UNIVERSITY OF Southampton
Southampton Statistical Sciences Research Institute

Working Paper M09/07
Methodology

Optimal Sample Coordination Using Controlled Selection

Alina Matei, Chris Skinner

Abstract

Sample coordination maximizes or minimizes the overlap of samples selected from overlapping populations. We propose a coordination method for the case where units are to be selected with maximum overlap using two designs with given unit inclusion probabilities. Our method is based on some theoretical conditions on the joint selection probability of two samples and on the application of a controlled selection method, implemented with linear programming.

Optimal sample coordination using controlled selection

Alina Matei * Chris Skinner †

Abstract

Sample coordination maximizes or minimizes the overlap of two or more samples selected from overlapping populations. It can be applied to designs with simultaneous or sequential selection of samples. We propose a method for sample coordination in the former case. We consider the case where units are to be selected with maximum overlap using two designs with given unit inclusion probabilities. The degree of coordination is measured by the expected sample overlap, which is bounded above by a theoretical bound, called the absolute upper bound, and which depends on the unit inclusion probabilities. If the expected overlap equals the absolute upper bound, the sample coordination is maximal. Most of the methods given in the literature consider fixed marginal sampling designs, but in many cases, the absolute upper bound is not achieved. We propose to construct optimal sampling designs for given unit inclusion probabilities in order to realize maximal coordination. Our method is based on some theoretical conditions on joint selection probability of two samples and on the controlled selection method with linear programming implementation. The method can also be applied to minimize the sample overlap.

Key words: sample surveys, sample coordination, simultaneous selection of samples, joint selection probability of two samples, linear programming.

1 Introduction

There are many sample survey applications where it is desirable to coordinate the selection of different samples.The maximization of overlap between

*Institute of Statistics, University of Neuchâtel, Pierre à Mazel 7, 2000, Neuchâtel, Switzerland, alina.matei@unine.ch

†Southampton Statistical Sciences Research Institute, University of Southampton, Southampton SO17 1BJ, U.K., C.J.Skinner@soton.ac.uk

samples is referred to as positive coordination, whereas negative coordination refers to overlap minimization. One class of applications where positive coordination may be desirable is in repeated surveys, where samples are selected on successive occasions and sample overlap helps to improve precision in the estimation of change. Another advantage to positive coordination in either repeated surveys or surveys conducted simultaneously is the potential reduction in data collection costs. See Ernst (1999) for an example of the simultaneous selection of samples with overlap maximization in the U.S. Bureau of Labor Statistics: the Economic Cost Index and the Occupational Compensation Surveys Program. One important reason why negative coordination may be desirable to a survey-taking organisation is in order to reduce the burden on sample units. See Perry et al. (1993) for an example of the selection of samples with overlap minimization in the National Agricultural Statistics Service: the Farm Costs and Returns Survey and the Agriculture Survey and the Labor Survey.

The maximization or minimization of overlap between samples may often have to be offset by other requirements on the designs, which rule out the extreme options of complete or no overlap. For example, the inclusion probabilities of different units in each sample may be pre-specified and may differ between samples, thus preventing the selection of a single common sample. Thus, the typical design problem faced in positive or negative coordination is one of constrained optimization: how to construct a scheme for selecting the samples which maximizes/minimizes (expected) overlap, subject to certain constraints on how each of the samples is selected.

There is a long-established and large literature on sampling methods to achieve positive or negative coordination. See, for example, Ernst (1999); Ohlsson (1995); Mach et al. (2006) and the references therein. Many of the methods proposed in the literature do not produce optimal overlap, but do lead to more coordination than would be provided by the independent selection of samples. One approach, which is based on clearly defined optimization criteria, is formulating the problem as a linear programming task known as the transportation problem (see Causey et al., 1985). In this approach the probabilities assigned to each of the possible pairs of samples are treated as unknowns and constraints on the sampling schemes as well as the objective function (the expected overlap) can be expressed linearly. The problem with this approach, however, is that the number of possible samples can be very large and hence the computational requirements can be prohibitive. Some approaches to reducing the amount of computation when solving this transportation problem have been proposed. See Aragon and Pathak (1990), Pathak and Fahimi (1992), Ernst and Ikeda (1995), Mach et al. (2006).

Matei and Tillé (2005) developed an alternative approach, which is based upon the same optimization criteria as the transportation problem, but employs a quite different (and less intensive) computational method, applying a simple iterative proportional fitting (IPF) algorithm to an initial feasible solution. They demonstrated that under certain conditions, their method provides a solution to the same optimization problem addressed by the transportation approach. The optimal solution, referred to as 'maximal sample coordination', involves maximizing the expected overlap between two samples, subject to the 'marginal' sampling designs for each sample being fixed.

A problem with the method of Matei and Tillé (2005), however, is that it does not always generate a solution which meets the theoretical conditions required for optimality. The aim of this paper is to extend their approach to address this problem, by use of the method of controlled selection, as set out in Rao and Nigam (1990). We take into account the positive coordination, but the negative case can be treated similarly.

Our extension of the method of Matei and Tillé (2005) does, however, involve a reformulation of the constraints in the optimization problem. Instead of requiring the 'marginal' sampling designs for each sample to be fixed, we only require that the first-order inclusion probabilities are fixed. By reducing the constraints on the problem, we seek to use controlled selection to reduce the probabilities (preferably to zero) of samples which would lead to the theoretical conditions being breached. Such samples will be called 'non-preferred' samples in the terminology of controlled selection.

In summary, the aim of the paper is to extend the method of Matei and Tillé (2005) using the method of controlled selection in order to obtain coordinated designs which meet specified optimality criteria, in situations where the former method does not generate an optimal solution.

The paper is organized as following. In Section 2, the formal framework and the design problem are set out. Section 3 reviews conditions needed to realize maximal sample coordination. Section 4 shows the controlled selection method may be applied to sample coordination. In Sections 5 and

6, our proposed approach is presented and exemplified. Finally, Section 7 provides a discussion of the practical implications of the paper. A review of the transportation problem approach is given in the Appendix.

