# The Maximization and Minimization of Sample Overlap Problems: A Half Century of Results

Lawrence R. Ernst
*U.S. Bureau of Labor Statistics, Office of Compensation and Working Conditions*
*2 Massachusetts Avenue, NE, Room 3160*
*Washington, DC 20212, U.S.A.*
*Ernst_L@BLS.gov*

## 1. Introduction

In the nearly half century since Keyfitz's (1951) pioneering work, research on the problem of maximizing and the related problem of minimizing overlap of sampling units has progressed in many different directions and under several different names, such as optimal integration of surveys, sample coordination, and overlapping maps.

A typical application of sample overlap is as follows. Units are selected for a survey from a stratified design with probability proportional to size (pps) without replacement. At a later date a new sample is to be selected using new size measures and a different stratification. To reduce costs it may be desirable to maximize the expected number of units common to the two samples while preserving prespecified selection probabilities for the units in the new design, either selection probabilities for individual units or selection probabilities for the possible sets of sample units in a new stratum. For example, when the units being overlapped are primary sampling units (PSUs), which are geographic areas, an overlap maximization procedure can reduce the costs associated with hiring a new interviewer; when the units are ultimate sampling units, such a procedure can reduce the extra costs of an initiation interview. Minimization of overlap, in contrast, is typically employed as method of reducing respondent burden.

Overlap procedures have been developed for application in many different situations. In Section 2 we detail eight properties of overlap procedures corresponding to the columns in Table 1. We discuss a number of key overlap procedures developed for the case when the samples to be overlapped are selected sequentially in Section 3, and for the case when the samples are selected simultaneously in Section 4. The discussion includes all the overlap procedures in Table 1 and some additional overlap procedures.

There is one class of overlap procedures that is omitted in Table 1 and which will only be discussed briefly in Section 3. These are procedures employing permanent random numbers (PRNs). An excellent discussion of these procedures is found in Ohlsson (1995).

## 2. Properties

*1. Sequential or simultaneous selection.* In the redesign illustration above, the two samples must be selected sequentially since the designs are for different points in time. However, in some applications samples are to be selected at the same point in time for two or more surveys, with different measures of size and possibly different stratifications. Although overlap procedures developed for sequential selection can also be used in the case of simultaneous selection, some overlap procedures have been developed specifically to be used for simultaneous selection and generally produce a better overlap than procedures developed for sequential selection or are computationally more efficient.

*2. Number of units (n) per stratum.* The procedures are either intended for 1, a small number (S), or a large number (L) of units per stratum designs. In all cases the sampling is without replacement. In theory, any of the procedures other than those restricted to 1 per stratum designs can be used for any value of *n*. However, those procedures developed for small *n*, many of which use linear programming (LP), typically cannot be used when $n$ is large because the solution algorithms become computationally infeasible. Also, all S procedures, which are typically used for selecting PSUs in a multistage design, most commonly for a household survey, preserve prespecified joint selection probabilities for units in a stratum in the new design for sequential procedures and for all designs for simultaneous procedures. All L procedures, which are typically used for selecting sample units in a single stage design, most commonly for establishment surveys, only preserve prespecified individual selection probabilities for the new design for sequential procedures and prespecified individual selection procedures for all designs for simultaneous procedures. Furthermore, although it is not emphasized in this paper, $n$ can generally vary among strata in the same design and among the designs being overlapped, except for $n=1$ procedures.

3. *Maximization/minimization.* Some procedures were developed for the maximization problem, some for minimization, and some for both, as indicated. Typically, the LP procedures are easily adaptable for both problems by maximizing or minimizing the same objective function. A "max" or "min" in parentheses indicates that the procedure was not described for this application but can be easily modified to apply to it.

4. *Restratification.* Some procedures allow for different stratifications (Y) in the designs being overlapped, while others do not (N).

5. *Optimal.* The procedure produces the optimal overlap (Y) or generally does not (N).

6. *Linear programming.* The procedure either does not use LP (N) or uses one of the following types of LP: transportation problem (TP), real-valued LP (LP), or integer LP (IP). Schrijver (1986) is a reference for LP.

7. *Independence of sampling from stratum to stratum.* This applies to the new design in sequential selection and to each design in simultaneous selection. Satisfying this condition (Y) is important for obtaining unbiased variance estimates. However, few procedures that allow for different stratifications in the designs overlapped satisfy it. For, as illustrated in Ernst (1986), consider the following situation. Two units, denoted units 1 and 2, were in the same initial stratum and are in different new strata, where $n=1$ for both designs. Furthermore, unit 1 was in the initial sample. Generally in that situation unit 1 would have a conditional selection probability for the new design greater than its unconditional selection probability, while the opposite would be true for unit 2, and these units would not be selected independently in the new design.

In the case of sequential procedures, if this independence condition is not satisfied and the procedure is used for two successive redesigns, then the sampling of the initial strata for the second redesign, which were the new strata in the first redesign, would not have been done independently

***Table 1. Properties of Overlap Procedures***

| Procedure | Seq/sim | $n$ | Max/min | Restrat. | Optimal | LP | Indepen. | Surveys |
|---|---|---|---|---|---|---|---|---|
| Keyfitz | Seq | 1 | Max | N | Y | N | Y | 2 |
| Perkins | Seq | 1 | Max | Y | N | N | JNR | 2 |
| K&S | Seq | 1 | Max | Y | N | N | IR | 2 |
| Sunter | Seq | S | Max | N | N | N | Y | 2 |
| Ohlsson | Seq | S | Max (min) | Y | N | N | Y | $\geq 2$ |
| CCE | Seq | S | Both | Y | Y | TP | JR | 2 |
| Ernst86 | Seq | S | Both | Y | N | LP | JNR | 2 |
| E&I | Seq | S | Max | Y | N | TP | IR | 2 |
| Pollock | Seq | L | Max (min) | Y | Y | N | JNR | 2 |
| Ernst95 | Seq | L | Both | Y | N | N | JNR | 2 |
| M&P | Sim | 1 | Max | N | $\leq 3$ surveys | N | Y | $\geq 2$ |
| PBI | Sim | L | Min (max) | Y | N | IP | N | $\geq 2$ |
| Ernst96 | Sim | 1 | Both | Y | Y | TP | N | 2 |
| Ernst98 | Sim | L | Both | N | Y | TP | Y | 2 |
| Ernst99 | Sim | S | Both | N | Y | LP | Y | $\geq 2$ |

from stratum to stratum. Depending on the procedure, this can make it difficult or impossible to use the procedure in the selection of the new units for the second redesign and still preserve the prespecified selection probabilities in the new design. A sequential procedure that does not satisfy this independence can be classified into one of the following three categories in order of decreasing desirability, and are listed in the column by the indicated abbreviation.

