SAMPLE REDESIGN ISSUES FOR ROTATING PRIMARY SAMPLING UNITS

Lawrence R. Ernst, Bureau of the Census*
Washington, DC 20233

KEY WORDS: household surveys, PSU rotation, maximizing overlap

1. INTRODUCTION

The household surveys conducted by the Census Bureau are redesigned approximately every ten years. Some primary sampling units (PSUs) are too small to provide enough distinct sample housing units for a survey for a decade. To avoid additional respondent burden that would result from reuse of sample during the decade, such PSUs are permitted to be in sample for only a portion of the decade. If selected in the current design, they are rotated in and out of sample using a procedure known as the Random Arc Method (RAM), described in Alexander, Ernst and Haas (1982). As a result, PSUs with insufficient sample for a decade are commonly referred to as "rotating" PSUs. Such PSUs will be referred to here as "small" PSUs instead, because RAM sometimes requires that PSUs with sufficient sample for a decade also rotate.

The rotation of PSUs is expensive and inconvenient, usually requiring the training of a new interviewer. Consequently, procedures which reduce the expected number of PSUs rotated into sample during the decade are desirable. Two such procedures are considered in this paper.

First, in Section 2, a new procedure for rotation of small PSUs is presented, which is applicable when each small PSU in a stratum has a large enough population to provide sample for at least half the decade. This procedure, which we call the Half-interval Method (HIM), is shown to minimize the expected number of rotations under these conditions, which RAM does not.

In addition to avoiding reuse of sample for the same household survey during a decade, the Census Bureau avoids reusing sample in other household surveys. Section 3 is concerned with PSUs that are large enough to provide distinct sample for a decade for any one survey, but not for all surveys. A procedure is presented for insuring that no PSU of this type is selected for more surveys than it is able to provide sufficient distinct sample, thus avoiding the need to rotate such PSUs.

The final topic discussed in this paper, in Section 4, is a modification of the procedure for maximizing the expected overlap of PSUs when a stratum has small PSUs. The Census Bureau has used in the past, and is planning to use in the 1990s' redesign, a methodology which increases the probability, in comparison with independent selection, of sample PSUs from the previous design being retained in sample (Causey, Cox and Ernst 1985), (Ernst 1986), thereby reducing additional costs, such as the expense of training a new interviewer, generally incurred with each change of sample PSU. For small PSUs it may be considered undesirable to maximize overlap, since the retention of a small PSU could result in many sample housing units from the previous design being also selected in the new design with a short time gap. In Section 4 options are presented which allow for the overlap of large PSUs to be maximized, while either minimizing overlap of small PSUs, or treating small PSUs in a neutral fashion with respect to maximization and minimization of overlap.

Due to space limitations a portion of Section 3 is omitted. The complete paper (Ernst 1991) is available from the author.

2. HALF-INTERVAL METHOD (HIM)

HIM is presented for one PSU per stratum designs in Section 2.1. HIM and RAM are compared in Section 2.2. In Section 2.3 an outline of HIM for two PSUs per stratum designs is presented.

2.1 HIM for One PSU Per Stratum Designs

In presenting the results for RAM in Alexander, Ernst, and Haas (1982), simplifications resulted by assuming that sampling was done continuously and uniformly in an interval of T months, $[0,T)$. Although HIM can readily be explained in terms of a discrete number of sampling periods, the continuous convention will be adopted here to conform with the description of RAM. For both procedures, R denotes the number of times a PSU is rotated into sample during $(0,T)$, that is with the PSU entering sample at time 0 not counted as a rotation.

Let S be a stratum consisting of u PSUs, ordered in increasing size. Let n denote the number of sample housing units required for the entire interval $[0,T)$. It is assumed that only the first v PSUs consist of fewer than n housing units. These PSUs will be denoted as small PSUs, the remaining PSUs as large PSUs. Let m_i be the population of the i-th PSU, $i=1,...,u$, and p_i the probability of selection of the i-th PSU, which is assumed proportional to m_i.