2 Framework and design problem

Consider the selection of two samples, each from a separate population. Let U^1 and U^2 denote the sets of labels of units in the populations from which the two samples are drawn. Let s^1 and s^2 denote the sets of labels of the two corresponding samples so that $s^1 \subset U^1$ and $s^2 \subset U^2$. Let $U = U^1 \cup U^2$ and, without loss of generality write $U = \{1, \dots, k, \dots, N\}$. In many applications the two populations might be identical so that $U = U^1 = U^2$. The sets of possible samples s^1 and s^2 are denoted $\mathcal{S}^1$ and $\mathcal{S}^2$, respectively. Let $m = |\mathcal{S}^1|$ and $q = |\mathcal{S}^2|$, where $|\ s\ |$ denotes the cardinality of a general set s, and write $\mathcal{S}^1 = \{s^1_1, ..., s^1_m\}$ and $\mathcal{S}^2 = \{s^2_1, ..., s^2_q\}$. The pair of samples $s_{ij} = (s^1_i, s^2_j)$ is referred to as the bi-sample. The set of all possible bi-samples is denoted $\mathcal{S} = \{s_{ij} | s_{ij} = (s^1_i, s^2_j), s^1_i \in \mathcal{S}^1, s^2_j \in \mathcal{S}^2, i = 1, \dots, m, j = 1, \dots q\}$. The overall sampling design is represented by the probability that bi-sample s_{ij} is selected, denoted $p_{ij} = p(s^1_i, s^2_j) = p(s_{ij})$ for $s_{ij} \in \mathcal{S}$. The resulting matrix $\mathbf{P} = (p_{ij})$ of dimension $m \times q$ is used to denote the overall design. The 'marginal' sampling designs for s^1 and s^2 may be derived from the joint probabilities and are represented by the probability $p^1(s^1_i)$ that sample $s^1_i \in \mathcal{S}^1$ is selected and the probability $p^2(s^2_j)$ that sample $s^2_j \in \mathcal{S}^2$ is selected, where $\sum_{s^1_i \in \mathcal{S}^1} p^1(s^1_i) = 1$ and $\sum_{s^2 \in \mathcal{S}^2} p^2(s^2_j) = 1$. We have $\sum_{j=1}^{q} p_{ij} = p^1(s^1_i)$, and $\sum_{i=1}^{m} p_{ij} = p^2(s^1_j)$. The marginal sampling design for s^1 is represented by the $m \times 1$ vector $\mathbf{p}^1$ which contains the values $p^1(s^1_i)$ for $s^1_i \in \mathcal{S}^1$. Similarly, the second marginal sampling design is represented by the $q \times 1$ vector $\mathbf{p}^2$ which contains the values $p^2(s^2_i)$ for $s^2_i \in \mathcal{S}^2$. The overall sampling design is said to be coordinated if $p(s^1_i, s^2_j) \neq p^1(s^1_i) p^2(s^2_j)$ (see Cotton and Hesse, 1992; Mach et al., 2006), i.e. if the two samples are not selected independently.

The size of the overlap between two samples is denoted $c_{ij} = |s^1_i \cap s^2_j|$ and, in general, this is random. We therefore measure the degree of coordination by the expected sample overlap, given by:

$$E(c_{ij}) = \sum_{i=1}^{m} \sum_{j=1}^{q} c_{ij} p_{ij} = \sum_{k \in U} \pi_k^{1,2},$$

where

$$\pi_k^{1,2} = \sum_{\substack{s_{ij} = (s_i^1 \ni k, s_j^2 \ni k) \\ s_{ij} \in \mathcal{S}}} p_{ij}$$

is the probability that unit $k \in U$ is included in both samples.

To obtain an upper bound on this expected overlap, let

$$\pi_k^1 = \sum_{\substack{s_i^1 \ni k \\ s_i^1 \in \mathcal{S}^1}} p^1(s_i^1)$$

be the first-order inclusion probability of unit $k \in U$ for the first design and let

$$\pi_k^2 = \sum_{\substack{s_j^2 \ni k \\ s_j^2 \in \mathcal{S}^2}} p^2(s_j^2)$$

be the first-order inclusion probability of unit $k \in U$ for the second design. If $k \in U^1 \setminus U^2, \pi_k^2 = 0$ and if $k \in U^2 \setminus U^1, \pi_k^1 = 0$. Since $\pi_k^{1,2} \leq \min(\pi_k^1, \pi_k^2)$, the expected sample overlap is bounded above by

$$E(c_{ij}) \leq \sum_{k \in U} \min(\pi_k^1, \pi_k^2). \tag{1}$$

Matei and Tillé (2005) call $\sum_{k \in U} \min(\pi_k^1, \pi_k^2)$ the Absolute Upper Bound (AUB) and say that *maximal sample coordination* occurs when equality holds in (1).

The general design problem of interest is how to construct a sampling design $\mathbf{P}$ such that maximal sample coordination holds. This construction will be under certain constraints. The standard constraints considered in the transportation problem and in the method of Matei and Tillé (2005) is that the marginal designs $\mathbf{p}^1$ and $\mathbf{p}^2$ are given (and thus fixed). We shall also consider the weaker constraint that only the marginal inclusion probabilities,

$\pi_k^1, \pi_k^2, k \in U$, are given. We denote the vectors of these values by $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ respectively.

One approach to maximization of the expected overlap $\sum_{i=1}^m \sum_{j=1}^q c_{ij} p_{ij}$ is to use linear programming, treating the p_{ij} as the unknowns. This linear programming problem, under the constraint that the marginal sampling designs are given, is in the form of the transportation problem (see Causey et al., 1985, and the Appendix for a description).

An alternative approach is the method set out in Matei and Tillé (2005). To specify this method, we first consider conditions under which maximal coordination occurs. These conditions are fundamental to the specification of our proposed method.

3 Conditions for maximal sample coordination

Matei and Tillé (2005) discuss conditions under which maximal sample coordination occurs, i.e.

$$E(c_{ij}) = \sum_{k \in U} \min(\pi_k^1, \pi_k^2). \tag{2}$$

Given arbitrary marginal sampling designs, $\mathbf{p}^1$ and $\mathbf{p}^2$, the AUB cannot always be achieved (see Ohlsson, 2000, for some examples using different methods for sample coordination). Maximal sample coordination for unequal probability designs can be achieved for Poisson sampling with permanent random numbers (Brewer et al., 1972, 1984) for each design or for Keyfitz's method (Keyfitz, 1951). These two methods suffer, however, from some drawbacks: the random sample size for the former, and the limitation to one unit drawn per stratum for the latter. For stratified simple random sampling without replacement (srswor), the sequential srswor with permanent random numbers (Ohlsson, 1995) can yield maximal expected overlap for positive coordination and does not suffer from the drawbacks mentioned above. The method of Mach et al. (2006), based on a reduced transportation problem, yields optimal expected overlap for positive coordination of several surveys and negative coordination of two surveys with stratified srswor designs, subject to constraints on marginal sampling designs.