Selection probabilities in the new design are not conditioned on joint selection probabilities in the initial design of units in different initial strata (JNR).

Selection probabilities in the new design are conditioned on joint selection probabilities in the initial design of units in different initial strata (JR).

The procedure requires that the sampling of the initial strata be independent from stratum to stratum in order to preserve selection probabilities in the new design (IR).

Thus, the IR procedures cannot be used in two successive redesigns if the selection probabilities in the new design are to be preserved. In practice JR procedures cannot be used then either, since it typically is not feasible to compute these joint probabilities.

8. *The number of surveys that can be overlapped.* This is two for most sequential procedures.

## 3. Sequential Procedures

We consider first the procedures developed for sequential selection. The various procedures for overlap maximization of two samples when the samples are selected sequentially are based on the same general principle, namely that these procedures do not alter the unconditional probability of selecting a unit in the new sample (or the probability of selecting a particular sample when the joint selection probabilities are prespecified), but condition the probability of selection of a new sample on the set of initial sample units in such a manner that the conditional probability of selection of a unit in the new sample is in general greater than its unconditional selection probability when the unit was in the initial sample and less otherwise. Overlap minimization follows the same principle with the obvious modifications.

For sequential selection, overlap procedures can be put in the following context. Let $S$ be a stratum in the new design consisting of $N$ units and let $p_i, p_{\cdot i}$, $i = 1,...,N$, denote the probability that the $i$-th unit in $S$ is in the initial and new samples, respectively. Let $I, J$ denote the random sets consisting of all units in $S$ in the initial and new samples, respectively, and let $I_1,....,I_M$, $J_1,...,J_{N'}$, denote all possibilities for $I$ and $J$, respectively. Typically, $J_1,...,J_{N'}$ are all subsets of $S$ consisting of exactly $n$ units and $N' = \binom{N}{n}$. As for the $I_i$'s, a subset $I^*$ of $S$ can be among the $I_i$'s only if no more than $n$ units from each initial stratum are in $I^*$ and at least $n$ units in each initial stratum are not in $S \sim I$. For $i = 1,...,M$, $j = 1,...,N'$, let $p'_i = P(I = I_i)$, $p'_{\cdot j} = P(J = J_j)$, $p'_{\cdot j|i} = P(J = J_j \mid I = I_i)$, and let $n_{ij}$ denote the number of units in $I_i \cap J_j$. Let $p_{\cdot j|i} = P(j \in J \mid I = I_i)$, $j = 1,...,N$. Now, there are two cases. In Case 1 there are prespecified probabilities, $p'_{\cdot j}$, for each $J_j$. For example, when $n = 2$ the $p'_{\cdot j}$'s might be the Brewer-Durbin probabilities (Cochran, 1977). In Case 2 only the individual selection probabilities, $p_{\cdot j}$, are prespecified. Case 1 applies to all the small $n$ procedures in Table 1; Case 2 applies to all the large $n$ procedures; and for $n = 1$ both cases are the same, with $p'_{\cdot j} = p_{\cdot j}$, $p'_{\cdot j|i} = p_{\cdot j|i}$.

In Case 1 we wish to obtain a set of $p'_{\cdot j|i}$'s which maximize or minimize

$$(1) \qquad \sum_{i=1}^{M} \sum_{j=1}^{N'} n_{ij} p'_i p'_{\cdot j|i}$$

subject to the constraints

(2) $$\sum_{i=1}^{M} p'_i p'_{j|i} = p'_j\,, \quad j = 1,\ldots,N'$$

(3) $$\sum_{j=1}^{N'} p'_{j|i} = 1\,, \quad i = 1,\ldots,M$$

Here (1) is the expected number of units in common to the initial and new samples; (2) simply states that the conditional selection probabilities do not alter the unconditional selection probabilities for the $J_j$'s; and (3) is required since exactly one $J_j$ is selected as the set of new sample units regardless of which $I_i$ was the set of initial sample units. In fact, although not all overlap procedures take this approach, maximization and minimization of overlap can be viewed as a LP problem, that is the maximization or minimization of objective function (1) subject to the constraints (2) and (3), where (1)-(3) are all linear functions of the only variables, the $p'_{j|i}$'s. Only an optimal procedure actually optimizes (1), but all small *n* sequential procedures satisfy (2) and (3).

In Case 2, analogously, we wish to obtain a set of $p_{j|i}$'s which maximize or minimize

(4) $$\sum_{i=1}^{M} \sum_{j=1}^{N} p'_i p_{j|i}$$

subject to the constraints

(5) $$\sum_{i=1}^{M} p'_i p_{j|i} = p_j\,, \quad j = 1,\ldots,N$$

(6) $$\sum_{j=1}^{N} p_{j|i} = n\,, \quad i = 1,\ldots,M$$

Additionally, let *K* denote the number of initial strata with units in *S*, and for $k = 1,\ldots,K$, let $I'_k$ denote the set of units that are in both the *k*-th such stratum and in *S*.

To illustrate these concepts 1, consider the following example, denoted Example 1, which we will return to several times in this paper.

***Table 2. Unconditional Probabilities for Example 1***

| | | | $i$ | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| $p_i$ | .1 | .2 | .2 | .3 | .1 |
| $p_i$ | .1 | .26 | .18 | .36 | .1 |

In this example $n = 1$, $N = 5$, $K = 2$, $I'_1 = \{1,2,3\}$, $I'_2 = \{4,5\}$, and the sampling in the initial design is assumed independent from stratum to stratum. Since the sum of the $p_i$'s for units 1,2, and 3 is less than 1, as is the sum for units 4 and 5, neither $I'_1$ nor $I'_2$ constitute an entire initial stratum. Consequently, there are 12 $I_i$'s, which are listed in Table 3 with the corresponding $p_i'$'s.