Let i_1, i_2 be the random variables denoting the sample PSUs from S for the months $[0,T/2)$ and $[T/2,T)$ respectively. i_1 is always selected by choosing from among all u PSUs with probability proportional to size. i_2 is selected differently for the three cases, $v\geq 3$, $v=2$, $v=1$. We proceed for each case to present the selection method for i_2; demonstrate that $P(i_2=i)=p_i$, $i=1,...,u$; compute the expected number of rotations; and prove that this expected number is minimal among all rotation schemes.

Case 1. $v \geq 3$. If $i_1 > v$, then $i_2 = i_1$. If $i_1 \leq v$, then i_2 is selected from among the remaining rotating PSUs i, $i=1,...,v$, $i \neq i_1$, with probability proportional to

$$p_i'\left(\frac{1}{1-2p_{i_1}'}+\frac{1}{1-2p_i'}\right),$$
where

$$p_i'=\frac{p_i}{\sum_{j=1}^{v} p_j}, \qquad i=1,\ldots,v.$$

Note that clearly $P(i_2=i)=p_i$, $i>v$. For $i\le v$, it follows from Durbin's method for selecting 2 PSUs per stratum without replacement (Cochran 1977) that $p_i'=P(i_2=i\,|\,i_1\le v)$ and hence $p_i = P(i_2=i)$.

Also observe that

$$E(R)=\sum_{i=1}^{v} p_i, \tag{2.1}$$

since there is one rotation if $i_1 \le v$ and none otherwise. For any unbiased rotation scheme there must be at least one rotation whenever a small PSU is in sample at time 0, an event which occurs with probability $\sum_{i=1}^{v} p_i$. Consequently HIM minimizes the expected number of rotations for this case.

Case 2. $v=2$. If $p_1<p_2$, then $i_2=1$ cannot always occur whenever $i_1=2$. Consequently, unlike Case 1, we must allow for the possibility of a large PSU, which we arbitrarily take to be PSU 3, to rotate. The selection procedure is then as follows:

If $i_1 \ge 4$, then $i_2 = i_1$.
If $i_1 = 1$, then $i_2 = 2$.
For $i_1 = 2$,
$P(i_2=1\,|\,i_1=2)=p_1/p_2$,
$P(i_2=3\,|\,i_1=2)=1-p_1/p_2$.
For $i_1 = 3$,
$P(i_2=2\,|\,i_1=3)=(p_2-p_1)/p_3$,
$P(i_2=3\,|\,i_1=3)=1-(p_2-p_1)/p_3$.

It is straightforward to show that $P(i_2 = i) = p_i$, $i=1,\ldots,u$, for this procedure.

Also,

$$E(R) = 2p_2, \tag{2.2}$$

since there is one rotation if $i_1 = 2$ or $i_2 = 2$ and no rotations otherwise.

To establish that HIM minimizes $E(R)$ among all unbiased rotation schemes, observe that there is a rotation from PSU 2 to the next PSU in sample when PSU 2 is in sample at time 0, which occurs with probability p_2. Similarly, with probability p_2 there a rotation from the sample PSU proceeding PSU 2 whenever PSU 2 is the final sample PSU. Consequently $E(R) \ge 2p_2$ for any unbiased rotation scheme in this case.

Case 3. $v=1$. If $i_1 = 1$ or $i_2 = 1$ then a large PSU, which we take to be PSU 2, must be in sample for the other half of the time interval $[0,T)$. The conditional selection probabilities are then as follows:

If $i_1 \ge 3$, then $i_2 = i_1$.
If $i_1 = 1$, then $i_2 = 2$.
For $i_1 = 2$,
$P(i_2=1\,|\,i_2=2)=p_1/p_2$,
$P(i_2=2\,|\,i_1=2)=1-p_1/p_2$.

Again, it easy to show that $P(i_2 = i) = p_i$, $i=1,\ldots,u$. Since there is one rotation if $i_1 = 1$ or $i_2 = 1$ and no rotations otherwise, it follows that

$$E(R) = 2p_1. \tag{2.3}$$

The proof that HIM minimizes the expected number of rotations in this case is the same as in Case 2, except that PSU 1, now replaces PSU 2.

2.2 Comparison of HIM and RAM

A brief description of RAM will first be provided. Further details on this procedure appear in Alexander, Ernst and Haas (1982). The expected number of rotations for HIM and RAM will then be compared.

