MINIMIZING OVERLAP IN NCES SURVEYS

Sadeq Chowdhury, Adam Chu, Westat; Steve Kaufman, National Center for Education Statistics
Sadeq Chowdhury, Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Minimizing overlap, resampling procedures, respondent load, school surveys

1. Introduction

The National Center for Education Statistics (NCES) conducts numerous school-based surveys covering a variety of topics. Often two or more of these surveys are fielded in roughly the same time period. To avoid undue reporting burden on individual schools, it is usually desirable to avoid or minimize the likelihood of selecting the same school for more than one survey.

Various methods have been used to minimize overlap among two or more concurrent surveys. For example, Rust and Johnson (1992) describe a relatively straightforward approach that was used to minimize the sample overlap between the national and state NAEP (National Assessment of Educational Progress) school samples. Ohlsson (1995) describes a general approach using permanent random numbers (PRN) that has been adopted by some national statistical agencies to minimize overlap among their active establishment surveys. Perry, Burt, and Ewig (1993) describe a method for simultaneous selection of samples for more than two surveys based on integer linear programming. Papers by Keyfitz (1951), Kish and Scott (1971), Causey, Cox, and Ernst (1985), and Brick, Morganstein, and Wolters (1987) present a number of alternative methods designed to control (i.e., either minimize or maximize) sample overlap. Ernst (1999) provides an excellent comparative review of these and other available methods for overlap control.

In this paper, we present a general and relatively simple approach for minimizing overlap among completing sample surveys. The approach is an extension of procedures developed and used by NCES to control the sample overlap between the Schools and Staffing Survey (SASS), the National Assessment of Educational Progress (NAEP), and other NCES surveys (Kaufman, 1993). Unlike many of the methods currently available, the method can easily be extended to the case of three or more competing surveys. In the following sections, we present the details for minimizing overlap among up to four surveys, and provide an example involving four NCES school surveys.

2. Statement of Problem and Notation

It is assumed that there is a "current" survey for which a sample is to be drawn with specified probabilities. However, samples from the same frame have already been selected for one or more "previous" surveys, where the term "previous" is used to describe the point in time in which the sampling occurred, not necessarily the period of data collection. For example, even though two or more surveys may be conducted concurrently, the selection of the respective samples is typically done sequentially. In this case, the "previous" survey is the one for which the sample was selected first. The goal is to select the "current" sample in a way that preserves the desired selection probabilities while avoiding as many of the previously selected units as possible.

In general, let s_j and $\bar{s}_j$ represent the set of sampled units that are selected and not selected for the j-th survey, respectively. Similarly, let $P_i(s_j)$ denote the probability of selecting the i-th unit for the j-th survey, and let $P_i(\bar{s}_j) = 1 - P_i(s_j)$. The joint probability of selecting a unit for more than one survey will be denoted by terms such as $P_i(s_j s_k)$, $P_i(s_j s_k s_l)$, $P_i(s_j s_k \bar{s}_l)$, etc. For example, $P_i(s_j s_k)$ is the probability that unit i is selected for both survey j and survey k. Similarly, $P_i(s_j s_k s_l)$ is the probability that unit i is selected for survey j, survey k, and survey l, whereas $P_i(s_j s_k \bar{s}_l)$ is the probability that unit i is selected for survey j and survey k, but not selected for survey l. The set of overlapping units among several samples is simply the intersection ($\cap$) of the samples, while the combined sample for two or more surveys is the union ($\cup$) of the samples.

An essential step in the overlap minimization procedures described in this paper involves partitioning the sampling frame into mutually exclusive subsets that depend on the inclusion status of the sample units for the previous surveys. These subsets are then combined and arranged using a "response load indicator." The response load indicator for a particular unit in the sampling frame is defined to be the number of times the unit was included in one or more previous surveys. For example, a response load of 0 indicates the unit was not included in any of the previous surveys, while a response load of 3 means the unit was included in exactly three of the previous surveys.

3. A Sampling Scheme to Minimize Overlap Between Two Surveys

Table 1 summarizes the notation and quantities needed to minimize overlap between two surveys. In this case, there are only two possible subsets reflecting the selection status of units for the previous survey, i.e., the set of units that were selected for the previous survey, s_1, and the complementary set, $\bar{s}_1$. The two subsets are arranged in ascending order of the response load indicator so that the priority in sample selection can be given to those units that have the smaller response load (in this case, a response load of 0). The "prior probabilities" in column (3) are quantities that relate to selection probabilities for the first survey, while the "target probabilities" in column (4) are functions of the probabilities of selection for both the previous and current surveys. In particular, the target probability in the first row of the table (corresponding to a response load of 0) is the desired unconditional probability of selection for the current survey, while the target probability in the second row (corresponding to a response load of 1) is the difference between the desired probability for the current survey and the probability that the unit was not selected for the previous survey (i.e., the prior probability indicated in the first row of the table).