***Table 3.*** $I_i$, $p_i'$, $n_{ij}$, ***and Optimal*** $p'_{j|i}$ ***for Example 1***

| | | | $n_{ij}$ | | | | | $p'_{j\|i}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | $j$ | | | | | $j$ | | |
| $i$ | $I_i$ | $p_i'$ | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 3 | 4 | 5 |
| 1 | {1} | .06 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 2 | {2} | .12 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 3 | {3} | .12 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 4 | {4} | .15 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 5 | {5} | .05 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 6 | {1,4} | .03 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 7 | {1,5} | .01 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 8 | {2,4} | .06 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 9 | {2,5} | .02 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 10 | {3,4} | .06 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 11 | {3,5} | .02 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 12 | ∅ | .3 | 0 | 0 | 0 | 0 | 0 | .133 | .4 | .133 | .2 | .133 |

As for the $J_j$'s, since $n = 1$ we have that $N' = N = 5$, $J_j = \{j\}$, and $p'_j = p_j$ for all *j*. The $n_{ij}$'s and an optimal set of $p'_{j|i}$'s are presented in Table 3. Generally there can be more than one set of optimal $p'_{j|i}$'s. The maximum value of (1) and (4), that is the value for these $p'_{j|i}$'s, is .700. This compares to an overlap probability of .216 when the new sample unit is selected independently of the initial sample.

In general, the probability that a unit *i* is in both samples can never exceed $\min\{p_i, p_i\}$, and consequently the maximum value of (1) and (4) can never exceed

$$(7) \qquad \sum_{i=1}^{N} \min\{p_i, p_i\}$$

which is .880 for Example 1. Analogously, the minimum value for (1) and (4) is never less than

$$(8) \qquad \sum_{i=1}^{N} \max\{p_i + p_i - 1, 0\}$$

In the following three subsections we consider sequential procedures for which $n = 1$, *n* is small, and *n* is large, respectively.

### 3.1 One Per Stratum Procedures

The concept of overlap maximization was first developed by Keyfitz (1951) who presented an optimal maximal overlap procedure, but only for $n = 1$ designs for which the initial and new stratifications are identical. Keyfitz's conditional probabilities of selection are as follows:

$$(9) \qquad p_{i|i} = \min\{p_i / p_i, 1\}$$

$$(10)\quad p_{j|i} = (1-\min\{p_i / p_i, 1\}) \frac{\max\{p_j - p_j, 0\}}{\sum_{l=1}^{N} \max\{p_l - p_l, 0\}}, \quad j \neq i$$

where it is understood that $I_i = \{i\}$ since the stratifications in the two designs are identical. Note that by (9), the probability of overlap for Keyfitz's procedure equals (7), and hence the procedure is optimal.

Perkins (1970), and Kish and Scott (K&S) (1971) considered the more general $n=1$ maximal overlap problem in which the strata definitions change in the new design. Perkins' procedure is a simple generalization of Keyfitz's procedure and was the overlap procedure used to select sample PSUs in the 1970s redesign of the household surveys conducted by the U.S. Bureau of the Census. K&S is slightly more complex. It generally yields a higher overlap than Perkins' procedure, but is more limited in its applicability because it is an IR procedure.

The following is the algorithm for Perkins' procedure. The first step is to determine from which $I'_k$ the new sample unit is to be selected. $I'_k$ is chosen from among $I'_1,...,I'_K$ with probability $y_k$, where $y_k$ is the sum of the $p_i$'s over all units *i* in $I'_k$. If $I'_k$ is selected and $I'_k \cap I = \{i\}$, then the new sample unit in *S* is chosen from among the units in $I'_k$ using conditional probabilities akin to the Keyfitz selection probabilities, with $y_k p_i$ substituted for $p_i$, and $y_k p_l$ substituted for $p_l$ in (9) and (10), and the summation in (10) only over units in $I'_k$. In addition, if $I'_k \cap I = \varnothing$, then the new sample unit is chosen from among the units *j* in $I'_k$ with probability proportional to $\max\{p_j - y_k p_j, 0\}$. To illustrate, for Example 1 we have $y_1 = .54$, $y_2 = .46$, and $y_k p_i \le p_i$ for all *i*. Consequently, if $I'_1$ is selected and unit 1, 2, or 3 was in the initial sample, then that unit would be retained with certainty in the new sample. If none of these three units were in the initial sample, then one of them, *j*, would be selected with probability proportional to $p_j - .54 p_j$. If $I'_2$ was selected, then either unit 4 or 5 would be selected in an analogous manner. The resulting overlap probability in this example with Perkins' method is .443.

Although Perkins' method does not directly produce a set of $p'_{j|i}$'s analogous to those in Table 3, these conditional probabilities can be calculated. For example, the values in the last five columns of the first row in Table 3 for Perkins' procedure are, respectively, .54, 0, 0, .09, .37. This is because if $I'_1$ is selected and $I = \{1\}$, then unit 1 is selected with certainty as the new sample unit in *S*; while if $I'_2$ is selected, then since neither unit 4 nor 5 was in the initial sample, one of these two units is selected with probability proportional to $p_j - .46 p_j$. Thus, if $I = \{1\}$, the probability of retaining unit 1 in the new sample with Perkins' method is only .54, in comparison with a probability of 1 using the original optimal conditional probabilities in Table 3, and .1 with independent selection.

There are several aspects of Perkins' procedure that keep it from producing the optimal overlap. First, using Example 1 as an illustration, if $I = \{1\}$ but $I'_2$ was selected, then Perkins' procedure would not make use of the knowledge that $1 \in I$. This cannot be helped, since if the procedure simultaneously used information about which units in $I'_1$ and $I'_2$ were in the initial sample, it would no longer be a JNR procedure. However, two other aspects of Perkins' procedure could be modified to improve the expected overlap while retaining JNR. In particular, if $I'_2$ was selected in Example 1 and neither unit 4 nor 5 was in the initial sample, then the procedure would select one of these two units anyway as the new sample unit, even though this offers no possibility of an overlap. It would be better to select one of the first three units in that situation, even without being able to use information on which, if any, of these units was in the initial sample, since this would offer a positive probability of overlap. Secondly, the probability, $y_k$, of selecting $I'_k$ used in

Perkins' procedure is generally not optimal. We return to these two issues later when discussing the Ernst (1986) procedure, which overcomes these two drawbacks.

