BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-84/31

MAXIMIZING THE OVERLAP BETWEEN SURVEYS WHEN INFORMATION IS INCOMPLETE

by

Lawrence R. Ernst
Statistical Research Division
Bureau of the Census
Room 3524, F.O.B. #3
Washington, D.C. 20233 U.S.A.

This series contains research reports, written by or in cooperation with staff members of the Statistical Research Division, whose content may be of interest to the general statistical research community. The views reflected in these reports are not necessarily those of the Census Bureau nor do they necessarily represent Census Bureau statistical policy or practice. Inquiries may be addressed to the author(s) or the SRD Report Series Coordinator, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233.

Recommended by: Paul P. Biemer

Report completed: November 29, 1984

Report issued: December 18, 1984

11/29/84

MAXIMIZING THE OVERLAP BETWEEN SURVEYS WHEN INFORMATION IS INCOMPLETE

by

Lawrence R. Ernst
Bureau of the Census
U.S. Department of Commerce
Washington, D.C. 20233, U.S.A.

ABSTRACT

When redesigning a survey with a multi-stage design, it is sometimes desired to maximize the number of first stage units retained in the new sample without altering unconditional selection probabilities. Using transportation theory, an optimal solution to this problem for a very general class of designs was recently presented by Causey, Cox and Ernst. However, that procedure has not yet been used in the redesign of any survey because it requires the knowledge of certain joint probabilities which are often not known in practice. In this paper an alternative linear programming procedure is presented which requires only probability information that should always be available, and which, under certain conditions, is optimum among all procedures requiring only this information. This procedure has recently been used in the redesign of two major surveys conducted by the U.S. Bureau of the Census, the Current Population Survey, and the National Crime Survey.

Abbreviated Title: Maximizing the Overlap Between Surveys

Keywords: Programming: linear; Statistics: sampling

AUTHOR'S FOOTNOTE

Lawrence R. Ernst is Mathematical Statistician, Statistical Research Division, Bureau of the Census, Washington, D.C. 20233, U.S.A.

1. INTRODUCTION

In national household surveys employing personal interviewing the sample selection process is typically a multi-stage process in which the first-stage sample units are contiguous geographic areas. These geographic areas are known as primary sampling units (PSU's) and commonly are of a size appropriate to be covered by a single interviewer. The selected PSU's are subsampled for one or more further stages until, at the final stage, the sample units are the desired analytic units, such as households or persons.

The sample PSU's are generally selected as follows. The set of all PSU's are partitioned into subsets known as strata. The partitioning places together PSU's with similar characteristics in order to increase the precision of the estimates. From each stratum a fixed number of sample PSU's are chosen. In most surveys the same number of PSU's are chosen from each stratum, with one or two often being that number. In order to further increase the precision of the estimates, these units are usually not chosen with equal probability but instead proportional to some measure of size associated with each PSU, such as the population of the PSU in the most recent census. Furthermore, in the case of m per stratum without replacement designs with $m>2$, not only is the individual probability of selection for each PSU in a stratum predetermined by the sampling procedure employed, but also the joint selection probability for any set of m PSU's in the stratum. The reader is referred to [3] and [9] for further information on multi-stage sampling and sampling proportional to size, and to [1] and [10] for examples of m PSU's per stratum without replacement selection procedures.

If the survey is periodic, at some point more current data may become available, such as data from a new census, which could be used to obtain an improved partitioning of the PSU's into strata and better selection probabilities. A redesign would than take place in which a new set of sample PSU's is chosen using the new stratification and selection probabilities. For example, such a redesign has been conducted at approximately ten year intervals for the household surveys conducted by the U.S. Bureau of the Census, using data obtained from the most recent U.S. Census of Population and Housing.