Table 1. Selection scheme to minimize overlap between two surveys

[1] Response load	[2] Subset	[3] Prior probability	[4] Target probability	[5] Conditional probability
0	$\bar{s}_1$	$P_i(\bar{s}_1)$	$P_i(s_2)$	Min{[4]/[3], 1}
1	s_1	$P_i(s_1)$	$P_i(s_2) - P_i(\bar{s}_1)$	Max{0, [4]/[3]}

Finally, the "conditional probabilities" of selection for the two subsets are presented in column (5) of the table. The conditional probabilities are those that will be used to select the current sample, and depend on the response load and the ratio of the target and prior probabilities in a row (denoted by the symbol, [4]/[3]). For example, a unit with a response load of 0 (corresponding to row 1 of the table) for which the ratio is less than 1 will be selected with a conditional probability equal to the ratio. On the other hand, if the ratio is greater than or equal to 1 the unit will be selected with a conditional probability of 1. For units with a response load of 1 (corresponding to row 2 of the table), the conditional probability of a unit for which the ratio is either zero or negative will be zero, whereas for a unit for which the ratio is positive, the conditional probability will be equal to the ratio.

In effect, the conditional probabilities are designed to select as many units as possible from the subset with a response load of 0. For example, in the case of equal probability sampling if the prior probability of every unit in $\bar{s}_1$ is greater than the corresponding target probability specified in the first row of Table 1 (i.e., the ratio [4]/[3] is always less than 1), then there are enough units in $\bar{s}_1$ from which to select the sample for the current survey. In this case, there is no need to select additional units from the complementary subset, s_1 in order to achieve the desired probability of selection. However, if there is at least one unit in $\bar{s}_1$ for which the prior probability is less than the corresponding target probability (i.e., the ratio [4]/[3] is more than 1), then the probability available for selecting that unit is not large enough to achieve the desired probability of selection for the second survey. In this case, all such units in $\bar{s}_1$ and some of the units in s_1 will be selected for the current survey. This clearly indicates the optimality of the scheme in terms of minimizing overlap.

Moreover, it can easily be seen that the unconditional probability of selecting a unit for the current survey is equal to the desired probability of selection. For those units in the frame for which $P_i(\bar{s}_1) > P_i(s_2)$, the probability of selecting the unit for the current survey is:

$$P_i(\bar{s}_1)\left\{\frac{P_i(s_2)}{P_i(\bar{s}_1)}\right\} + P_i(s_1)(0) = P_i(s_2).$$

Similarly, for those units in the frame for which $P_i(\bar{s}_1) \le P_i(s_2)$, the probability of selecting the unit for the current survey is:

$$P_i(\bar{s}_1)(1) + P_i(s_1)\left\{\frac{P_i(s_2) - P_i(\bar{s}_1)}{P_i(s_1)}\right\} = P_i(s_2).$$

The approach can also be applied to situations where there are changes in the frame or when the stratification schemes for the surveys are different. For instance, if a new unit has been added to the frame, that unit had no chance of selection for the first survey. This unit will belong to the first row of Table 1 (corresponding to a response load of 0) with a prior probability of 1, and hence the conditional probability will equal to the desired unconditional probability of selection for the current survey. The approach accommodates differences in stratification schemes since the conditional selection probabilities required to select the current sample depend only on the prior and target probabilities in the new stratum.

4. A Sampling Scheme to Minimize Overlap Among Three or More Surveys

Table 2 presents an analogous selection scheme to minimize overlap among three surveys. As in the case of two surveys, all possible subsets defined in terms of the inclusion status in the previous two surveys are arranged in ascending order of the response load indicator. Note that subsets with a response load of 1 have been combined in Table 2. The prior probabilities and the target probabilities are specified in column (3) and column (4), respectively. As defined earlier, the target probability in the first row (corresponding to a response load of 0) is equal to the desired probability for the current survey, i.e., $P_i(s_3)$. In the remaining rows, the target probability is the difference between the target in the given row and the prior probability in the previous row. The conditional probabilities to be used to select the current sample, which are functions of the target and the prior probabilities given in the table, are presented in the last column.