For two fixed (but arbitrary) marginal sampling designs, theoretical conditions to achieve the AUB are given in Matei and Tillé (2005) and are summarised in Proposition 1.

Proposition 1 *Let (p_{ij}) denote an arbitrary sampling design for which the marginal designs $\mathbf{p}^1$ and $\mathbf{p}^2$ are given. Let $I = \{k \in U | \pi_k^1 \leq \pi_k^2\}$ be the set of 'increasing' units, and let $D = \{k \in U | \pi_k^1 > \pi_k^2\}$ be the set of 'decreasing' units, with $U = I \cup D$, and $I \cap D = \emptyset$. The AUB is achieved by the design iff the following two relations are fulfilled:*

a) if $(s_i^1 \backslash s_j^2) \cap I \neq \varnothing$ then $p_{ij} = 0$,

b) if $(s_j^2 \backslash s_i^1) \cap D \neq \varnothing$ then $p_{ij} = 0$,

for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, q\}$.

Matei and Tillé (2005) proposed the use of Iterative Proportional Fitting (IPF) (Deming and Stephan, 1940) to obtain a design p_{ij} for which conditions *a)* and *b)* of Proposition 1 are satisfied, subject to given marginal designs $\mathbf{p}^1$ and $\mathbf{p}^2$. We now outline this method.

The values p_{ij} are obtained using an iterative process. To simplify notation, however, we do not use subscripts for the iteration steps. Let $\mathbf{P} = (p_{ij})$ be initially any matrix of dimension $m \times q$, for which the implied marginal designs correspond to the given values $\mathbf{p}^1$ and $\mathbf{p}^2$. For example, p_{ij} might be obtained from applying $\mathbf{p}^1$ and $\mathbf{p}^2$ independently, i.e. $p_{ij} = p^1(s_i^1)p^2(s_j^2)$.

Then modify $\mathbf{P}$ by assigning zero values to elements p_{ij} of $\mathbf{P}$ implied by conditions *a)* and *b)* of Proposition 1. More precisely, let $\mathbf{P}^0$ denote the matrix of dimension $m \times q$, with elements ${p^0}_{ij}$, where

$${p^0}_{ij} = 0, \text{ if } (s_i^1 \backslash s_j^2) \cap I \neq \varnothing, \text{ or } (s_j^2 \backslash s_i^1) \cap D \neq \varnothing,$$

and ${p^0}_{ij} = p_{ij}$, otherwise, and modify $\mathbf{P}$ by replacing it by $\mathbf{P}^0$. Note that the transformation from $\mathbf{P}$ to $\mathbf{P}^0$ depends upon the sets I and D, which themselves depend upon the inclusion probabilities $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$. These are taken here to be fixed but we could be more explicit about this dependence by writing $\mathbf{P}^0 = \{\mathbf{P}; \boldsymbol{\pi}^1, \boldsymbol{\pi}^2\}^0$.

As a result of this transformation, the row and column totals of $\mathbf{P}^0$ will differ, in general, from the given values $\mathbf{p}^1$ and $\mathbf{p}^2$, respectively, so the constraints on the joint probabilities p_{ij} will no longer be respected, i.e.

$\sum_{j=1}^{q} {p^0}_{ij} \neq p^1(s_i^1)$, $\sum_{i=1}^{m} {p^0}_{ij} \neq p^2(s_j^2)$. To ensure that these constraints are respected, the non-zero values of $\mathbf{P}^0$ are modified using the IPF procedure. The IPF procedure iteratively modifies the matrix $\mathbf{P}^0$ and is applied until convergence is reached. The final matrix $\mathbf{P}$ has the property that $\sum_{i=1}^{m} \sum_{j=1}^{q} c_{ij} p_{ij}$ equals the AUB, if the latter can be achieved, and the marginal constraints are respected.

Remark 1 *If, however, there exist either row(s) i' and/or column(s) j' of the matrix $\mathbf{P}^0$ which consist entirely of zeros (so that $\sum_{j=1}^{q} {p^0}_{i'j} = \sum_{i=1}^{m} {p^0}_{ij'} = 0$) then the AUB cannot be achieved by the above IPF procedure since the corresponding row or column sums must all be strictly positive.*

Hence, the procedure of Matei and Tillé (2005) cannot be guaranteed to construct a solution which achieves the AUB. Sub-optimal solutions (when the AUB is not achieved) can be obtained by replacing zeros on the row(s) i' and/or columns j' with a small quantity, say ε, defined as 10^{-6}, to assure the marginal constraints and the convergence of the IPF procedure. Our goal is not, however, to obtain sub-optimal solutions. One situation when the IPF procedure does provide theoretically an optimal solution is when the samples are completely overlapping, since there is then at least one non-zero value on each row and column of $\mathbf{P}^0$ ($s_i^1 \setminus s_j^2 = s_j^2 \setminus s_i^1 = \emptyset$ if $s_i^1 = s_j^2$ and $p_{ij} \neq 0$). Another situation when the AUB is achieved is as follows.

Remark 2 *Consider what happens when the conditions $(s_i^1 \backslash s_j^2) \cap I \neq \varnothing$ or $(s_j^2 \backslash s_i^1) \cap D \neq \varnothing$ in Proposition 1 are not fulfilled for all pairs (i, j) and it is not possible to set p_{ij} to zero. If the following condition is satisfied*

$$(s_i^1 \backslash s_j^2) \cap I = \varnothing, \text{ for all } i = 1, \ldots, m, \text{ and for all } j = 1, \ldots, q, \tag{3}$$

then, if $k \in (s_i^1 \backslash s_j^2)$ we have $k \in D$ and $\pi_k^{1,2} = \sum_{\substack{s_{ij}=(s_i^1 \ni k, s_j^2 \ni k) \\ s_{ij} \in \mathcal{S}}} p_{ij} = \pi_k^2 - \sum_{\substack{s_{ij}=(s_i^1 \ni k, s_j^2 \not\ni k) \\ s_{ij} \in \mathcal{S}}} p_{ij}$. Since we have supposed $(s_j^2 \backslash s_i^1) \cap D = \varnothing$, it follows that $\pi_k^{1,2} = \pi_k^2 = \min(\pi_k^1, \pi_k^2)$. Similarly, if the following condition is satisfied

$$(s_j^2 \backslash s_i^1) \cap D = \varnothing, \text{ for all } j = 1, \ldots, q, \text{ and for all } i = 1, \ldots, m, \tag{4}$$

then $\pi_k^{1,2} = \pi_k^1 = \min(\pi_k^1, \pi_k^2)$.