K&S actually present several procedures. The authors claim that their Method II produces the highest overlap, and we present an outline of this method only. The $I'_k$ are ordered in decreasing order of $y_k$, where $y_k$ is as defined for Perkins' procedure. An attempt is first made to retain the unit in $I'_1 \cap I$ if this set is nonempty. If this attempt is successful, then the algorithm stops. If this attempt is not successful or if $I'_1 \cap I = \emptyset$, then an attempt is made to retain the unit in $I'_2 \cap I$, and so forth. For each *k*, if the algorithm has not stopped before reaching $I'_k$ and there is a unit *i* in $I'_k \cap I$, then the probability of it being retained is given by (9) with $p_i$ replaced by $Q_{k-1} p_i$, where $Q_0 = 1$ and $Q_k$, $k = 1,...,K$, is the probability that no unit has been selected for retention from among $I'_1,...,I'_k$. If no unit has been selected after passing through all of $I'_1,...,I'_K$, then a unit is selected from among all units *j* in *S* with probability proportional to $\min\{p_j - Q_{k-1} p_j, 0\}$, where *k* satisfies $j \in I'_k$. Note that it is in the calculation of the $Q_k$ that the IR assumption is used. The probability of overlap for Example 1 for this method is .688, the deviation from optimality occurring because $P(J = \{3\} \mid I = \{3\}) = .9$, not 1.

## 3.2 Small Number of Units Per Stratum Procedures

We will first discuss procedures for this type of problem which do not use LP techniques and then discuss LP procedures.

Fellegi (1966) was the first to develop a procedure for maximizing overlap when $n > 1$. He presented two procedures for the case when $n = 2$, the stratifications in the initial and new designs are identical, and a specific sampling procedure that he developed is used. The second procedure is readily generalized to any small value of *n* and any without replacement sampling procedure, which Sunter (1989) accomplished. Only Sunter's procedure is listed in Table 1. This procedure can also be viewed as a direct generalization of Keyfitz's to $n > 1$. In fact, the conditional probabilities for Sunter's method can be obtained from (9) and (10) by simply replacing $p_i, p_j, p_{j|i}$ by $p'_i, p'_j, p'_{j|i}$. Thus, for example, if $n = 2$ and a pair of units has a higher joint selection probability in the new design than in the initial design and that pair was in the initial sample, then it is retained with certainty in the new sample by the modified form of (9). However, unlike Keyfitz's procedure, Sunter's generalization to $n > 1$ is not optimal. This is because if, for example, $n = 2$ and the pair in the initial sample is not retained, Sunter's procedure does not assign a higher conditional probability to selecting a pair for the new sample with one unit in common with the initial sample, an outcome which is better than selecting a pair with no units in common.

Recently, Ohlsson (1999) developed a procedure for the maximization problem that is applicable when the strata definitions change in the new design, with the key feature that it selects the units in the new design independently from stratum to stratum. This is a generalization of the Ohlsson (1996) procedure, which only considered the case $n = 1$. Ohlsson's methodology uses transformed random numbers. This procedure, which he has dubbed exponential sampling, is innovative, simple, and applicable to more than two designs. In the $n = 1$ case he assigns a PRN $X_i$ in the interval (0,1) to each unit *i* in the frame. For each design *k* the transformed number $x_{ik} = [\log(1 - X_i)] / p_{ik}$ is assigned to unit *i*, where $p_{ik}$ is the probability of selecting unit *i* in design *k*. The unit with the smallest transformed number is selected. He also shows how to assign a PRN $X_i$ retrospectively if the initial sample had already been selected independently from stratum to stratum. (However, if the initial sample had been selected using an overlap procedure that destroyed this independence, then Ohlsson's procedure is not usable.) For $n = 2$ he combines his procedure for the $n = 1$ case with Brewer's method for selecting two units pps by adjusting the draw probabilities (Cochran 1977). Analogously, for $n > 2$ Ohlsson's method for $n = 1$ can be combined with Sampford's method (Cochran 1977). Although Ohlsson does not mention minimization of overlap, it appears that this can be achieved for two designs by selecting units for one of these designs exactly as Ohlsson describes, while in the second design using the set of transformed random numbers obtained by replacing $1 - X_i$ by $X_i$ in the formula for $x_{ik}$. The probability of overlap for Ohlsson's procedure for Example 1, calculated using equation (3.3) in Ohlsson (1996), is .607.

The remaining small *n* procedures employ LP techniques. Advantages of LP approaches include easy formulation, optimality, and flexibility in what to optimize. As an example of this flexibility, in Ernst (1986) the concept of partial success was incorporated into the objective function. This occurred when the PSUs that were being overlapped were redefined in some situations, allowing for a PSU in the new design to partially intersect a PSU in the initial design. In addition, LP approaches developed for the maximization problem typically can also be applied to the minimization problem by simply minimizing instead of maximizing the appropriate objective function. The main disadvantage of using LP procedures for sequential overlap problems is that they typically require intensive use of computer resources. For large *n*, the LP problems can become impracticably large, which is why these procedures are listed as applying only to small *n* problems, even though in theory they are applicable to large *n* problems also.

Des Raj (1956) first presented an LP approach, but only for the case considered by Keyfitz. Arthanari and Dodge (A&D) (1981) and Causey, Cox, and Ernst (CCE) (1985) generalized this approach to $n > 1$. A&D only considered the case of identical stratifications, while CCE allowed for restratification. Only the most general formulation, CCE, is listed in Table 1.

CCE, which is an optimal procedure, has essentially already been discussed. Basically, it is the solution of the LP problem (1)-(3) above, with the $p'_{j|i}$'s as the variables. Table 3 provides an illustration of CCE for Example 1. The one difference between CCE and (1)-(3) as presented, is that the former is a TP while the latter is not. A TP can be viewed as a particular type of LP problem in which the set of variables forms a two-dimensional array, with the constraints specifying the row and column marginals. Although (1)-(3) is not of the required form for a TP because of the $p'_i$ coefficients in (2), it can be easily converted to this form with the transformation $x_{ij} = p'_i p'_{j|i}$.

There are two drawbacks to CCE. First, because CCE is a JR procedure, it is not applicable to a redesign of a survey for which the previous redesign used an overlap procedure that did not preserve the independence of sampling from stratum to stratum. The second drawback is that even if the sample units in the initial design were selected independently from stratum to stratum, CCE may not be usable because the TPs involved can become impracticably large even when $n = 1$. That is, the computer resource issue discussed above, which can occur with all overlap procedures using LP, is most acute for CCE. In the extreme case when a stratum consists of *N* units, all of which were in different initial strata, the number of variables in the TP is $N2^N$ for $n = 1$. This is because in that case there is an $I_i$ corresponding to each of the $2^N$ subsets of *S* and a $J_j$ corresponding to each of the *N* units in *S*, and hence there are $N2^N$ variables $p_{j|i}$.