To implement RAM for a stratum S, a cluster is first formed consisting of all small PSUs in S. If the set of all small PSUs in S is not sufficient in size to provide sample for all of $[0,T)$, then a single large PSU is added to the cluster. For PSU selection, the cluster is initially treated as a single PSU with probability of selection equal to the sum of the probabilities of selection of the individual PSUs in the cluster. If the cluster is selected, then at any time during $[0,T)$ a PSU in the cluster will be in sample. The method of determining which PSU in the cluster is in sample at a particular time is as follows. Form a circle with each PSU in the cluster corresponding to an arc in the circle. The length of the arc for the i-th PSU in the cluster is Tm_i/n which is the number of months of sample that the i-th PSU can provide. Select a random starting point x on the circle. At any time $t\in[0,T)$, the location on the circle will be t units measured clockwise from x. If this point is located on the arc corresponding to PSU i, then this PSU will be in sample at time t.

In Theorem 3.3 of Alexander, Ernst and Haas (1982), it is proven that if the rotation cluster consists of the first k PSUs in the stratum, then the expected number of rotations conditioned on the cluster being selected is $nk/\sum_{i=1}^{k} m_i$ for RAM, and consequently the unconditional expected number of rotations is

$$E(R) = \frac{nk\sum_{i=1}^{k} p_i}{\sum_{i=1}^{k} m_i}, \tag{2.4}$$

or alternatively, since p_i/m_i is the same for all i,

$$E(R) = \frac{nkp_i}{m_i}, \qquad i=1,...,k. \qquad (2.5)$$

Now assuming the conditions for which HIM is applicable, that is $m_i \geq n/2$, $i=1,...,u$, we have $k=v$ if $v\geq 2$, and $k=2$ if $v=1$, since a large PSU need only be included in the cluster for RAM if $v=1$.

The ratio of $E(R)$ for HIM to $E(R)$ for RAM will now be obtained for the three cases in Section 2.1. From (2.1) and (2.4) it follows that in Case 1 this ratio is

$$\frac{\sum_{i=1}^{v} m_i/v}{n},$$

or the ratio of the mean population per small PSU to n. Since $m_i < n$, $i=1,...,v$, this ratio is less than 1.

Since $k=2$ for both Cases 2 and 3, we obtain from (2.2), (2.3) and (2.5) that the ratio of $E(R)$ for HIM to $E(R)$ for RAM is m_2/n for Case 2 and m_1/n for Case 3. In both cases these ratios are less than 1.

2.3 HIM for Two PSU Per Stratum Designs

The simplest approach is to treat all the PSUs involved in the rotation as a single cluster. The cluster will consist of all small PSUs if $v\geq 3$. If $v<3$ then it will consist of all small PSUs and the smallest large PSU. Select the pair of PSUs, using any appropriate selection method, with the cluster treated as a single PSU with probability equal to the sum of the probabilities of the individuals PSUs in the cluster. If the cluster is one of the selections, then i_1 is chosen by selecting a PSU in the cluster with probability proportional to size, and i_2 is chosen using the conditional selection probabilities in Section 2.1. Note that the cluster size must be less than half of the stratum size if the two PSUs are chosen without replacement with probability proportional to size.

The major drawback to the approach just presented is that unbiased variance estimates would not be possible, since the joint probability of any pair of PSUs in the cluster being in sample at the same time would be zero. An alternative approach which avoids this problem will be sketched. The idea is to form two clusters from the small PSUs in which half the small PSUs are assigned to one cluster and half to the other, with one cluster receiving an extra PSU if there are an odd number of small PSUs. Any cluster with less than three small PSUs would have a large PSU added to it. The method of assignment of these PSUs to the two clusters must be such that each pair of PSUs has a positive probability of being in different clusters and hence in sample together. The size of each cluster must be less than half the stratum size. If there is only one small PSU in the entire stratum, then two clusters of small PSUs cannot be formed. Instead, a single cluster consisting of the small PSU and a large PSU is formed. The large PSU in the cluster would not be fixed, but instead would be selected by any procedure for which at least two of the large PSUs have a positive probability of being in the cluster with the small PSU. This insures that all large PSUs have a positive probability of not being in the cluster and hence of being in sample together with the small PSU.