Table 2. Selection scheme to minimize overlap among three surveys

[1] Response load	[2] Subset	[3] Prior probability	[4] Target probability	[5] Conditional probability
0	$\bar{s}_1 \cap \bar{s}_2$	$P_i(\bar{s}_1\bar{s}_2)$	$P_i(s_3)$	Min{[4]/[3], 1}
1	$(\bar{s}_1 \cap s_2) \cup (s_1 \cap \bar{s}_2)$	$P_i(\bar{s}_1 s_2)+P_i(s_1\bar{s}_2)$	$P_i(s_3)-P_i(\bar{s}_1\bar{s}_2)$	Max{0, Min{[4]/[3],1}}
2	$s_1 \cap s_2$	$P_i(s_1 s_2)$	$P_i(s_3)-P_i(\bar{s}_1\bar{s}_2)-P_i(\bar{s}_1 s_2)-P_i(s_1\bar{s}_2)$	Max{0,[4]/[3]}

There are three different ways in which a unit can be selected for the current survey under the scheme presented in Table 2. In each of these cases, the overall unconditional probability of selection is same as the desired probability of selection for the third survey as shown below.

Case 1: If $P_i(\bar{s}_1\bar{s}_2) \geq P_i(s_3)$ the conditional probability in the first row will be less than or equal 1, and the conditional probabilities in the second and third rows will be zero. Therefore, the overall probability of selection in this case is:

$$P_i(\bar{s}_1\bar{s}_2) \times \frac{P_i(s_3)}{P_i(\bar{s}_1\bar{s}_2)} + \{P_i(\bar{s}_1 s_2) + P_i(s_1\bar{s}_2)\} \times 0$$
$$+ P_i(\bar{s}_1\bar{s}_2) \times 0 = P_i(s_3).$$

Case 2: If $P_i(\bar{s}_1\bar{s}_2) < P_i(s_3)$ but $P_i(\bar{s}_1 s_2) + P_i(s_1\bar{s}_2) \geq P_i(s_3) - P_i(\bar{s}_1\bar{s}_2)$, the conditional probability will be 1 in the first row, less than or equal to 1 in the second row, and zero in the third row. In this case, the overall probability of selection is:

$$P_i(\bar{s}_1\bar{s}_2) \times 1 + \{P_i(s_1\bar{s}_2) + P_i(\bar{s}_1 s_2)\} \times$$
$$\frac{P_i(s_3) - P_i(\bar{s}_1\bar{s}_2)}{P_i(s_1\bar{s}_2) + P_i(\bar{s}_1 s_2)} + P_i(\bar{s}_1\bar{s}_2) \times 0 = P_i(s_3).$$

Case 3: If $P_i(\bar{s}_1\bar{s}_2) < P_i(s_3)$ but $P_i(\bar{s}_1 s_2) + P_i(s_1\bar{s}_2) < P_i(s_3) - P_i(\bar{s}_1\bar{s}_2)$, the conditional probabilities will be 1 in the first and the second rows and will be less than 1 in the third row. The overall probability of selection is equal to the desired probability of selection for the third survey:

$$\{P_i(\bar{s}_1\bar{s}_2) + P_i(s_1\bar{s}_2) + P_i(\bar{s}_1 s_2)\} \times 1 + P_i(\bar{s}_1\bar{s}_2) \times$$
$$\frac{P_i(s_3) - P_i(\bar{s}_1\bar{s}_2) - P_i(s_1\bar{s}_2) - P_i(\bar{s}_1 s_2)}{P_i(\bar{s}_1\bar{s}_2)} = P_i(s_3).$$

Table 3 presents the selection scheme for minimizing overlap among four surveys. As in Tables 1 and 2, the possible subsets are formed based on the inclusion status of the units in the previous three surveys. The subsets are then arranged in the table in ascending order of the response load indicator. The prior and the target probabilities are included in columns (3) and (4). The expressions for the conditional probabilities in the first and the last rows are the same as the first and the last rows of Tables 1 and 2. The expressions for the conditional probabilities in the intermediate rows are all same and equal the expression in the middle row of Table 2. Therefore, no additional expressions for conditional probabilities are required. It can be easily verified that the overall unconditional selection probabilities in all possible cases are the same as the desired selection probability in the fourth survey.