The new set of sample PSU's may of course be selected independently of the initial sample PSU's. However, additional costs, such as the expense of training a new interviewer, are generally incurred with each change of sample PSU. Consequently, it may be considered desirable to maximize the expected number of sample PSU's retained in the new design, while strictly maintaining the requirements of probabililty sampling. That is, such a procedure would not alter the unconditional selection probabilities for any set of PSU's in a new stratum, but would condition each such probability on the set of initial sample PSU's in such a manner that the conditional probability of a new PSU being selected would in general be greater than its unconditional probability when the PSU was in the initial sample and less otherwise.

Keyfitz [6] presented an optimum procedure for one PSU per stratum designs in the special case when the initial and new strata are identical, with only the selection probabilities changing. For the more general one PSU per stratum problem for which the strata definitions can change in the redesign, Perkins [8], and Kish and Scott [7] presented procedures that are not optimum. Fellegi [4] considered a particular two PSU's per stratum design, but his procedure is also not optimum. None of these authors used the techniques of mathematical programming.

Causey, Cox and Ernst [2] obtained an optimal solution to the overlap problem by formulating it as a transportation problem. Their procedure is very general with no restrictions on changes in strata definitions or number of PSU's per stratum. (Raj [9] had previously employed the transportation problem approach, but only with the restrictive assumptions considered by Keyfitz.) However, the procedure of [2] requires that the joint selection probabilities is the initial sample be known for any set of PSU's that are in the same stratum in the new design. If the initial sample was not chosen independently from stratum to stratum, this information may not be available. This is explained further in the next section. Such a situation existed for the recently completed redesign of the household surveys conducted by the U.S. Bureau of the Census, and consequently, the procedure of [2] could not be used in the selection of new sample PSU's for these surveys.

In this paper an alternative overlap procedure is presented which only requires knowledge of the joint selection probabilities in the initial sample for sets of PSU's that are in the same initial and new strata, and which, in certain circumstances, is optimum among all procedures which require only this amount of information. This alternative procedure formulates the overlap problem as a linear programming problem, but not a transportation problem. Not only was this procedure usable for the recently completed redesign of the U.S. Bureau of the Census's household surveys, but it was indeed used for selecting PSU's for the only two surveys that employed an overlap procedure, the Current Population Survey (CPS) and the National Crime Survey (NCS).

In Section 2 the need for an alternative to the procedure of [2] is explained further. The new procedure is presented in Section 3 and illustrated by a simple example in Section 4. Finally, in Section 5 we discuss the application of this procedure to the redesign of the CPS and the NCS, which includes a description of how the objective function can be modified in situations in which not only the strata definitions and selection probabilities change in the new design, but also the PSU definitions.

2. NEED FOR AN ALTERNATIVE OVERLAP PROCEDURE

It will be assumed throughout this paper that the initial design is $m>1$ PSU's per stratum without replacement and that the new design is $m'>1$ PSU's per stratum without replacement. The overlap procedure of [2] and the overlap procedure to be presented in Section 3 can both readily be modified to apply to other designs. For example, if the new design is m' PSU's with replacement then the selection of sample PSU's in each new stratum can be treated as m' identical one PSU per stratum problems.

The optimal overlap procedure of [2] requires that the joint selection probabilities in the initial sample be known for any set of PSU's in the same stratum in the new design. If the initial sample was chosen independently from stratum to stratum, then these joint probabilities are easy to compute. Simply partition the set of PSU's in question into subsets determined by the initial stratum in which each PSU was placed. The joint selection probability in the initial sample for each such subset should always be known, since the probability for subsets consisting of exactly m PSU's is predetermined, from which probabilities for subsets of fewer PSU's can be obtained. The joint probability for the entire set would then be the product of the joint probabilities for each subset.

Unfortunately, there are several sampling techniques which when used in the PSU selection process in general destroy this independence. One such technique is controlled section, which is described in [5] and [2]. More interesting, perhaps, is the fact that all overlap procedures destroy this independence in the following manner. Suppose that the initial set of sample PSU's was chosen independently from stratum to stratum and that a new set of sample PSU's is chosen using an overlap procedure. The new set of sample PSU's, as will be illustrated below, would not have been chosen independently from stratum to stratum. Consequently, if a subsequent redesign were to take place, what were the new sample PSU's would then become the initial sample PSU's which would not have been selected with the desired independence.