3. KEEPING SMALL PSUs FROM BEING SELECTED FOR TOO MANY SURVEYS

Currently the PSUs for several of the key demographic survey are selected independently of each other. In this section an alternative approach is presented that limits the number of surveys for which a PSU can be selected to the maximum number for which the PSU can provide sample for an entire decade without reusing sample. This procedure would insure that a rotation procedure would not be required for any PSU large enough to provide a decade of sample for any single survey. For a PSU too small to provide a decade's sample for even a single survey, rotation would be required, but this procedure could still be used even with this type of PSU to guarantee that the PSU is not in sample for more than one survey. It could, therefore, remain in sample for the selected survey longer than if it was in sample for other surveys.

The set of surveys to which this procedure can be applied includes any combination of surveys with one PSU per stratum and two PSU per stratum designs. Let r be the number of surveys to which this procedure is to be applied. It is assumed for simplicity that the population of PSUs is the same for each survey. Let N be the number of PSUs in the population and r_i, $i=1,...,N$, the maximum number of surveys for which the i-th PSU can provide a decade's sample if this PSU is selected for the surveys with the largest sample requirements. If this maximum is 0, set $r_i=1$ instead of 0. Assume that $r_i<r$ for the first M PSUs only.

The selection of the sample PSUs for the r surveys becomes a two step process. First a subset R_i, $i=1,...,M$, of size r_i of the r surveys are selected independently for each of the small PSUs. The selection probabilities $\alpha_{ik}=P(k\varepsilon R_i)$, $i=1,...,M$, $k=1,...,r$, must satisfy the conditions $\alpha_{ik}>0$ and $\sum_{j=1}^{r} \alpha_{ij}=r_i$ for all i,k. During the second step of the selection process the i-th PSU can only be selected as a sample PSU for the k-th survey if $k\varepsilon R_i$.

Let S_k be a stratum from the k-th survey, consisting of u PSUs of which $r_i<r$ only for the first s PSUs. Let $R=R_1\times R_2\times...\times R_s$. In the second step of the selection process, sample PSUs are selected from S_k with selection probabilities conditioned on R. At this point, the cases for which the k-th survey is a one PSU per stratum design and for which it is a two PSUs per stratum design are considered separately.

One PSU per stratum design. Let p_i, $i=1,...,u$, be the unconditional selection probability for the i-th PSU. Then $p_{iR} = P(i \text{ is selected}|R)$ is defined as follows:

$$\begin{aligned} p_{iR} &= p_i/\alpha_{ik}, && i=1,...,s, \ k\varepsilon R_i, \\ &= 0, && i=1,...,s \ \ k\notin R_i, \\ &= g_R p_i, && i=s+1,...,u, \end{aligned}$$

where

$$g_R = \frac{1-\sum_{j=1}^{s} p_{jR}}{\sum_{j=s+1}^{u} p_j}.$$

Note that this definition satisfies the required relationship $E(p_{iR}) = p_i$, i=1,...,u. This is clearly true for i=1,...,s, while it also hold for i=s+1,...,u, since

$$E(g_R) = \frac{1-\sum_{j=1}^{s} p_j}{\sum_{j=s+1}^{u} p_j} = 1.$$

It remains only to show that $0 \le p_{iR} \le 1$ for i=1,...,u and all possible R. These inequalities will hold, provided for all R:

$$p_i/\alpha_{ik} \le 1, \qquad i=1,...,s, \tag{3.1}$$

$$g_R p_i \le 1, \qquad i=s+1,...,u \tag{3.2}$$

$$g_R \ge 0. \tag{3.3}$$

Now (3.2) holds, since for i=s+1,...,u, and all R,

$$g_R p_i \le \frac{p_i}{\sum_{j=s+1}^{u} p_j}.$$

Furthermore (3.3) implies (3.1), since $g_R \ge 0$ for all R if and only if

$$\sum_{i=1}^{s} p_i/\alpha_{ik} \le 1. \tag{3.4}$$

Thus it remains only to establish (3.4). Unfortunately, this inequality does not hold for every possible stratification. It would not generally hold, for example, if $s=u$, since then

