632 with an average of 221. Had improvements not been made, they would have ranged from 56 to 1,010 with an average of 588.

The small block cluster subsampling had several differences. For instance, having the same target weight for small clusters in every state reduces the amount of weight variation between the states. This was an important change since unlike in the Dress Rehearsal, information is combined across states for Census 2000 estimates. A second change from the Dress Rehearsal was having a maximum weight that small clusters did not exceed. If the small block cluster subsampling plan had been used from the Dress Rehearsal, the maximum weight would have been as large as 6,170, rather than the current 1,200 maximum. Finally, there was the introduction of graduated sampling rates based on cluster size. In the Dress Rehearsal, clusters were either selected with certainty or at a rate of 1 in 10, where the cutoff between the rates was 10 housing units. Thus, a cluster with 9 housing units would have a weight that was ten times that of a cluster with 10 housing units. For 2000, the weight differential between small clusters differing by one unit is only about two or three.

Much of the A.C.E. operational planning was based on 1990 Census data. For instance, the estimated number of housing units for creating the Independent List for each state was estimated from 1990 information, and then used for renting office space and hiring staff in different areas of the country. Hence, exceeding these numbers may cause workload problems. To help keep the actual listing close to the estimated listing, two adjustments were built into the 2000 design that were not needed for the Dress Rehearsal. The first involves an adjustment prior to selecting a sample. If it appears the listing would be too much based on the preliminary sampling rate, then the sampling rate was decreased. The second adjustment comes in the form of a two step sample. If the clusters selected during the first step surpass the expected listing, a second sample from the first sample is selected. Without these two procedures, the actual listing would have surpassed the estimated listing by over 7.5 percent.

The improvements made from the Dress Rehearsal to Census 2000 illustrate the evolution of census coverage evaluations. There is no doubt the design will continue to evolve with the 2010 Census and beyond. However, the current A.C.E. sample design provides the framework to produce reliable estimates of coverage for Census 2000.

References

Bureau of the Census. (1997). Report to Congress: Plan for Census 2000. Washington, D.C.: Bureau of the Census.

Department of Commerce v. United States House of Representatives, No. 98-404 (U.S. filed Jan. 25, 1999).

Fay, R.E. (1988), "Evaluation of Census Coverage from the 1980 Post Enumeration Program (PEP): Census Omissions as Measured by the P Sample," Census Bureau Memorandum, March 10, 1988.

Fay, R. E. (1998), "Small Blocks in the 1990 PES," Census Bureau Memorandum, August 1998 (DRAFT).

SAS Institute Inc., *SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 1*, Cary, NC: SAS Institute Inc., 1989. 943 pp.

Navarro, F. and Olson, D. (2000), "2000 Census Accuracy and Coverage Evaluation Targeted Extended Search," *Proceedings of Survey Research Methods Section, American Statistical Association*, Alexandria, VA, American Statistical Association, to appear.

Schindler, E. (1998), "Allocation of the ICM Sample to the States for Census 2000," *Proceedings of Survey Research Methods Section, American Statistical Association*, Alexandria, VA, American Statistical Association, pp. 593-598.

Schindler, E. and Haines, D. (2000), "Poststratification for the Census 2000 Accuracy and Coverage Evaluation Survey," *Proceedings of Survey Research Methods Section, American Statistical Association*, Alexandria, VA, American Statistical Association, to appear.

Thompson, J. (1992), "CAPE Processing Results," Census Bureau Memorandum, March 20, 1992.

ZuWallack, R., Salganik, M. and Mule Jr., V.T. (1999), "Sample Design for the Census 2000 Accuracy and Coverage Evaluation," *Proceedings of Survey Research Methods Section, American Statistical Association*, Alexandria, VA, American Statistical Association, pp. 501-506.