To illustrate the fact that the new sample PSU's would not be selected independently from stratum to stratum when they are chosen with an overlap procedure, consider the following situation. Two PSU's, which we denote by PSU 1 and PSU 2, were in the same initial stratum but are in different new strata. Furthermore, the initial design was one PSU per stratum and PSU 1 was in the initial sample. Then, in general, in that situation PSU 1 would have a conditional selection probability in the new sample greater than its unconditional selection probability, while the opposite would be true for PSU 2 since it could not have been in the intial sample. Thus, these two PSU's would not be selected independently in the new sample.

In the case of the household surveys conducted by the U.S. Bureau of the Census, both controlled selection and overlap procedures had been used in the last two redesigns prior to the current one, and hence the independence assumption did not hold for the current redesign. Although theoretically the joint probabilities required to use the procedure of [2] could still have been computed, it would be a laborious task in general, and an impossible task for these surveys, since some of the data needed to perform the computations were no longer available. Consequently, it was necessary to use an alternative overlap procedure, which will be described in the next section.

3. THE PROCEDURE

Note first that for any overlap procedure, each stratum S in the new design represents a separate problem. Let $T_1, \ldots, T_r$ denote the initial strata that contains PSU's in S. For each i, unconditional selection probabilities in the initial sample are known for every subset of T_i. The problem noted in the last section can only arise when such probabilities are also needed for sets of PSU's that were in more than one intital stratum. It is this observation that motivates the procedure to be described. The general idea is to select one of the T_i and then condition the selection of the set of new sample PSU's in S on the set of initial sample PSU's that are in $T_i \cap S$. Furthermore, in general, one specific T_i is not chosen with certainty, but instead probabilities $y_i, \ldots, y_r$ are assigned to $T_1, \ldots, T_r$ respectively. The chosen T_i is designated by T. The y_i's are variables in the optimization process. Thus, the selection of a set of new sample PSU's in S is a three stage process. First T is chosen, then the set of intitial sample PSU's that are in $T \cap S$ is noted, and finally the set of new sample PSU's in S is chosen conditioned on the outcome of the first two steps of the process.

To proceed further some more notation is required. For i=1, ..., r let I_{ij}, $j=1, \ldots, u_i$, denote the possible outcomes for the set of PSU's in $T_i \cap S$ that were initial sample PSU's; designate by I_i the actual outcome; and denote by p_{ij} the probabililty that $I_i = I_{ij}$. Similarly let $N_1, \ldots, N_n$ denote all possible outcomes for the set of new sample PSU's in S, designate by N the actual outcome, and denote by π_k the probability that $N = N_k$. Note that each I_{ij} contains no more than m elements, that each N_k contains exactly m' elements, and that the p_{ij}'s and π_k's are known values.

Returning to the three stage event described above, let x_{ijk}, $i=1, \ldots, r$, $j=1, \ldots, u_i$, $k=1, \ldots, n$, denote the joint probability that the initial stratum T_i is chosen, that I_{ij} is the set of initial sample PSU's in $T_i \cap S$ and that N_k is the set of new sample PSU's in S. That is, with the notation just described,

$$x_{ijk} = P(T=T_i,\ I_i=I_{ij},\ N=N_k).$$

The x_{ijk}'s will be the only other variables besides the y_i's in the linear programming problem to be described. After an optimal set of values is obtained by solving the linear programming problem, the desired probabilities, which are the probabilities of selection of each of the N_k's conditioned on the entire set of initial sample PSU's in S, can be expressed in terms of the optimal x_{ijk}'s and the known p_{ij}'s as follows. Because T is selected independently of the initial sample we have that for each i, j, k,

$$P(N=N_k|T=T_i,\ I_i=I_{ij}) = \frac{P(T=T_i,\ I_i=I_{ij},\ N=N_k)}{P(T=T_i,\ I_i=I_{ij})} = \frac{x_{ijk}}{y_i p_{ij}},$$

from which it follows that if $1< j_i <u_i$ for $i=1, \ldots, r$, then

$$P(N=N_k|I_1=I_{1j_1},\ \ldots,\ I_r=I_{rj_r})$$
$$= \sum_{i=1}^{r} y_i\ P(N=N_k|T=T_i,\ I_i=I_{ij_i}) = \sum_{i=1}^{r} \frac{x_{ij_ik}}{p_{ij_i}} \qquad (3.1)$$