$$\sum_{i=1}^{s} p_i/\alpha_{ik} = \sum_{i=1}^{u} p_i/\alpha_{ik} > \sum_{i=1}^{u} p_i = 1,$$

unless $\alpha_{ik}=1$, $i=1,...,s$. However, we will be particularly concerned only with obtaining a set of α_{ik}'s, which together with constraints on the size of s, satisfy the requirements of this procedure for the set of four major household surveys to be fully redesigned by the Census Bureau, namely the Current Population Survey (CPS), the National Crime Survey (NCS), the National Health Interview Survey (NHIS), and the Survey of Income and Program Participation (SIPP). Furthermore, two of these surveys, SIPP and NHIS, have a two PSUs per stratum design, and as will be shown, a more restrictive inequality, (3.11), must be met for such designs. A numerical illustration of how either (3.4) or (3.11) can be satisfied, as appropriate, for each of these four surveys simultaneously will be postponed until the two PSUs per stratum case is considered.

The remainder of this section is omitted due to lack of space. It is contained in the complete paper (Ernst 1991), available from the author.

4. MODIFICATIONS OF MAXIMIZATION OF OVERLAP DUE TO SMALL PSUs

A quick background on the solution to overlap maximization problems will first be presented. The problem of maximizing the expected number of PSUs retained in sample when redesigning a survey with a stratified redesign for which the PSUs are selected with probability proportional to size was first introduced to the literature by Keyfitz (1951). Causey, Cox and Ernst (1985) were able to obtain an optimal solution to this problem under very general conditions by formulating it as a transportation problem. Unlike previous approaches, this procedure imposes no restrictions on changes in strata definitions or number of PSUs per stratum. However, this procedure in practice is usable only when the initial sample of PSUs was selected independently from stratum to stratum, a condition met at the Census Bureau by the 1980s' NHIS and SIPP designs, but not by the CPS and NCS designs. An alternative linear programming procedure for use when this independence condition is not met was developed by Ernst (1986) and will be used in the 1990s' redesign, as it was in the previous redesign, to maximize overlap for CPS and NCS. For NHIS and SIPP a modification of the procedure of Causey, Cox and Ernst (1985) will be used. This procedure, presented in Ernst (1989), drastically reduces the size of the transportation problems that must be solved.

All previous known uses of overlap procedures have attempted to maximize overlap with respect to all PSUs in each new design stratum. In this section it is demonstrated how the objective function in the linear programming problem can be modified so that overlap is maximized for large PSUs, but some other goal is achieved for small PSUs.

The modifications will be presented with respect to the procedure in Ernst (1986). The modifications for other linear programming overlap procedures are analogous. The reader of this section is urged to read Ernst (1986) to facilitate understanding of the work to be presented, since only a sketchy description of that procedure will be provided here.

A one PSU per stratum design will be assumed. Modifications for other designs will be described later. The notation used in this section will mostly conform to the notation in Ernst (1986), rather than the notation of previous sections of this paper.

For any overlap procedure, each stratum S in the new design represents a separate problem. Let n denote the number of PSUs in S, of which the first v are assumed small. Let T_1,...,T_r denote the set of strata in the initial design that contains PSUs in S. Let u_i, i=1,...,r, denote the number of possibilities for the subset of $T_i \cap S$ consisting of all initial sample PSUs in $T_i \cap S$. Since a one PSU design is assumed, each singleton subset of $T_i \cap S$ is among these possible subsets. Unless $T_i \subset S$, there is one additional possibility, that the set of initial sample PSUs in $T_i \cap S$ is empty.

The overlap procedure in Ernst (1986) involves selection of a single initial stratum from among $T_1,...,T_r$ with probabilities that are determined by the optimization procedure, and then conditioning the selection probabilities for the new sample PSU on which of the u_i possibilities occurred for the initial sample PSUs in $T_i \cap S$, $i=1,...,r$.