Although the approach is unbiased and optimal for minimizing overlap, the realized sample sizes can vary from the desired sample sizes if there are differences in stratification or changes in the sampling frame. The magnitude of the difference will depend on the extent of these changes. As long as the changes are not large the difference between the achieved and the target sample sizes should not be unduly large. If achieving the exact sample sizes is important, then a modified conditional probability can be used to ensure that the exact number of units by stratum is selected. However, the use of modified conditional probabilities could result in less than optimal overlap control.

Table 3. Selection scheme to minimize overlap among four surveys

[1] Response load	[2] Subset	[3] Prior probability	[4] Target probability	[5] Conditional probability
0	$\bar{s}_1 \cap \bar{s}_2 \cap \bar{s}_3$	$P_i(\bar{s}_1\bar{s}_2\bar{s}_3)$	$P_i(s_3)$	Min{[4]/[3], 1}
1	$(\bar{s}_1 \cap \bar{s}_2 \cap s_3)$ $\cup(\bar{s}_1 \cap s_2 \cap \bar{s}_3)$ $\cup(s_1 \cap \bar{s}_2 \cap \bar{s}_3)$	$P_i(\bar{s}_1\bar{s}_2 s_3)+$ $P_i(\bar{s}_1 s_2\bar{s}_3)+$ $P_i(s_1\bar{s}_2\bar{s}_3)$	$P_i(s_3)-P_i(\bar{s}_1\bar{s}_2\bar{s}_3)$	Max[0, Min{[4]/[3],1}]
2	$(\bar{s}_1 \cap s_2 \cap s_3)$ $\cup(s_1 \cap \bar{s}_2 \cap s_3)$ $\cup(s_1 \cap s_2 \cap \bar{s}_3)$	$P_i(\bar{s}_1 s_2 s_3)+$ $P_i(s_1\bar{s}_2 s_3)+$ $P_i(s_1 s_2\bar{s}_3)$	$P_i(s_3)-P_i(\bar{s}_1\bar{s}_2\bar{s}_3)-$ $P_i(\bar{s}_1\bar{s}_2 s_3)-P_i(\bar{s}_1 s_2\bar{s}_3)-$ $P_i(s_1\bar{s}_2\bar{s}_3)$	Max[0, Min{[4]/[3],1}]
3	$s_1 \cap s_2 \cap s_3$	$P_i(s_1 s_2 s_3)$	$P_i(s_3)-P_i(\bar{s}_1\bar{s}_2\bar{s}_3)-$ $P_i(\bar{s}_1\bar{s}_2 s_3)-P_i(\bar{s}_1 s_2\bar{s}_3)-$ $P_i(s_1\bar{s}_2\bar{s}_3)-P_i(\bar{s}_1 s_2 s_3)-$ $P_i(s_1\bar{s}_2 s_3)-P_i(s_1 s_2\bar{s}_3)$	Max{0,[4]/[3]}

Similar overlap minimization schemes can be developed for more than four surveys by following the schemes above.

5. Implementation

To apply the general approach outlined above, the selection status and selection probabilities corresponding to each of the previous samples are required. For example, in the case of three surveys, in addition to the selection status, the probabilities required from the previous two surveys are $P_i(s_1)$, $P_i(s_2)$, and

$$P_i(\bar{s}_1\bar{s}_2)=P_i(\bar{s}_1)P_i(\bar{s}_2|\bar{s}_1)=P_i(\bar{s}_1)\{1-P_i(s_2|\bar{s}_1)\},$$

$$P_i(s_1 s_2)=P_i(s_1)P_i(s_2|s_1),$$

$$P_i(s_1\bar{s}_2)=P_i(s_1)-P_i(s_1 s_2),$$

$$P_i(\bar{s}_1 s_2)=P_i(s_2)-P_i(s_1 s_2).$$

All of the above probabilities can be derived if the following four probabilities are available:

$$P_i(s_1),\ P_i(s_2),\ P_i(s_2|s_1), \text{and } P_i(s_2|\bar{s}_1).$$

In the case of four surveys, in addition to $P_i(s_3)$ and those listed above, the following probabilities will be required:

$$P_i(\bar{s}_1\bar{s}_2\bar{s}_3)=P_i(\bar{s}_1\bar{s}_2)P_i(\bar{s}_3|\bar{s}_1\bar{s}_2)=P_i(\bar{s}_1\bar{s}_2)\{1-P_i(s_3|\bar{s}_1\bar{s}_2)\},$$

$$P_i(\bar{s}_1\bar{s}_2 s_3)=P_i(\bar{s}_1\bar{s}_2)P_i(s_3|\bar{s}_1\bar{s}_2),$$