To overcome the first drawback, an alternative JNR, LP procedure was developed by Ernst (1986), which generally produces a better overlap than Perkins' procedure for $n = 1$, but which is not restricted to $n = 1$. It also is typically operationally feasible when $n = 1$ or $n = 2$. This procedure was used by the U.S. Bureau of the Census to select the sample PSUs in both the 1980s and 1990s redesigns for two major household surveys which required a JNR procedure, the Current Population Survey (CPS), and the National Crime Victimization Survey (NCVS). As an example of the results, the proportion of new sample PSUs that overlapped with the Ernst (1986) procedure for the 1980s redesign of NCVS was .81. Perkins' procedure and independent selection of the new PSUs would have produced overlap proportions of .67 and .59, respectively.

The Ernst (1986) procedure uses the basic idea of the Perkins procedure, that is, it also selects one of the initial strata $I'_k$ and conditions the selection probabilities for the $J_j$ on the set of units in $I \cap I'_k$. Because the Ernst (1986) procedure, like the Perkins procedure, only uses information from the selected $I'_k$ in determining the conditional selection probabilities for the new design, it does not generally produce as high an overlap as CCE. However, it takes advantage of LP to overcome the two drawbacks of Perkins' procedure mentioned previously. To illustrate, we present the results of the Ernst (1986) procedure for Example 1. $I'_1, I'_2$ are selected with probabilities $y_1 = .9$ and $y_2 = .1$, respectively. No matter which $I'_k$ is selected, if there is a unit in $I \cap I'_k$, then that unit is the new sample unit. If $I'_2$ is selected and $I \cap I'_2 = \varnothing$, then unit 2 is selected with certainty. If $I'_1$ is selected and $I \cap I'_1 = \varnothing$, then the conditional probabilities of selecting units 1-5 are .022, .044, 0, .733, .2, respectively. This results in an overlap probability of .610. The larger overlap probability for the Ernst (1986) procedure than Perkins' procedure is a result of the following. First, the larger value for $y_1$ with the Ernst (1986) procedure increases the overlap probability, since there is a greater chance of there being a unit in $I \cap I'_1$ than $I \cap I'_2$. Raising $y_1$ above .9 reduces the overlap, however, because $y_1 p_3 = p_3$ for $y_1 = .9$, and hence if $y_1 > .9$, then unit 3 would sometimes not be selected for the new sample when $I'_1$ is selected and unit 3 was in the initial sample. Also, for the Ernst (1986) procedure, whenever $I'_2$ is selected and $I \cap I'_2 = \varnothing$, the new sample unit is in $I'_1$, resulting in some probability of overlap; similarly, when $I'_1$ is selected and $I \cap I'_1 = \varnothing$, the new sample unit is in $I'_2$ with high probability.

When the sample units in the initial design were chosen independently from stratum to stratum, but CCE is still not usable because the TPs are too large, there are two general approaches to overcoming this size problem. The approach taken by Aragon and Pathak (1990), not listed in Table 1, is to use the CCE transportation model, and hence retain optimality, while using an algorithm to reduce the size of the TPs. In the case when the stratifications of the two designs being overlapped are identical, their algorithm reduces the size of each TP to 1/4 it original size. However, when the stratifications are very different, which is when the size problem becomes most acute, the percentage reductions in the size of the TPs become negligible with this approach. Pathak and Fahimi (1992) generalized this approach.

Ernst and Ikeda (E&I) (1995), in contrast, sacrificed optimality for potentially huge reductions in the size of the TPs by using a different TP model. They developed this procedure for, the 1990s redesign of a U.S. Bureau of the Census survey, the Survey of Income and Program Participation (SIPP), with an $n = 2$ design for which an IR procedure was usable since the PSUs were selected independently in the 1980s redesign. E&I is also applicable to other values of *n*. The size of the TP for E&I is $(N+1)N$ for $n = 1$ and $(N(N-1)/2 + N + 1)(N(N-1)/2)$ for $n = 2$, a striking reduction for large *N* from the maximum size of the TPs for CCE.

E&I generally yields a higher overlap than the Ernst (1986) procedure, which is why it was used in the SIPP application, but E&I is more limited in its applicability since it is an IR procedure. The general idea of the E&I procedure for $n = 1$ is as follows. The *N* singleton subsets of *S* are ordered as detailed in the paper. The idea is to place earlier in the ordering those units that have a better chance of being retained in the new sample if they were selected in the initial sample. (The ordering has been developed only for the maximization problem, which is why E&I, unlike the other LP procedures, is listed in Table 1 as only applying to the maximization problem even though the TP used in this procedure is applicable to both the maximization and minimization problems. The procedure can be used with a random ordering of the singleton subsets, but that would generally result in a lower overlap probability.) These *N* single subsets are followed in the ordering by $\varnothing$, resulting in a total of $N+1$ subsets of *S*. Then, instead of obtaining conditional probabilities $P(J = J_i \mid I = I_i)$, the procedure conditions on the first singleton in the ordering whose unit was in the initial sample; if there is no such unit, then the procedure conditions on $I = \varnothing$.

To illustrate, for Example 1 the ordering of the singletons is $\{2\},\{4\},\{1\},\{5\},\{3\}$.

Consequently, if $I = \{3,4\}$, for example, then the procedure would condition on the fact that {4} is the first element in the ordering in *I*, that is the procedure would attempt to retain unit 4, and would not make use of the information that 3 is also in *I*. Furthermore, if $I \neq \varnothing$ for Example 1, then with certainty the new sample unit would be the first element of *I* in the ordering. Thus, the new sample unit would always have been in the initial sample when $I \neq \varnothing$, and, consequently, the overlap when using the E&1 procedure is .700, the same as for CCE.