All that now remains is to state the linear programming problem that yields the optimal x_{ijk}'s. The constraints will be presented first and then the objective function. Since new unconditional selection probabilities must be preserved, the probabilities of all the three stage events with N_k as the set of new sample PSU's in S sum to π_k, that is

$$\sum_{i=1}^{r} \sum_{j=1}^{u_i} x_{ijk} = \pi_k,\quad k=1, \ldots, n. \qquad (3.2)$$

Similarly, since $P(T=T_i,\ I_i=I_{ij}) = p_{ij}\ y_i$, we also have

$$\sum_{k=1}^{n} x_{ijk} = p_{ij}\ y_i, \quad i=1,\ \ldots,\ r,\ j=1,\ \ldots,\ u_i. \tag{3.3}$$

The final constraint is

$$\sum_{i=1}^{r} y_i = 1, \tag{3.4}$$

which arises from the fact that exactly one initial stratum is chosen.

As for the objective function, if for each i, j, k, a constant c_{ijk} could be determined which would be the conditional expected number of sample PSU's in N_k that were in the initial sample given that $T = T_i$ and $I_i = I_{ij}$, then

$$\sum_{i=1}^{r} \sum_{j=1}^{u_i} \sum_{k=1}^{n} c_{ijk}\ x_{ijk} \tag{3.5}$$

would be the unconditional expected number of PSU's in S that are in both the initial and new samples, which is what we seek to maximize.

Hence the linear programming problem is to maximize (3.5) subject to (3.2), (3.3) and (3.4), and the only remaining task is to specify the c_{ijk}'s used in (3.5). To do this, we first for k=1, ..., n, let N_{kh}, h=1, ..., m´, denote the PSU's in N_k; let p'_{kh} denote the unconditional probability that N_{kh} was in the initial sample; and let p''_{ijkh}, i=1, ..., r, j=1, ..., u_i, denote the conditional probability that N_{kh} was in the initial sample given that $T=T_i$ and $I_i=I_{ij}$.

Then clearly

$$c_{ijk} = \sum_{h=1}^{m'} p''_{ijkh}.$$

Furthermore, the following known values are to be used for p''_{ijkh}:

$$p''_{ijkh} = \begin{array}{ll} 1 \text{ if } N_{kh} \in I_{ij}, \\ 0 \text{ if } N_{kh} \in T_i{\sim}I_{ij}, \\ p'_{kh} \quad \text{otherwise.} \end{array} \tag{3.6}$$

The values given for p''_{ijkh} in the first two cases above are obviously the correct ones. As for the third case, if the initial sample had been selected independently from stratum to stratum we would indeed have $p''_{ijkh} = p'_{kh}$, that is the conditional probability of selection would simply be the unconditional probability. Although we are specifically not making this independence assumption, and hence $p''_{ijkh} = p'_{kh}$ is not necessarily true, it is used anyway in this case. This is because if $N_{kh} \not\in T_i$ then we lack a better estimate of p''_{ijkh}, since knowing that $I_i = I_{ij}$ does not in general provide any information concerning whether N_{kh} was in the initial sample. In addition, it is believed that the use of p'_{kh} instead of the unknown true value of p''_{ijkh} would not alter the expected overlap greatly. Furthermore, because (3.2) holds irrespective of the values used for the p''_{ijkh}'s, this procedure always preserves the new unconditional probabilities of selection, that is the p''_{ijkh}'s affect only the objective function not the constraints. This contrasts with the method of [2], where joint probabilities that are not necessarily known are needed in the constraints, and if incorrect probabilities are used then the requirement that the new unconditional selection probabilities be preserved would not be met.