Thus, if conditions (3) and (4) are simultaneously satisfied, maximal sample coordination is possible since $\pi_k^{1,2} = \min(\pi_k^1, \pi_k^2)$, for all $k \in U$. In this case, we can obtain a joint probability p_{ij}, using the IPF procedure directly with the marginal distributions $\mathbf{p}^1$ and $\mathbf{p}^2$, without setting any values to zero, and the AUB is always achieved.

4 Maximal sample coordination using controlled selection

As mentioned in Remark 1, there are cases where the application of the Matei and Tillé (2005) procedure leads to some rows and/or some columns in $\mathbf{P}$ which consist entirely of zeros so that the constraints on the marginal sample probabilities are violated. To address this problem, we propose the use of controlled selection. In order to achieve the AUB we shall consider relaxing the constraints used in the method of Matei and Tillé (2005) that the marginal designs $\mathbf{p}^1$ and $\mathbf{p}^2$ are fixed. Instead, we shall consider methods which only require that the inclusion probabilities $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ for each of the marginal designs are fixed.

As discussed in Rao and Nigam (1990), the method of controlled selection begins by classifying all possible samples as either 'preferred' or 'nonpreferred'. The method then takes an initial sampling design $\widetilde{\mathbf{p}}$, and then defines a new design $\widetilde{\mathbf{p}}_*$ which selects 'nonpreferred' samples s with probability $\widetilde{p}_*(s) \leq \widetilde{p}(s)$, while maintaining the assigned inclusion probability of each unit in the population. The resulting sampling design $\widetilde{\mathbf{p}}_*$ is called 'controlled', while the initial one is called 'uncontrolled'. In what follows the subscript '$_*$' will denote a controlled design.

Consider the coordination of two designs with simultaneous selection and fixed inclusion probabilities $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$. Suppose that any initial design $\mathbf{P}$ is used for which the marginal designs imply these given inclusion probabilities. And suppose that zeros are assigned as discussed in the previous section resulting in $\mathbf{P}^0$. Suppose that this matrix has some row(s) i' (with corresponding sample denoted $s_{i'}^1$) and/or columns j' (with the corresponding sample denoted $s_{j'}^2$) with only zero values.

For illustration, consider the following simple example: suppose that two samples of size two are to be drawn from U^1 and U^2, respectively, where $U^1 = \{1,2,3\}$ and $U^2 = \{1,2,3,4\}$. The inclusion probabilities $\boldsymbol{\pi}^1$ and

$\boldsymbol{\pi}^2$ determine the sets I and D and suppose these inclusion probabilities are fixed to take values which imply that $I = \{1, 2, 4\}$ and $D = \{3\}$. All possible samples for the first marginal design are listed on the rows of $\mathbf{P}^0$, displayed below in Expression (5). Similarly, all possible samples for the second marginal design are listed on the columns of $\mathbf{P}^0$. The non-zero values of the matrix $\mathbf{P}^0$ are denoted x. Note that $j' \in \{6\}$, with $s_6^2 = \{3, 4\}$, and there are no i' rows in $\mathbf{P}^0$.

$$\mathbf{P}^0 = \left(\begin{array}{c|cccccc} & \{1,2\} & \{1,3\} & \{1,4\} & \{2,3\} & \{2,4\} & \{3,4\} \\ \hline \{1,2\} & x & 0 & 0 & 0 & 0 & 0 \\ \{1,3\} & x & x & x & 0 & 0 & 0 \\ \{2,3\} & x & 0 & 0 & x & x & 0 \end{array} \right) \tag{5}$$

In general, the AUB cannot be achieved using the method of Matei and Tillé (2005) if there exist samples of the form $s_{i'}^1$ and $s_{j'}^2$. For the controlled selection method, these define the 'nonpreferred' samples. We now introduce controlled selection algorithms to reduce the probability of these samples arising. If the probabilities can be reduced to zero, then the AUB can be achieved, whilst preserving the fixed values of $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$.

We introduce two linear programming algorithms designed to reduce the probabilities of the 'nonpreferred' samples. In each case, the algorithm begins with a matrix $\mathbf{P}^0$ constructed from some feasible initial design $\mathbf{P}$ which implies the given inclusion probabilities $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$. The first algorithm will be a plug-in application of the method of Rao and Nigam (1990) in our context. It provides a solution (a 'controlled' design) for each of the two 'uncontrolled' plans, $\mathbf{p}^1$ and $\mathbf{p}^2$. The second algorithm will be defined on the set of bi-samples. The two initial sampling designs are not involved directly. Here the 'uncontrolled' plan is the initial joint sampling design, which is not explicitly defined, and has as marginal distributions the two plans, $\mathbf{p}^1$ and $\mathbf{p}^2$.

To formulate the first algorithm, we define the sets of 'nonpreferred' samples for $\mathbf{P}^0$ as: $\mathcal{S}_*^1 = \{s_i^1 \in \mathcal{S}^1 \,|\, p_{ij}^0 = 0, \text{ for all } j = 1, \ldots, q\}$ and $\mathcal{S}_*^2 = \{s_j^2 \in \mathcal{S}^2 \,|\, p_{ij}^0 = 0, \text{ for all } i = 1, \ldots, m\}$. The algorithm for marginal design $t = 1, 2$ is defined as the linear programming solution to the following problem

$$\min_{\mathbf{p}_*^t} \sum_{s_\ell^t \in \mathcal{S}_*^t} p_*^t(s_\ell^t), \tag{6}$$

subject to

$$\left| \begin{array}{l} \sum_{\substack{s_\ell^t \ni k \\ s_\ell^t \in \mathcal{S}^t}} p_*^t(s_\ell^t) = \pi_k^t,\ k \in U, \\ \sum_{\ell=1}^r p_*^t(s_\ell^t) = 1, \\ p_*^t(s_\ell^t) \geq 0,\ \ell = 1, \ldots, r, \end{array} \right.$$

where $r = m$ if $t = 1$ and $r = q$ if $t = 2$.