The objective function for this problem, when the goal is to maximize overlap for all PSUs, both large and small, is

$$\sum_{i=1}^{r}\sum_{j=1}^{u_i}\sum_{k=1}^{n} c_{ijk}x_{ijk}. \tag{4.1}$$

The variable x_{ijk} in (4.1) is the joint probability that T_i is the selected initial stratum, the j-th outcome occurred for the initial sample PSUs in $T_i \cap S$, and the k-th PSU in S is selected as the new sample PSU. The cost coefficient c_{ijk} is the conditional probability that the k-th PSU in S was an initial sample PSU given that j-th outcome occurred for the initial sample PSUs in $T_i \cap S$. (4.1) is therefore the probability that the new sample PSU was in the initial sample, and maximization of (4.1) maximizes expected overlap for S.

In order to maximize overlap for large PSUs while treating small PSUs in a neutral fashion with respect to maximization and minimization of overlap, simply modify (4.1) by replacing c_{ijk} in (4.1) by c'_{ijk}, where for all i,j,k,

$$c'_{ijk}=0,\ k=1,...,v, \qquad c'_{ijk}=c_{ijk},\ k=v+1,...,n. \tag{4.2}$$

The objective function with these cost coefficients is the probability that a large PSU in S is selected as a new sample PSU and was an initial sample PSU, so that maximization of this objective function will maximize overlap for the set of large PSUs while not attempting to control overlap of small PSUs. It is cost coefficients of this form that the Census Bureau plans to use in overlapping all the surveys in the 1990s' redesign.

If the goals are to maximize overlap of large PSUs and minimize overlap of small PSUs then (4.1) can be modified by replacing c_{ijk} by c''_{ijk}, where for all i,j,

$$c''_{ijk}=\beta(1-c_{ijk}),\quad k=1,...,v, \qquad c''_{ijk}=c_{ijk},\quad k=v+1,...,n, \tag{4.3}$$

with β a positive constant. The objective function would then be the expected overlap of large PSUs plus β times the expected nonoverlap of small PSUs. If this objective function is maximized, then the procedure would attempt to attain the twin goals of maximizing the overlap of large PSUs and minimizing the overlap of small PSUs. The constant β can help to handle any conflict between these two goals. For example, if maximization of overlap of large PSUs is the primary objective, with minimization of overlap of small PSUs a secondary objective, then a small value of β would be appropriate. This is because

$$\sum_{i=1}^{r}\sum_{j=1}^{u_i}\sum_{k=1}^{n} c''_{ijk}x_{ijk}\le\beta,$$

and therefore an optimal solution with the cost coefficients (4.3) must result in an overlap for the set of large PSUs that is within β of the maximum overlap of large PSUs obtained with the cost coefficients (4.2).

The computation of the cost coefficients c_{ijk} for designs of more than one PSU per stratum is described in Ernst (1986, p. 195). For such designs, the k in c_{ijk} corresponds to a set of new sample PSUs instead of a single PSU. For example, for a two PSUs per stratum design, k corresponds to a pair of PSUs. For designs of more than one PSU per stratum, instead of modifying c_{ijk} directly, the contribution to c_{ijk} from each PSU in the set of PSUs corresponding to k is separately modified analogously to (4.2) and (4.3), and the resulting modified contributions summed to obtain a modification of c_{ijk}.

To ascertain the effect of the cost coefficients options (4.1)-(4.3) on an actual survey, the overlap procedure was run on each of the 55 strata for the 1980s' CPS design which contained small PSUs. These strata were treated as strata from the "new" design. The "initial" design was obtained from a simulated 1970s' stratification. (The actual 1970s' design consisted of more than one sample of PSUs. Working with it would have required a more complex overlap procedure.) For each of the 55 strata, the overlap procedure was run four times, with cost coefficients (4.1), (4.2), (4.3) for $\beta=1/2$ and (4.3) for $\beta=1$. The expected overlap for large PSUs and small PSUs are presented in Table 1 for each of the four sets of cost coefficients and also for selection of the new sample PSUs independently of the initial sample of PSUs. Each entry in the first numerical column of this table is the arithmetic mean over the 55 strata of the probability that a large PSU is selected in the new sample and was also in the initial sample. For Table 2, each entry in the first column is obtained by first dividing the expected overlap for large PSUs for each stratum by the probability that a large PSU is selected in the new design for that stratum, and then taking the arithmetic mean of the quotient over the 55 strata. The result is the expected proportion of large new sample PSUs in these strata that were in the initial sample. Column 2 of Tables 1 and 2 are obtained analogously for small PSUs.