However, E&I does not always result in a large an overlap as CCE. To see this, consider the following very simple example, designated as Example 2. In this example $N = 2$, $n = 1$, $K = 2$, $p_1 = p_2 = .8$, $p_1 = p_2 = .5$. Then one solution for CCE is $P(J = \{i\} | I = \{i\}) = 1$, $i = 1,2$; $P(J = \{j\} | I = \{1,2\} \text{ or } I = \varnothing) = .5$, $j = 1,2$. Consequently, for Example 2 there is an overlap unless $I = \varnothing$, and hence the probability of overlap for CCE is .96. Now for E&I, because of the symmetry, the probability of overlap for Example 2 is the same whether {1} or {2} appears first in the ordering, so the ordering {1},{2},$\varnothing$ is assumed. Then for E&I, $P(J = \{1\} \mid I = \{1\} \text{ or } I = \{1,2\}) = .625$, $P(J = \{2\} \mid I = \{1\} \text{ or } I = \{1,2\}) = .375$, $P\{J = \{1\} | I = \{2\} \text{ or } I = \varnothing) = 0$, $P\{J = \{2\} | I = \{2\} \text{ or } I = \varnothing) = 1$. Therefore, for Example 2 the overlap probability is .9 for E&I. To understand why E&I produces a lower overlap than CCE for this example, note that since $p_1 < p_1$, we must sometimes have that $1 \in I$ but $J = \{2\}$ for both CCE and E&I. However, for CCE this only occurs when $I = \{1,2\}$, in which case there is an overlap, not when $I = \{1\}$. Since E&I cannot distinguish between $I = \{1,2\}$ and $I = \{1\}$, the conditional probability that $J = 2$ for these two possibilities for *I* must be the same, and hence $P(I = \{1\} \text{ and } J = \{2\}) = .06$ for E&I, which accounts for the deviation from optimality.

The following is outline of E&I for $n = 2$. The $N(N-1)/2$ pairs of units in *S* are ordered as described in the paper, where again the general idea is to place earlier in the ordering those pairs that have a higher probability of being retained. These pairs are followed by the $N$ singleton subsets of *S* in any order, and finally by $\varnothing$. The selection for each of the $N(N-1)/2$ possible $J_j$ is conditioned on the first of the $N(N-1)/2 + N + 1$subsets in the ordering that is contained in *I*. For $n > 2$ the procedure is essentially the same with the *n*-tuples ordered.

The proportion of new sample PSUs overlapped in the 1990s redesign of SIPP with the E&I procedure was .76. This compares to an overlap proportion of .29 with independent selection and an upper bound of .82 with CCE. The actual overlap proportion for CCE could not be computed because the size of the TPs for CCE would have been too large for many of the strata, but an upper bound was calculated using a simple formula described in Ernst and Ikeda (1994).

For the 2000s SIPP redesign it as planned, as of this writing, to use the Ernst (1986) procedure, since E&I cannot be used again. This indicates one approach to overlapping when periodically redesigning a survey, use E&I the first time and subsequently Ernst (1986). An alternative would be to use Ohlsson’s procedure, which is much simpler operationally and preserves the independence of sampling from stratum to stratum. As for the relative performance in terms of expected overlap between the Ohlsson and Ernst (1986) procedures, this author is not aware of any empirical study comparing these methods. However, it is surmised that the Ernst (1986) procedure would typically produce a higher overlap when the designs being overlapped have similar stratifications, since in the case when the stratifications are identical the Ernst (1986) procedure yields an optimal overlap. (In fact, of all the procedures discussed so far, Ohlsson’s is the only one that does not produce an optimal overlap when $n = 1$ and the stratifications are identical. In particular, for this type of design the other procedures always retain a unit in the initial sample if it has a higher selection probability in the new design than in the initial design, but Ohlsson’s procedure may not if the selection probability for another unit in the stratum increases at a faster rate in the new design.) However, when the stratifications are very different and *K* may be large, Ohlsson’s procedure might perform better than the Ernst (1986) procedure since the latter procedure only can use information for the selected $I'_k$, which may be of limited use when *K* is large.

### 3.3 Large Number of Units Per Stratum Procedures

Sequential overlap procedures for large $n$ most commonly have been used for equal probability sampling and have employed PRNs. Most PRN procedures that are applicable to unequal probability sampling, such as Brewer, Early, and Joyce (BEJ) (1972) which uses Poisson sampling, do not guarantee a fixed sample size. The PRN procedures are described in detail in Ohlsson (1995) and are omitted from Table 1, except for Ohlsson's procedure, which has very different properties that have already been discussed.

Among procedures that do not use PRNs, Pollock's (1984) procedure is similar to BEJ in that it is applicable to pps sampling, employs Poisson type sampling, and does not guarantee a fixed sample size. Otherwise, Pollock's procedure has a desirable set of properties including simplicity and optimality. That is, the procedure maximizes (4) subject to (5), but does not satisfy (6). The $p_{j|i}$'s for Pollock's procedure for the maximization problem are quite simply defined, that is:

$$
(11) \quad \begin{aligned} p_{j|i} &= \min\{p_j / p_j, 1\} \text{ if } j \in I_i \\ &= \max\{(p_j - p_j)/(1 - p_j), 0\} \text{ if } j \notin I_i \end{aligned}
$$

It follows from (11) that the expected overlap for Pollock's procedure is equal to (7). In particular, the expected overlap is .880 for Example 1, higher than for CCE. This is because with Pollock's procedure if two units in $S$ were in the initial procedure, they both can be retained, while with CCE they cannot. That is, CCE is optimal among sequential overlap procedures with fixed $n$, while Pollock's procedure is optimal without this restriction. Pollock discusses a modification of his procedure to force a fixed sample size in certain situations. He does not discusses the minimization problem, but the same general approach can be used except that $p_{j|i}$ would be defined by

$$
(12) \quad \begin{aligned} p_{j|i} &= \max\{(p_j + p_j - 1)/ p_j, 0\} \text{ if } j \in I_i \\ &= \min\{p_j /(1 - p_j), 1\} \text{ if } j \notin I_i \end{aligned}
$$

The procedure of Brick, Morganstein and Wolters (BMW) (1987), which is omitted from Table1, is similar to Pollock's, except that BMW only considers equal probability sampling.

Ernst (1995) is a fixed sample size procedure for pps sampling but, unlike the Pollock and BMW procedures, is not optimal. The basic idea of the procedure in the case of the maximization problem is that $p_{j|i}$ is obtained by adding a positive amount $a_{ij}$ to $p_i$ for each $j \in I_i$ and subtracting a positive amount $b_i$ for each unit in $S$. The set of $a_{ij}$, $b_i$ must be such that $0 \le p_{j|i} \le 1$ for all $i, j$, and (5), (6) are satisfied. For the minimization problem, the only difference is that $a_{ij}$ is added to $p_i$ for each $I_i$ $a_{ij}$'s are chosen to be as large as possible. Further details are omitted here. The overlap for Example 1 with the Ernst

## 4. Simultaneous Procedures

Finally, we consider procedures developed for simultaneous selection. Mitra and Pathak's

surveys, although it is optimal only for the overlap of two or three surveys. Perry, Burt and Iwig's (PBI) (1993) procedure has fewer restrictions in applicability than M&P but, unlike the other

algorithms for either of these procedures.