Thus, in summary, the overlap procedure presented in this section always preserves the new unconditional probabilities of selection without any knowledge of joint selection probabilities in the initial sample for any set of PSU's not contained in a single initial stratum, and is optimum among all such overlap procedures provided the initial sample had been selected independently from stratum to stratum.

4. AN EXAMPLE

In this example, which illustrates the method presented in the previous section, both the initial and new designs are one PSU per stratum. S consists of five PSU's, designated S_1, ..., S_5, with new selection probabilities .40, .15, .05, .30, .10. Then n=5, $N_k=\{S_k\}$, k=1, ..., 5, and the given probabilities are the values of $\pi_1, \ldots, \pi_5$ respectively.

We are further given that the initial selection probabilities for the five PSU's were .50, .06, .04, .60, .10, and that S_1, S_2, S_3 are in one initial stratum and S_4, S_5 in a second initial stratum. Consequently r=2, $S\cap T_1 = \{S_1, S_2, S_3\}$, $S\cap T_2 = \{S_4, S_5\}$, $u_1 = 4$, $u_2 = 3$, and the I_{ij}'s are: $I_{11} = \{S_1\}$, $I_{12} = \{S_2\}$, $I_{13} = \{S_3\}$, $I_{14} = \emptyset$, $I_{21} = \{S_4\}$, $I_{22} = \{S_5\}$, $I_{23} = \emptyset$. Furthermore, from the given initial selection probabilities we have $p_{11} = .5$, $p_{12} = .06$, $p_{13} = .04$, $p_{21} = .60$, $p_{22} = .10$, while $p_{14} = 1 - p_{11} - p_{12} - p_{13} = .4$ and $p_{23} = 1 - p_{21} - p_{22} = .3$. In addition, since $m'=1$ it follows that $c_{ijk} = p''_{ijk1}$ for all i, j, k and that $N_{k1} = S_k$. Consequently, the c_{ijk}'s are as given in Table 1.

Table 1. c_{ijk}'s

(i,j) \ k	1	2	3	4	5
(1,1)	1	0	0	.6	.1
(1,2)	0	1	0	.6	.1
(1,3)	0	0	1	.6	.1
(1,4)	0	0	0	.6	.1
(2,1)	.5	.06	.04	1	0
(2,2)	.5	.06	.04	0	1
(2,3)	.5	.06	.04	0	0

Upon maximizing (3.5) subject to (3.2), (3.3) and (3.4) with the p_{ij}'s, π_k's, and c_{ijk}'s as above, an optimal set of x_{ijk}'s is obtained, which is presented in Table 2.

Table 2. x_{ijk}'s which maximize (3.5)

(i,j) \ k	1	2	3	4	5
(1,1)	.400	.000	.000	.000	.000
(1,2)	.000	.048	.000	.000	.000
(1,3)	.000	.000	.032	.000	.000
(1,4)	.000	.042	.018	.180	.080
(2,1)	.000	.000	.000	.120	.000
(2,2)	.000	.000	.000	.000	.020
(2,3)	.000	.060	.000	.000	.000

The corresponding y_i's are y_1 = .800 and y_2 = .200, and the maximum value of the objective function is .740. This compares with an overlap probability of .401 if the new sample had been selected independently of the initial sample, and an overlap proability of .880 if the initial sample had been selected independently from stratum to stratum and the optimal method of [2] had been used.

Finally from (3.1), Table 2 and the p_{ij}'s above, an optimal set of conditional probabilities are obtained, as given in Table 3.