$$P_i(s_1 s_2\bar{s}_3)=P_i(s_1 s_2)\{1-P_i(s_3|s_1 s_2)\},$$

$$P_i(s_1\bar{s}_2 s_3)+P_i(\bar{s}_1 s_2 s_3)=\{P_i(s_1\bar{s}_2)+P_i(\bar{s}_1 s_2)\}P_i\{s_3|(s_1\bar{s}_2)\cup(\bar{s}_1 s_2)\},$$

$$P_i(s_1\bar{s}_2\bar{s}_3)+P_i(\bar{s}_1 s_2\bar{s}_3)=\{P_i(s_1\bar{s}_2)+P_i(\bar{s}_1 s_2)\}[1-P_i\{s_3|(s_1\bar{s}_2)\cup(\bar{s}_1 s_2)\}],$$

$$P_i(s_1 s_2 s_3)=P_i(s_1 s_2)P_i(s_3|s_1 s_2).$$

Note that all of the above probabilities can be calculated if the following four probabilities are available:

$$P_i(s_3),\ P_i(s_3|\bar{s}_1\bar{s}_2),\ P_i(s_3|s_1 s_2),\ P_i\{s_3|(s_1\bar{s}_2)\cup(\bar{s}_1 s_2)\}.$$

If the samples for the previous surveys in Tables 2 or 3 were selected by following the overlap minimization approach given in Table 1, all of the required conditional probabilities can be derived from the unconditional selection probabilities. In this case, only the selection status and selection probabilities in the previous surveys are required to implement the method.

6. Application to NCES Surveys

In the spring of 2000, a stratified sample of public schools was designed and selected for the first biennial School Survey on Crime and Safety (SSOCS). However, school samples had been previously selected for the National Assessment of Educational Progress (NAEP), Early Childhood Longitudinal Study (ECLS-K), the Schools and Staffing Survey (SASS), and a Fast Response Survey System (FRSS) survey on teacher quality issues. The SASS sample had been selected using the scheme outlined in Table 1 to minimize overlap with the combined samples for NAEP and ECLS-K. Similarly, the sample for the FRSS survey was selected by minimizing the overlap with NAEP/ECLS-K and SASS using the scheme specified in Table 2. Therefore, to minimize overlap with all three of the previously-selected samples, the SSOCS sample was selected using the minimization scheme presented in Table 3.

In Tables 4 and 5, we summarize some results of the overlap minimization process for the FRSS and SSOCS samples. For comparison, we also present rough estimates of the expected overlap assuming independent sampling (i.e., in the absence of any overlap control). Table 4 shows that if the NAEP/ECLS-K, SASS, and FRSS samples had been selected independently, an expected 24 percent of the FRSS sample would have overlapped with either the

NAEP/ECLS-K or SASS sample (but not both). Under the overlap minimization selection procedure, however, only 5.6 percent of the FRSS sample overlapped with one of the previous samples.

Table 4. Distribution of FRSS sample under alternative sampling schemes by response load

Response load	Independent selection*		Overlap minimized selection (FRSS)†	
	Frame	Sample	Frame	Sample
0 (not selected for any survey)	60,359 (74.1%)	1,636 (74.1%)	59,501 (73.1%)	2,085 (94.4%)
1 (NAEP/ECLS-K only)	11,515 (14.1%)	312 (14.1%)	12,373 (15.2%)	90 (4.1%)
1 (SASS only)	8,004 (9.8%)	217 (9.8%)	8,862 (10.9%)	34 (1.5%)
2 (NAEP/ECLS-K and SASS)	1,527 (1.9%)	41 (1.9%)	669 (0.8%)	0 (0.0%)
Total	81,405 (100%)	2,207 (100%)	81,405 (100%)	2,209 (100%)

*Assumes all samples are selected independently.
†Assumes all samples are selected using overlap minimization procedures.

Table 5. Distribution of SSOCS sample under alternative sampling schemes by response load