Ernst (1996, 1998) developed optimal simultaneous procedures for two different situations.

In Ernst (1996), $n = 1$ but the designs may have different stratifications. In Ernst (1998) there are no restrictions on *n* but the stratifications must be identical. These two procedures employ the algorithm in Causey, Cox, and Ernst (1985) for solving the two-dimensional version of the controlled selection problem developed by Goodman and Kish (1950). This algorithm involves solving a sequence of TPs. (The result in Ernst (1996) had been obtained earlier by Pruhs (1989) using a much more complex graph theory approach.)

The basic idea of the algorithm for the Ernst (1996) procedure is as follows. A two-dimensional controlled selection problem is a two-dimensional tabular array **A,** that is, an array in which all the elements except those in the final row and column correspond to the internal elements of an additive table. The elements in the final row and those in the in the final column are marginal values of the array; in particular the element in the lower right hand corner of the array is the grand total. Let $D_1, D_2$ denote the two designs, with $T_1, T_2$ the number of strata in $D_1, D_2$, respectively. In Ernst (1996) a single controlled selection problem is used to select the sample for both designs for all strata together. **A** is a $(T_1 + 2) \times (T_2 + 2)$ tabular array. For the maximization problem, the value, $a_{ij}$, of cell $(i, j)$, $i = 1,...,T_1$, $j = 1,...,T$ , is obtained by letting $p_{ijk1}, p_{ijk2}$ denote the $D_1, D_2$ selection probabilities, respectively, for the *k*-th unit that is in both $D_1$ stratum *i* and $D_2$ stratum *j*, and then summing $\min\{p_{ijk1}, p_{ijk2}\}$ over all *k*. In addition, let

$$a_{i(T_2+2)} = 1, \ i = 1,...,T_1; \quad a_{(T_1+2)j} = 1, \ j = 1,...,T_2; \quad a_{(T_1+1)(T_2+1)} = 0$$

and let the remaining elements, which are in rows $T_1 + 1$, column $T_2 + 1$, and the grand total cell, be defined so that the additivity constraints of a tabular array are satisfied.

A solution to a controlled selection problem consists of a sequence of integer-valued, tabular arrays, $\mathbf{M}_l = (m_{ijl})$, $l = 1,....,r$ , of the same dimensions as **A,** and associated probabilities $p_l^*$ satisfying:

$$|m_{ijl} - a_{ij}| < 1 \ \text{for all} \ i, j, l; \quad \sum_{l=1}^{r} p_l^* = 1; \quad \sum_{l=1}^{r} p_l^* m_{ijl} = a_{ij} \ \text{for all} \ i, j$$

Such a solution can be obtained by solving a sequence of TPs of the same dimensions as **A**, as described in Causey, Cox, and Ernst (1985). A particular $\mathbf{M}_l$ is chosen using the associated probabilities. The internal elements of the array are all 0 or 1 with a single 1 in each of the first $T_1$ rows and each of the first $T_2$ columns. A 1 in cell $(i, j)$, $i = 1,...,T_1$, $j = 1,...,T_2$, indicates a unit which is in $D_1$ stratum *i* and $D_2$ stratum *j* is to be selected to be in both samples, in which case the particular unit is selected from among all such units with probability proportional to $\min\{p_{ijk1}, p_{ijk2}\}$. A 1 in cell $(i, T_2 + 1)$, $i = 1,...,T_1$, indicates that the unit to be selected in $D_1$ stratum *i* is not in the $D_2$ sample, with the particular unit selected from among all units in $D_1$ stratum *i* with probability proportional to $p_{ijk1} - \min\{p_{ijk1}, p_{ijk2}\}$. Similarly, a 1 in cell $(T_1 + 1, j)$, $j = 1,...,T_2$, indicates that the unit to be selected in $D_2$ stratum *j* is not in the $D_1$ sample, with the particular unit selected from all units in $D_2$ stratum *j* with probability proportional to $p_{ijk2} - \min\{p_{ijk1}, p_{ijk2}\}$. The Ernst (1996) procedure for the minimization problem is identical to that for the maximization problem except $\min\{p_{ijk1}, p_{ijk2}\}$ is replaced everywhere by $\max\{p_{ijk1} + p_{ijk2} - 1, 0\}$.

For the Ernst (1998) procedure, in contrast to the Ernst (1996) procedure, there is a separate controlled selection problem **A** for each stratum *S*, with **A** an $(N + 1) \times 5$ array. Each of the first *N* rows corresponds to a unit in *S*. Let $p_{i1}, p_{i2}$ denote the $D_1, D_2$ selection probabilities, respectively, for unit *i*. Then for the maximization problem let

$$a_{i3} = \min\{p_{i1}, p_{i2}\}\,; \quad a_{ij} = p_{ij} - a_{i3},\ \ j = 1,2\,; \quad a_{i5} = 1$$

with the remaining elements of **A**, which are in row $N+1$ or column 4, defined so that the additivity constraints of a tabular array are satisfied. The selected array arising from the solution to the controlled selection problem has a single 1 and three 0's in the first four columns of the first *N* rows. If for row *i* the 1 is in the first column, then unit *i* is in the $D_1$ sample only; if the 1 is in the second column the unit is in the $D_2$ sample only; if it is in the third column then the unit is in both samples; and if it is in the fourth column the unit is in neither sample. For the minimization problem the only change is that $a_{i3} = \max\{p_{i1} + p_{i2} - 1, 0\}$.