For $t = 1$, the algorithm defined by (6) may be used to reduce the probability of selecting samples $s_{i'}^1$ for which $p_{i'j}^0 = 0$, for all $j = 1, \ldots, q$. Similarly, for $t = 2$, the algorithm is used to reduce the probability of selecting samples $s_{j'}^2$ for which $p_{ij'}^0 = 0$, for all $i = 1, \ldots, m$. The inclusion probabilities π_k^1 and π_k^2 are preserved via the first constraint. A solution to the linear programming Problem (6) always exists. If the value of the objective function equals zero, the AUB is achieved.

To formulate the second algorithm, define the set of 'nonpreferred' bi-samples for the joint sampling design by $\mathcal{S}_* = \{s_{ij} = (s_i^1, s_j^2) \in \mathcal{S} \mid \text{ with } p_{ij}^0 = 0, i = 1, \ldots, m, j = 1, \ldots, q\}$. In this case the set of 'nonpreferred' bi-samples is determined by all zeros in $\mathbf{P}^0$, and not only those from rows i' and columns j'. The goal is to reduce the impact of the bi-samples with $p_{ij}^0 = 0$. The algorithm is the linear programming solution to the following problem:

$$\min_{\mathbf{p}_*} \sum_{s_{ij} \in \mathcal{S}_*} p_*(s_{ij}), \tag{7}$$

subject to

$$\left| \begin{array}{l} \sum_{\substack{s_i^1 \ni k \\ s_i^1 \in \mathcal{S}^1}} \sum_{s_j^2 \in \mathcal{S}^2} p_*(s_{ij}) = \pi_k^1, k \in U, \\ \sum_{\substack{s_j^2 \ni k \\ s_j^2 \in \mathcal{S}^2}} \sum_{s_i^1 \in \mathcal{S}^1} p_*(s_{ij}) = \pi_k^2, k \in U, \\ \sum_{i=1}^m \sum_{j=1}^q p_*(s_{ij}) = 1, \\ p_*(s_{ij}) \geq 0, i = 1, \ldots, m, j = 1, \ldots, q. \end{array} \right.$$

For any feasible solution of Problem (7), the AUB is achieved if the objective function is 0. The inclusion probabilities π_k^1 and π_k^2 are preserved by the first two constraints of this problem.

Remark 3 *Rao and Nigam (1992) extended the approach given in Rao and Nigam (1990), and proposed a method to construct a controlled plan which matches the variance of a general linear unbiased estimator of a total associated with a specified uncontrolled plan. The controlled plan is constructed* *using linear programming. The general linear unbiased estimator of the total* $Y = \sum_{i=1}^{N} y_i$ *has the form*

$$\widehat{Y} = \sum_{k \in s} d_k(s) y_k, \ \ s \in \widetilde{\mathcal{S}},$$

where $\widetilde{\mathcal{S}}$ *is the set of all possible samples* s*, and the weights* $d_k(s)$ *may depend on the unit* k *or* s *or both.* $\widehat{Y}$ *is unbiased under the uncontrolled plan. The unbiasedness condition imposed by Rao and Nigam is* $\sum_{s \ni k} d_k(s) \tilde{p}_*(s) = 1, k = 1, \ldots, N$ *where* $\tilde{p}_*(s)$ *is the controlled plan. For the Horvitz-Thompson estimator with* $d_k(s) = 1/\pi_k$*, where* $\pi_k = Pr(k \in s)$*, and a large class of sampling designs the unbiased condition is simply* $\pi_k = \sum_{s \ni k} \tilde{p}_*(s), k = 1, \ldots, N$*. This condition is included as a constraint in Problems (6) and (7).*

5 The proposed method

We propose two methods for obtaining maximal sample coordination. In each case, we suppose that $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ are fixed and we want to obtain joint selection probabilities p_{ij} to achieve the AUB. The two methods are as follows:

- *Method I:*

  a) if conditions (3) and (4) are both satisfied, directly apply the IPF procedure to obtain p_{ij}, using as marginal distributions the initial plans, $\mathbf{p}^1$ and $\mathbf{p}^2$;

  b) otherwise, set the values of p_{ij} to zero as described in Section 3, i.e. replace $\mathbf{P}$ by $\mathbf{P}^0$;

  c) if there exist rows i' and/or columns j' in $\mathbf{P}^0$ containing only zero values:

     c1) apply the first algorithm defined by (6) to obtain $\mathbf{p}^1_*$ and/or $\mathbf{p}^2_*$ with minimum values of $p^1_*(s^1_{i'})$ and/or $p^2_*(s^2_{j'})$;

     c2) if $\mathcal{S}^1_* = \emptyset$ then set $\mathbf{p}^1_* = \mathbf{p}^1$; if $\mathcal{S}^2_* = \emptyset$ then set $\mathbf{p}^2_* = \mathbf{p}^2$;

     c3) apply the IPF procedure to obtain all p_{ij} different from zero with the marginal distributions $\mathbf{p}^1_*$ and $\mathbf{p}^2_*$;

d) else if there are no rows or columns of $\mathbf{P}^0$ consisting of zeros, obtain p_{ij} using the IPF procedure based on the initial marginal plans, $\mathbf{p}^1$ and $\mathbf{p}^2$.

- *Method II*

  – apply the second algorithm based upon (7) to compute $p_{ij} = p(s_{ij}) = p_*(s_{ij})$;

  – the new probabilities, $\mathbf{p}^1_*$ and $\mathbf{p}^2_*$, are obtained as marginal probabilities of the solution (p_{ij}).