Table 3. $P(N=N_k \mid I_1 = I_{1j_1}, I_2 = I_{2j_2})$

				k		
j_1	j_2	1	2	3	4	5
1	1	.800	.000	.000	.200	.000
1	2	.800	.000	.000	.000	.200
1	3	.800	.200	.000	.000	.000
2	1	.000	.800	.000	.200	.000
2	2	.000	.800	.000	.000	.200
2	3	.000	1.000	.000	.000	.000
3	1	.000	.000	.800	.200	.000
3	2	.000	.000	.800	.000	.200
3	3	.000	.200	.800	.000	.000
4	1	.000	.105	.045	.650	.200
4	2	.000	.105	.045	.450	.400
4	3	.000	.305	.045	.450	.200

5. APPLICATIONS

The procedure described in Section 3 was used twice in the recent redesign of the household surveys conducted by the U.S. Bureau of the Census. It was first used to maximize the number of sample PSU's common to the old and new designs of the Current Population Survey (CPS), which is a monthly survey that provides estimates of labor force characteristics, including the U.S. unemployment rate. The procedure was subsequently used in the redesign cf the National Crime Survey (NCS), a survey that produces estimates of annual crime victimizations. However, the set of new sample PSU'S for NCS was not chosen to maximize the overlap with the old NCS sample PSU's, but instead to maximize the overlap with the new CPS sample PSU's, which had been selected first. This was done primarily because of the belief that greater operational efficiencies could be achieved if the new NCS and new CPS samples had as many PSU's in common as possible, since the interviewing for both surveys in any such PSU can generally be performed by the same interviewer.

In overlapping the new NCS design with the new CPS design, a modification of the procedure presented in Section 3 was necessary, which will be detailed here because of its potential general applicability. The new CPS and new NCS are both one PSU per stratum designs. However, not only are the set of strata different for the two surveys, but some of the NCS PSU's are not identical to any CPS PSU, that is the set of NCS PSU's and the set of CPS PSU's constitute two somewhat different geographical partitionings of the United States. Consequently, the objective of maximizing the set of sample PSU's in common to the two designs no longer has a precise meaning. The description of the resulting neccessary modification of the overlap procedure will be presented in generality. That is, we will refer to the initial and the new sample instead of the new CPS sample and the new NCS sample respectively, and the presentation will not be restricted to one PSU per stratum designs. (In the one PSU per stratum case, the changes to be described are due to Alexander and Roebuck [11].)

When the set of PSU's in the two designs are not identical, our modified goal is to maximize the expected number of PSU's in the new sample in which at least one initial sample interviewer resides, since any such new sample PSU would not requiring hiring a new interviewer. This, in turn, requires that the T_i's, I_{ij}'s and p''_{ijkh}'s have a new, more general meaning. $T_1, \ldots, T_r$ are now the initial strata that have at least one PSU that intersects a PSU in S; I_{ij}, $j=1, \ldots, u_i$, denotes all the possible outcomes for the set of PSU's in T_i that intersect PSU's in S and were initial sample PSU's; and p''_{ijkh} is the conditional probability that the interviewer from at least one initial sample PSU resides in N_{kh} given $T = T_i$ and $I_i = I_{ij}$. With this more general meaning for p''_{ijkh} , c_{ijk} is now the conditional expected number of PSU's in N_k in which at least one initial interviewer resides given $T = T_i$ and $I_i = I_{ij}$, and (3.5) remains the desired objective function, which is now the expected number of the new sample PSU's in S in which at least one initial sample interviewer resides.

It remains to explain how p''_{ijkh} is to be computed with its new meaning. First, let I_{ijt}, $t=1, \ldots, v_{ij}$, denote the PSU's in I_{ij} and let f_{ijtkh} denote the proportion of I_{ijt} that is in N_{kh} based on the new measure of size. Now, any information that may be known concerning where the initial sample interviewers actually reside cannot be used in computing the p''_{ijkh}'s, since the objective function must be completely independent of the initial sample. Instead, we take the conditional probability that an interviewer resides in $I_{ijt} \cap N_{kh}$ given that I_{ijt} was in the initial sample to be f_{ijtkh}, and then let

$$p''_{ijkh} = 1-[\prod_{t=1}^{v_{ij}} (1-f_{ijtkh})] \prod_{\substack{q=1\\ q\neq i}}^{r} (1-\sum_{w=1}^{u_q} p'_{qw}[1-\prod_{t=1}^{v_{qw}}(1-f_{qwtkh})]). \qquad (5.1)$$