Response load	Independent selection*		Overlap minimized selection (FRSS)†	
	Frame	Sample	Frame	Sample
0 (not selected for any survey)	58,723 (72.1%)	2,422 (72.1%)	57,416 (70.5%)	3,172 (94.3%)
1 (NAEP/ECLS-K only)	11,203 (13.8%)	462 (13.8%)	12,283 (15.1%)	144 (4.3%)
1 (SASS only)	7,787 (9.6%)	321 (9.6%)	8,828 (10.8%)	38 (1.1%)
1 (FRSS only)	1,636 (2.0%)	68 (2.0%)	2,085 (2.6%)	8 (0.2%)
2 (NAEP/ECLS-K and SASS)	1,486 (1.8%)	61 (1.8%)	669 (0.8%)	0 (0.0%)
2 (NAEP/ECLS-K and FRSS)	312 (0.4%)	13 (0.4%)	90 (0.1%)	0 (0.0%)
2 (SASS and FRSS)	217 (0.3%)	9 (0.3%)	34 (0.0%)	0 (0.0%)
3 (NAEP/ECLS-K SASS, FRSS)	41 (0.1%)	2 (0.0%)	0 (0.0%)	0 (0.0%)
Total	81,405 (100%)	3,358 (100%)	81,405 (100%)	3,362 (100%)

*Assumes all samples are selected independently.
†Assumes all samples are selected using overlap minimization procedures.

Similar results are presented for the SSOCS sample in Table 5. Under the overlap minimization procedure only 5.7 percent of the SSOCS sample overlapped with one of the previous samples and the remaining 94.3 percent did not overlap with any of the previous samples. Under independent sampling, an expected 25.3 percent of the SSOCS sample would have overlapped with the samples of one of the previous surveys, and 2.5 percent would have overlapped with the samples of two or more the previous surveys.

7. Discussion

The overlap minimization schemes presented in Tables 1 to 3 assume that the response loads of all surveys are equal. However, the general approach presented in this paper can also be applied to the case where different surveys have different levels of response burden. By applying appropriate weights to the different surveys, a weighted response load indicator can be derived. For example, if the reporting burden associated with the first or second survey is half of that associated with the third survey, the response load indicator can be derived by giving a weight of 0.5 to each of the first two surveys and a weight of 1.0 to the third survey. In this case, the response load indicator for a unit included in both the first and the second surveys will be 1.0. For units included in both the second and the third surveys, the response load indicator will be 1.5, and so on. Then combining and arranging the subsets based on the weighted response load indicator, a minimization scheme can be developed. For instance, the units selected only for the third survey can be combined with the units selected for both the first and the second surveys, and the associated conditional probabilities can be calculated accordingly.

With a few minor modifications, the approach can also be adapted for use in maximizing sample overlap. The basic difference in the setup of the sampling schemes is that the subsets will be arranged in descending instead of ascending order of the response load indicator.

Finally, we note that the procedures given here generally will not guarantee that the realized sample size equals the desired sample size. In particular, changes in the sampling frame and differences in stratification schemes will contribute to the variation in sample size. However, in practice, the difference between the desired and actual sample sizes are expected to be small. For the NCES surveys described above, there were only minor differences in the sampling frames, but major differences in the way the schools were stratified for the various studies. However, despite these differences, the achieved sizes were generally close to the target numbers. For example, only two additional schools were selected for the FRSS sample (compared with the target sample size), and only four additional schools were selected for the SSOCS sample. The variation in sample size at the stratum level was greater, but again, the differences were generally small (i.e., the difference between the actual and desired sample sizes were no more than 2 or 3 in most cases).

8. References

Brick, J.M., Morganstein, D.R., and Wolters, C.L. (1987). Additional Uses for Keyfitz Selection. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 787-791.

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

Ernst, L.R. (1999). The Maximization and Minimization of Sample Overlap Problems: A Half Century of Results. A Paper Presented at the International Statistical Conference, 1999.

Keyfitz, N. (1951). Sampling with Probability Proportional to Size: Adjustment for Changes in Probabilities. *Journal of the American Statistical Association*, **46**, pp. 105-109.

Kaufman, S. (1993). 1990-91 Schools and Staffing Survey: Sample Design and Estimation (Technical Report NCES 93-449). Washington, DC: National Center for Education Statistics.

Kish, L. and Scott, A. (1971). Retaining Units After Changing Strata and Probabilities. *Journal of the American Statistical Association*, **66**, pp. 461-470.

Ohlsson, E. (1995). Coordination of Samples Using Permanent Random Numbers. In Business Survey Methods, eds. B.G. Cox, D.A. Binder, D.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott. New York: John Wiley & Sons, pp. 153-169.

Perry, C.R., Burt, J.C., and Iwig, W.C. (1993). Methods of Selecting Samples in Multiple Surveys to Reduce Respondent Burden. *Proceedings of the International Conference on Establishment Surveys, American Statistical Association*, pp. 345-351.

Rust, K.F. and Johnson, E.G. (1992). Sampling and Weighting in the National Assessment. *Journal of the Educational Statistics*, Vol. 17, **2**, pp. 111-129.