Remark 4 *As mentioned in Rao and Nigam (1992), to select samples s^1_i or s^2_j with controlled probabilities $p^1_*(s^1_i)$ or $p^2_*(s^2_j)$, one can use the cumulative sums method or Lahiri's method (see Lahiri, 1951; Cochran, 1977).*

Remark 5 *The proposed method assumes the simultaneous selection of samples for two designs. However, this procedure can also be used for successive sampling (where samples are drawn on two successive occasions) if there are no rows i' in matrix $\mathbf{P}_0$ with only zero values. Consequently, the second design can be modified applying Problem (6), and becomes an optimal design for the first one (which is considered fixed), in order to achieve the AUB.*

Remark 6 *Since $\pi^{1,2}_k \geq \max(0, \pi^1_k + \pi^2_k - 1), k = 1, \ldots, N,$ the expected sample overlap is bounded below by*

$$E(c_{ij}) \geq \sum_{k \in U} \max(0, \pi^1_k + \pi^2_k - 1). \tag{8}$$

Minimal sample coordination occurs when the expected overlap equals the Absolute Lower Bound (ALB) defined by $\sum_{k \in U} \max(0, \pi^1_k + \pi^2_k - 1)$ in Expression (8). To reach the ALB similar conditions as in Proposition 1 are defined below:

Proposition 2 *Let (p_{ij}) denote an arbitrary sampling design for which the marginal designs $\mathbf{p}^1$ and $\mathbf{p}^2$ are given. Let $\widetilde{I} = \{k \in U | \pi^1_k + \pi^2_k - 1 \leq 0\}$ and let $\widetilde{D} = \{k \in U | \pi^1_k + \pi^2_k - 1 > 0\}$, with $U = \widetilde{I} \cup \widetilde{D}$, and $\widetilde{I} \cap \widetilde{D} = \emptyset$. The ALB is achieved by the design iff the following two relations are fulfilled:*

a1) if $(s^1_i \cap s^2_j) \cap \widetilde{I} \neq \varnothing$ then $p_{ij} = 0$, *b1) if* $\left(U \setminus (s_i^1 \cup s_j^2)\right) \cap \widetilde{D} \neq \varnothing$ *then* $p_{ij} = 0$,

for all $i \in \{1, \ldots, m\}$ *and* $j \in \{1, \ldots, q\}$.

The proposed method can be applied for ALB by solving Problems (6) or (7) based now on Proposition 2 instead of Proposition 1.

Remark 7 *As in other approaches based on mathematical programming (Causey et al., 1985; Ernst and Ikeda, 1995; Ernst, 1986, 1996, 1998; Ernst and Paben, 2002; Mach et al., 2006), the sizes of Problems (6) and (7) can increase very fast. Thus, the procedure is operationally feasible to implement only for moderate* m *and* q.

6 Example

We now apply the two methods proposed in Section 5 to an example taken from Causey et al. (1985) (see also the Appendix). Suppose $U = U^1 = U^2 = \{1, 2, 3, 4, 5\}$ and consider the following two initial sampling plans. There are $m = 12$ possible samples s_i^1 in the first initial sampling plan, as listed below:

$$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 4\}, \{1, 5\}, \{2, 4\}, \{2, 5\}, \{3, 4\}, \{3, 5\}, \{\},$$

with the selection probabilities

$$\mathbf{p}^1 = (0.15, 0.018, 0.012, 0.24, 0.04, 0.3, 0.05, 0.036, 0.006, 0.024, 0.004, 0.12).$$

For the second design, there are $q = 5$ possible samples $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}$ with selection probabilities

$$\mathbf{p}^2 = (0.4, 0.15, 0.05, 0.3, 0.1).$$

It follows that the inclusion probabilities of the marginal designs are

$$\boldsymbol{\pi}^1 = (\pi_k^1)_k = (0.5, 0.06, 0.04, 0.6, 0.1),$$

$$\boldsymbol{\pi}^2 = (\pi_k^2)_k = (0.4, 0.15, 0.05, 0.3, 0.1),$$

and these determine the sets $I = \{2, 3, 5\}$ and $D = \{1, 4\}$.

Given an arbitrary initial design $\mathbf{P}$ for which the marginal designs are as above, the matrix $\mathbf{P}^0$ takes the form set out below in (9). The rows correspond to the possible samples for the first design and the columns correspond to the possible samples for the second design. The non-zero values of matrix $\mathbf{P}^0$ are denoted x. Note that there are two rows consisting of zeros and corresponding to: $s_9^1 = \{2,5\}$ and $s_{11}^1 = \{3,5\}$, and there are no columns consisting just of zeros.

$$\mathbf{P}^0 = \left(\begin{array}{c|ccccc|} & \{1\} & \{2\} & \{3\} & \{4\} & \{5\} \\ \hline \{1\} & x & x & x & 0 & 0 \\ \{2\} & 0 & x & 0 & 0 & 0 \\ \{3\} & 0 & 0 & x & 0 & 0 \\ \{4\} & 0 & x & x & x & 0 \\ \{5\} & 0 & x & x & 0 & x \\ \{1,4\} & x & x & x & x & 0 \\ \{1,5\} & x & x & x & 0 & x \\ \{2,4\} & 0 & x & 0 & 0 & 0 \\ \{2,5\} & 0 & 0 & 0 & 0 & 0 \\ \{3,4\} & 0 & 0 & x & 0 & 0 \\ \{3,5\} & 0 & 0 & 0 & 0 & 0 \\ \{\} & 0 & x & x & 0 & 0 \end{array}\right) \qquad (9)$$

We now apply *Method 1* given in Section 5. The samples $\{2,5\}$ and $\{3,5\}$ have only zero values in the corresponding rows of $\mathbf{P}^0$ and so the application of IPF to this matrix cannot lead to maximal coordination for the given marginal designs using the method of Matei and Tillé (2005) . We apply the first algorithm defined by (6) for the case $t = 1$, with $\mathcal{S}_*^1 = \{\{2,5\},\{3,5\}\}$. The result is a new probability design

$$\mathbf{p}_*^1 = (0.3, 0, 0, 0.4, 0, 0.1, 0.1, 0.06, 0, 0.04, 0, 0).$$

The zero values in $\mathbf{p}_*^1$ correspond to the samples $\{2\}, \{3\}, \{5\}, \{2,5\}, \{3,5\}$ and $\{\}$. The application of the algorithm is not necessary for the case $t = 2$ since there is no single column in $\mathbf{P}^0$ consisting of zeros. We apply the IPF procedure for the samples with non-zero selection probabilities in the first design $\mathbf{p}_*^1$ and for all samples of the second design $\mathbf{p}^2$. The marginal distributions are now $\mathbf{p}_*^1$ and $\mathbf{p}^2$. The resulting matrix $\mathbf{P} = (p_{ij})_{6\times 5}$ is given in (10).

$$\mathbf{P} = \left(\begin{array}{c|ccccc|c} & \{1\} & \{2\} & \{3\} & \{4\} & \{5\} & \mathbf{p}_*^1 \\ \hline \{1\} & 0.3 & 0.00 & 0.00 & 0.0 & 0.0 & 0.30 \\ \{4\} & 0.0 & 0.09 & 0.01 & 0.3 & 0.0 & 0.40 \\ \{1,4\} & 0.1 & 0.00 & 0.00 & 0.0 & 0.0 & 0.10 \\ \{1,5\} & 0.0 & 0.00 & 0.00 & 0.0 & 0.1 & 0.10 \\ \{2,4\} & 0.0 & 0.06 & 0.00 & 0.0 & 0.0 & 0.06 \\ \{3,4\} & 0.0 & 0.00 & 0.04 & 0.0 & 0.0 & 0.04 \\ \hline \mathbf{p}^2 & 0.4 & 0.15 & 0.05 & 0.3 & 0.1 & 1 \end{array} \right) \qquad (10)$$

The matrix of the sample overlaps $\mathbf{C} = (c_{ij})_{6\times 5}$ is given in (11).