To understand (5.1), first note that $\prod_{t=1}^{v_{ij}}(1-f_{ijtkh})$ is the conditional probability given $T=T_i$ and $I_i=I_{ij}$ that no initial sample interviewer from a PSU in T_i resides in N_{kh}. For $q\neq i$, $p'_{qw}\, [1-\prod_{t=1}^{v_{qw}}(1-f_{qwtkh})]$ is the unconditional probability that $I_q=I_{qw}$ and at least one initial sample interviewer in I_{qw} resides in N_{kh}, and then

$$1-\sum_{w=1}^{u_q} p'_{qw}[1-\prod_{t=1}^{v_{qw}}(1-f_{qwtkh})]$$

is the unconditional probability that no initial sample interviewer in any PSU in T_q resides in N_{kh}. It then follows, upon combining these observations, that (5.1) is the desired expression for p''_{ijkh} provided the initial sample PSU's had been selected independently from stratum to stratum. (The use of this independence assumption in obtaining 5.1 is analogous to its use in Section 3 to obtain (3.6).)

The results of using the overlap procedure in selecting the new CPS and NCS samples are as follows. For the new CPS sample the proportion of overlap with the old CPS sample is .56. This compares with an expected overlap proportion of .39 if the new CPS sample had been selected independently of the old CPS sample. (Some PSU's, because of their large size, were selected in the new CPS sample with certainty and are omitted from these calculations.) The corresponding proportions for the overlap of the new NCS sample with the new CPS sample are .81 and .59 respectively.

REFERENCES

[1] Bayless, D.L. and Rao, J.N.K., "An Empirical Study of Stabilities of Estimators and Variance Estimators in Unequal Probability Sampling (n=3 or 4)," Journal of the American Statistical Association, 65 (1970), 1645-1667.

[2] Causey, B.D., Cox, L. H. and Ernst L.R., "Applications of Transportation Theory to Statistical Problems," American Statistical Association - Proceedings of the Section on Survey Research Methods, 1983, 112-117.

[3] Cochran, William G., Sampling Techniques, 3rd ed., New York: John Wiley and Sons, 1977.

[4] Fellegi, Ivan P., "Changing the Probabililties of Selection when Two Units Are Selected with PPS Without Replacement," American Statistical Association - Proceedings of the Social Statistics Section, 1966, 434-442.

[5] Goodman, Rao and Kish, Leslie, "Controlled Selection - A Technique in Probability Sampling," Journal of the American Statistical Association, 45 (1950), 350-372.

[6] Keyfitz, Nathan, "Sampling with Probabilities Proportionate to Size: Adjustment for Changes in Probabilities," Journal of the American Statistical Association, 46 (1954), 105-109.

[7] Kish, Leslie and Scott, Alastair, "Retaining Units After Changing Strata and Probabilities," Journal of the American Statistical Association, 66 (1971), 461-470.

[8] Perkins, Walter M., "1970 CPS Redesign: Proposed Method for Deriving Sample PSU Selection Probabilities Within 1970 NSR Strata," U.S. Bureau of the Census memorandum to Joseph Waksberg, August 5, 1970.

[9] Raj, Des, Sampling Theory, New York: McGraw Hill, 1968.

[10] Rao, J.N.K., and Bayless D.L., "An Empirical Study of the Stabilities of Estimators and Variance Estimators in Unequal Probability Sampling of Two Units Per Stratum," Journal of the American Statistical Association, 64, (1969), 540-559.

[11] Shapiro, Gary M., "NCS/GPS Redesign: Speciciations for PSU Selection Using Overlap with 1980 CPS Design," U.S. Bureau of the Census Memorandum to Robert T. O'Reagan, May 12, 1983. Prepared by Charles H. Alexander and Michael J. Roebuck.