From either table it can be seen that the ordering of the amount of overlap for both large and small PSUs for these four sets of coefficients is as would be anticipated. For large PSUs the overlap is highest for cost coefficients (4.2) and lowest for (4.1). For small PSUs it highest for (4.1) and lowest for (4.3) with $\beta=1$. While the four overlap procedures differ little in the amount of overlap for large PSUs, the proportion of small new sample PSUs that were in the initial sample is much higher for (4.1) than for the other three sets of cost coefficients.

Separate tables of overlap results for each of the 55 strata were obtained, although they are not presented here. However, it is interesting to note that for all but 16 of these strata, the overlap of large PSUs was the same for all four sets of cost coefficients. Among these 16 strata, for 11 the large PSU overlap for (4.1) was lower than for the other three sets of cost coefficients, while these three sets yielded the same overlap for large PSUs. For three strata, (4.1) resulted in the smallest overlap of large PSUs, (4.2) and (4.3) with $\beta=1/2$ had the highest overlap of large PSUs, with (4.3) with $\beta=1$ in between. For one stratum, (4.1) had the smallest overlap of large PSUs, (4.2) had the highest, with (4.3) for both $\beta=1/2$ and $\beta=1$ equal at an intermediate value. Finally, for one stratum, (4.3) with $\beta=1$ had the smallest overlap of large PSUs, (4.2) and (4.3) with $\beta=1/2$ had the highest, with (4.1) at an intermediate value. For the small PSUs, the overlap for each of the 55 strata for (4.3) with $\beta=1$ was less than or equal to the overlap for (4.3) with $\beta=1/2$, which in turn was less than or equal to (4.2), which was less than or equal to (4.1).

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

REFERENCES

Alexander, C.H., Ernst, L.R., and Haas M.E. (1982), "A System for Replacing Primary Sampling Units When the Units Have Been Exhausted, Proceedings of the Section on Survey Research Methods, American Statistical Association, 211-216.

Causey, B.D., Cox, L.H., and Ernst, L.R. (1985), "Applications of Transportation Theory to Statistical Problems," Journal of the American Statistical Association, 80, 903-909.

Cochran, W.G. (1977), Sampling Techniques, Third Edition, New York: John Wiley and Sons.

Ernst, L.R. (1991), "Sample Redesign Issues for Rotating Primary Sampling Units," Report CENSUS/SRD/RR-91/06, U.S. Bureau of the Census, Statistical Research Division.

Ernst, L.R. (1986), "Maximizing the Overlap Between Surveys When Information Is Incomplete,"European Journal of Operational Research, 27, 192-200.

Roebuck, M.J. (1991), "1990 Redesign: PSU Selection Specifications," Doc #24-R-7, Working Draft 2, memorandum to B.M. Walter, U.S. Bureau of the Census.

ACKNOWLEDGEMENT

The author would like to thank Todd Williams who did the programming that produced the tables in Section 4.

Table 1. Expected Number Per Stratum of Large and Small PSUs in New Sample and Also in Initial Sample

Selection Procedure	Large	Small
Overlap of Large and Small PSUs Maximized (4.1)	0.5076	0.0572
Neutral Treatment of Small PSUs (4.2)	0.5158	0.0060
Overlap of Small PSUs Minimized (4.3) with $\beta=1/2$	0.5158	0.0037
Overlap of Small PSUs Minimized (4.3) with $\beta=1$	0.5152	0.0029
Independent Selection of New PSUs	0.2020	0.0064

Table 2. Expected Proportion of Large and Small New Sample PSUs in Initial Sample

Selection Procedure	Large	Small
Overlap of Large and Small PSUs Maximized (4.1)	0.5684	0.4735
Neutral Treatment of Small PSUs (4.2)	0.5781	0.0635
Overlap of Small PSUs Minimized (4.3) with $\beta=1/2$	0.5780	0.0326
Overlap of Small PSUs Minimized (4.3) with $\beta=1$	0.5773	0.0279
Independent Selection of New PSUs	0.2218	0.0490