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}. \qquad (11)$$

Finally, $\sum_{i=1}^{6}\sum_{j=1}^{5} c_{ij}p_{ij} = \sum_{k\in U}\min(\pi_k^2, \pi_k^2) = 0.9$ and the maximal sample coordination is possible. Using the transportation problem approach and the uncontrolled plans, $\mathbf{p}^1$ and $\mathbf{p}^2$, Causey et al. (1985) have given the solution 0.88. Yet, with our approach, the AUB 0.9 is achieved.

We next apply *Method II* in Section 5 to the same example. Suppose that π_k^1 and π_k^2 are fixed, $\forall k \in U$ and we use the second algorithm defined by (7) to compute the values $p_*(s_{ij}) = p_{*,ij}$. The solution $\mathbf{P}_* = (p_{*,ij})$ of this problem is given in (12).

$$
\mathbf{P}_* = \left(\begin{array}{c|ccccc|c}
 & \{1\} & \{2\} & \{3\} & \{4\} & \{5\} & \mathbf{p}_*^1 \\
\hline
\{1\} & 0.10 & 0.09 & 0.01 & 0 & 0 & 0.20 \\
\{2\} & 0 & 0.06 & 0 & 0 & 0 & 0.06 \\
\{3\} & 0 & 0 & 0.04 & 0 & 0 & 0.04 \\
\{4\} & 0 & 0 & 0 & 0.30 & 0 & 0.30 \\
\{5\} & 0 & 0 & 0 & 0 & 0.10 & 0.10 \\
\{1,4\} & 0.30 & 0 & 0 & 0 & 0 & 0.30 \\
\hline
\{1,5\} & 0 & 0 & 0 & 0 & 0 & 0 \\
\{2,4\} & 0 & 0 & 0 & 0 & 0 & 0 \\
\{2,5\} & 0 & 0 & 0 & 0 & 0 & 0 \\
\{3,4\} & 0 & 0 & 0 & 0 & 0 & 0 \\
\{3,5\} & 0 & 0 & 0 & 0 & 0 & 0 \\
\{\} & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
\mathbf{p}_*^2 & 0.40 & 0.15 & 0.05 & 0.30 & 0.10 & 1
\end{array}\right) \tag{12}
$$

The new marginal probabilities are obtained as $p_*^1(s_i^1) = \sum_{j=1}^{5} p_{*,ij}$ and $p_*^2(s_j^2) = \sum_{i=1}^{12} p_{*,ij}, i = 1, \ldots, 12, j = 1, \ldots, 5$. The last 6 rows in $\mathbf{P}_*$ have only zeros values. For the first design, the samples s_i^1 with $p_*^1(s_i^1) \neq 0$ are $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1,4\}$.

As for the first method, the AUB is achieved. And, as for the first method, the number of possible samples with a strictly positive probability (the sample support) for the first design is greatly reduced, since we obtain 6 possible samples instead of 12.

7 Discussion

In this paper, we have proposed a method for sample coordination, designed to maximize (or minimize) expected overlap between two samples, while constraining the inclusion probabilities for units in each sample to specified values. The method is primarily designed for use when two surveys are being conducted simultaneously and it is desired to maximize (or minimize) the overlap between the two samples. For example, both surveys may involve multistage stratified sampling and the primary sampling units (PSUs) might consist of geographical areas. If both surveys are to use face-to-face interviewing then inclusion of the same PSUs in both surveys might reduce travel and interviewer recruitment costs if the same interviewers can work on both surveys. The inclusion probabilities for the PSUs in each survey may be prespecified according to a probability proportional to size design and the size measures employed in the two surveys might differ so that it may not be feasible to employ the same sample of PSUs in both surveys.

The proposed method in such circumstances involves first specifying probability designs for the sampling of PSUs in each survey, which meet the conditions on the inclusion probabilities. Next, the method of Matei and Tillé (2005) is applied to maximize the expected overlap between the two designs, generating a new joint design for the two samples. Next, one of the methods of controlled selection described in this paper is applied to improve the expected overlap further, subject to the given inclusion probabilities. The extent to which the increase in expected overlap is practically useful requires further investigation in real survey applications. In the numerical example in this paper, the use of controlled selection increased the expected overlap from 0.88 to 0.9.

Although we have focussed on the problem of achieving maximal sample coordination in circumstances when two samples are to be selected simultaneously, under some circumstances the method can also be applied to the selection of successive samples in repeated surveys. In this case, the marginal sample design for the first sample will generally be given and the aim is to select the second sample with specified inclusion probabilities such that expected overlap is maximum. Our approach can then be applied by using first the method of Matei and Tillé (2005) with the sampling design on the first occasion given. As pointed out in Remark 5, however, our approach to controlled selection can then only be applied if the IPF algorithm of Matei and Tillé (2005) does not result in any of the samples which were feasible at the first occasion being made impossible.

Our final comment relates to computation. A key motivation for the method Matei and Tillé (2005) was to avoid the major computational demands of linear programming approaches to coordination. In our approach, however, we have reintroduced some element of linear programming. Some further research is required to compare the computational demands of our approach and the traditional linear programming approach via the transportation problem.

Acknowledgements

The authors would like to thank the anonymous referee for her/his helpful comments and suggestions which greatly improved the paper. This project was supported by the grant PI0I2–117192/1 from the Swiss National Science Foundation.