Although both the Ernst (1996) and (1998) procedures require the solution of a sequence of TPs, the dimensions of these TPs, given above, are typically not unreasonably large, so the procedures tend to be computationally efficient. Also both procedures have the desirable property that if the $D_1$ selection probability for each unit does not exceed the $D_2$ selection probability, then the $D_1$ sample is a subsample of the $D_2$ sample. In addition, the expected overlap for both procedures satisfy (7) for the maximization problem and (8) for the minimization problem, and hence both procedures are optimal. In particular, the overlap for Example 1 for the Ernst (1996) procedure is .880, greater than for CCE and the same as for Pollock's procedure, even though the Ernst (1996) procedure has the additional restriction of a fixed sample size. Ernst (1996) accomplishes its greater overlap than CCE in a different way than Pollock's procedure. For Example 1, with the initial design designated as $D_1$ and the new design as $D_2$, Ernst (1996) takes advantage of the simultaneous selection to increase the overlap by reducing the probability that two units in *S* are in the $D_1$ sample, since only one of them could be in the $D_2$ sample, while increasing the probability that at least one unit is in the $D_1$ sample. CCE, in contrast, since it is a sequential procedure, has no control over the $D_1$ sample. Thus CCE is an optimal sequential, fixed sample size procedure, while Ernst (1996) and (1998) are optimal simultaneous, fixed sample size procedures.

A problem for future research is to develop, if possible, an optimal procedure which combines the features of the Ernst (1996) and Ernst (1998) procedures, that is, which can be used when the stratifications in the two designs are different and for any *n*.

Finally, we present here for the first time a procedure, that we have designated as Ernst (1999), which preserves prespecified joint selection probabilities in each stratum in each design and is applicable to the overlap of more than two surveys. It is essentially a generalization of CCE to more than two surveys, but like the M&P generalization of Keyfitz's procedure, it is only applicable to simultaneous selection, and, additionally, only when the stratifications for all the designs are identical. To simplify notation we present it in terms of three designs, $D_1, D_2, D_3$, but the same approach holds for any number of designs. For a stratum *S*, let $p'_{il}$ denote the probability of selecting the *i*-th possible sample for $D_l$, $l = 1,2,3$. Let $x_{ijk}$ denote the joint probability of selecting the *i*-th possible $D_1$ sample, the *j*-th possible $D_2$ sample, and the *k*-th possible $D_3$ sample from *S*. Let $n_{ijk}$ be obtained by multiplying each unit that is in at least one of these three samples by one less than the number of samples that it is in, and summing the product over all such units. (Other values for $n_{ijk}$ are possible.) Then the LP problem is to maximize or minimize

$$\sum_{i,j,k} n_{ijk} x_{ijk} \ \text{ subject to the constraints } \sum_{j,k} x_{ijk} = p'_{i1}\,, \quad \sum_{i,k} x_{ijk} = p'_{j2}\,, \quad \sum_{i,j} x_{ijk} = p'_{k3}$$

**REFERENCES**

Brewer, K. R. W., Early, L. J., and Joyce, S. F. (1972). Selecting Several Samples from a Single Population. Australian Journal of Statistics, 14, 231-239.

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, 787-791.

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, 3rd ed. New York: John Wiley.

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

Ernst, L. R. (1995). Maximizing and Minimizing Overlap of Ultimate Sampling Units. Proceedings of the Section on Survey Research Methods, American Statistical Association, 706-711.

Ernst. L. R. (1996). Maximizing the Overlap of Sample Units for Two Designs with Simultaneous Selection. Journal of Official Statistics, 12, 33-45.

Ernst, L. R. (1998). Maximizing and Minimizing Overlap When Selecting a Large Number of Units per Stratum Simultaneously for Two Designs. Journal of Official Statistics, 14, 297-314.

Ernst, L. R. and Ikeda, M. M. (1994). A Reduced-Size Transportation Algorithm for Maximizing the Overlap Between Surveys. Bureau of the Census, Statistical Research Division, Research Report Series, No. RR-93/02.

Ernst, L. R. and Ikeda, M. M. (1995). A Reduced-Size Transportation Algorithm for Maximizing the Overlap Between Surveys. Survey Methodology, 21, 147-157.

Fellegi, I. P. (1966). Changing the Probabilities of Selection When Two Units Are Selected with PPS Without Replacement. Proceedings of the Social Statistics Section, American Statistical Association, 434-442.

Goodman, R. and Kish, L. (1950), Controlled Selection—A Technique in Probability Sampling, Journal of the American Statistical Association, 45, 350-372.

Keyfitz, N. (1951). Sampling With Probabilities Proportionate to Size: Adjustment for Changes in Probabilities. Journal of the American Statistical Association, 46, 105-109.

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

Mitra, S. K., and Pathak, P. K. (1984). Algorithms for Optimal Integration of Two or Three Surveys. Scandinavian Journal of Statistics, 11, 257-263.

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, 153-169.

Ohlsson, E. (1996). Methods for PPS Size One Sample Coordination. Institute of Actuarial Mathematics and Mathematical Statistics, Stockholm University, No. 194.

Ohlsson, E. (1999). Comparison of PRN Techniques for Small Sample Size PPS Sample Coordination. Institute of Actuarial Mathematics and Mathematical Statistics, Stockholm University, No. 210.

Pathak, P. K., and Fahimi, M. (1992). Optimal integration of surveys. In Essays in Honor of D. Basu, eds. M. Ghosh, and P. K. Pathak. Hayward, California: Institute of Mathematical Statistics, 208-224.

Perkins, W. M. (1970). 1970 CPS Redesign: Proposed Method for Deriving Sample PSU Selection Probabilities Within 1970 NSR Strata. Memorandum to Joseph Waksberg, U.S. Bureau of the Census.

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, 345-351.

Pollock, J. (1984). PPES subsampling of Two Subdomains with Independent Probabilities. Proceedings of the Section on Survey Research Methods, American Statistical Association, 223-228.

Pruhs, K. (1989). The Computational Complexity of Some Survey Overlap Problems. Proceedings of the Section on Survey Research Methods, American Statistical Association, 747-752.

Raj, D. (1956). On the Method of Overlapping Maps in Sample Surveys. Sankhya, 17, 89-98.

Schrijver, A. (1986). Theory of Linear and Integer Programming. New York: John Wiley.

Sunter, A. B. (1989). Updating Size Measures in a PPSWOR Design. Survey Methodology, 15, 253-260.

The views expressed in this paper are attributable to the author and do not necessarily reflect those of the Bureau of Labor Statistics.

**RÉSUMÉ**

*Beaucoup de procédures ont été développées pour maximiser ou réduire au minimum la superposition des unités de prélèvement. Les propriétés des diverses procédures sont récapitulées. Ces propriétés incluent si le procédé est pour des sélections séquentielles ou simultanées des échantillons, le nombre d'unités des échantillons par strate, si les strates dans les conceptions superposées peuvent différer, et si le procédé est applicable quand l'échantillon initial n'a pas été choisi indépendamment d'une strate à l'autre.*