Appendix

Transportation problem

Sample coordination using the transportation problem approach was studied by Raj (1968), Arthnari and Dodge (1981), Causey et al. (1985), Ernst and Ikeda (1995), Ernst (1996), Ernst (1998), Ernst and Paben (2002) and more recently by Mach et al. (2006). The transportation problem was used to maximize (minimize) the overlap between samples for periodic surveys with a multistage stratified design. Causey et al. (1985) assumed the following conditions for applying the transportation problem approach: there are two surveys with different stratifications and each stratum $\mathfrak{S}$ in the second survey is a separate overlap problem. Samples using the two designs are selected sequentially, on two different time occasions. Let m be the number of initial strata which have a non-empty intersection with $\mathfrak{S}$. The intersection between $\mathfrak{S}$ and an initial stratum i is a random subset s_i^1, for $i = 1, \dots, m$. The probability $p^1(s_i^1)$ is known. On the second occasion, there are q possible samples s_j^2, with the selection probability $p^2(s_j^2), j = 1, \dots, q$.

Following Causey et al. (1985), the transportation problem to be solved for positive coordination is given in Problem (13):

$$\max \sum_{i=1}^{m} \sum_{j=1}^{q} c_{ij} p_{ij}, \tag{13}$$

subject to

$$\left| \begin{array}{l} \sum_{j=1}^{q} p_{ij} = p_i^1, i = 1, \dots, m, \\ \sum_{i=1}^{m} p_{ij} = p_j^2, j = 1, \dots, q, \\ \sum_{i=1}^{m} \sum_{j=1}^{q} p_{ij} = 1, \\ p_{ij} \geq 0, i = 1, \dots, m, j = 1, \dots, q, \end{array} \right.$$

where

$$c_{ij} = |s_i^1 \cap s_j^2|, p_i^1 = \Pr(s_i^1), p_j^2 = \Pr(s_j^2), p_{ij} = \Pr(s_i^1, s_j^2),$$

$s_i^1 \in \mathcal{S}^1$ and $s_j^2 \in \mathcal{S}^2$ denote all the possible samples in the first and second design, respectively, with $m = |\mathcal{S}^1|$ and $q = |\mathcal{S}^2|$. We suppose that $p_i^1 > 0, p_j^2 > 0$ in order to compute the conditional probabilities. So, given $\mathbf{p}^1$ and $\mathbf{p}^2$, finding the optimum expected overlap amounts to finding the maximum, over all $\mathbf{P} = (p_{ij})_{m \times q}$. Once a solution p_{ij} is obtained using Problem (13), conditional on the already selected sample $s^1 = s_i^1$, the selection probability for s_j^2 is p_{ij}/p_i^1. For negative coordination, the function max is replaced by min in Problem (13), keeping the same constraints.

References

Aragon, J. and Pathak, P. K. (1990). An algorithm for optimal integration of two surveys. *Sankyā*, Ser.B, 52:198–203.

Arthnari, T. and Dodge, Y. (1981). *Mathematical Programming in Statistics.* Wiley, Inc., New York.

Brewer, K., Early, L., and Hanif, M. (1984). Poisson, modified Poisson and Collocated sampling. *Journal of Statistical Planning and Inference*, 10:15–30.

Brewer, K., Early, L., and Joyce, S. (1972). Selecting several samples from a single population. *Australian Journal of Statistics*, 3:231–239.

Causey, B. D., Cox, L. H., and Ernst, L. R. (1985). Application of transportation theory to statistical problems. *Journal of the American Statistical Association*, 80:903–909.

Cochran, W. (1977). *Sampling Techniques.* Wiley, New York.

Cotton, F. and Hesse, C. (1992). Tirages coordonnés d'échantillons. Technical Report E9206, Direction des Statistiques Économiques, INSEE, Paris.

Deming, W. and Stephan, F. (1940). On a least square adjustment of sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11:427–444.

Ernst, L. R. (1986). Maximizing the overlap between surveys when information is incomplete. *European Journal of Operational Research*, 27:192–200.

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. (1999). The maximization and minimization of sample overlap problems: a half century of results. In *Proceedings of the International Statistical Institute, 52nd Session*, pages 168–182, Helsinki, Finland.

Ernst, L. R. and Ikeda, M. M. (1995). A reduced-size transportation algorithm for maximizing the overlap between surveys. *Survey Methodology*, 21:147–157.

Ernst, L. R. and Paben, S. P. (2002). Maximizing and minimizing overlap when selecting any number of units per stratum simultaneously for two designs with different stratifications. *Journal of Official Statistics*, 18:185–202.

Keyfitz, N. (1951). Sampling with probabilities proportional to size: adjustment for changes in the probabilities. *Journal of American Statistics Association*, 46:105–109.

Lahiri, D. (1951). A method for sample selection providing unbiased ratio estimates. *Bulletin of the International Statistical Association*, 33:133–140.

Mach, L., Reiss, P. T., and Ṣchiopu-Kratina, I. (2006). Optimizing the expected overlap of survey samples via the northwest corner rule. *Journal of the American Statistical Association*, 101(476):1671–1679.

Matei, A. and Tillé, Y. (2005). Maximal and minimal sample co-ordination. *Sankhyā*, 67, part 3:590–612.

Ohlsson, E. (1995). Coordination of samples using permanent random numbers. In Cox, B. G., Binder, D. A., Chinnapa, B. N., Christianson, A., Colledge, M. J., and Kott, P. S., editors, *Business Survey Methods*, chapter 9, pages 153–169. Wiley, New York, USA.

Ohlsson, E. (2000). Coordination of pps samples over time. In *Proceedings of the Second International Conference on Establishment Surveys*, pages 255–264. American Statistical Association.

Pathak, P. K. and Fahimi, M. (1992). Optimal integration of surveys. In Ghosh, M. and Pathak, P. K., editors, *Current issues in statistical inference: Essays in honor of D. Basu*, pages 208–224. Hayward, CA: Institute of Mathematical Statistics.

Perry, C., Burt, J., and Iwig, W. (1993). Methods of selecting samples in multiple surveys to reduce respondent burden. *Proceedings of the International Conference on Establishment Surveys, Alexandria, VA: American Statistical Association*, pages 345–351.

Raj, D. (1968). *Sampling Theory.* McGraw-Hill, New York.

Rao, J. N. K. and Nigam, A. K. (1990). Optimal controlled sampling designs. *Biometrika*, 77:807–814.

Rao, J. N. K. and Nigam, A. K. (1992). 'Optimal' controlled sampling: A unified approach. *International Statistical Review*, 60, no. 1:89–98.