Published in International Statistical Review 76, issue 3, 368-386, 2008
which should be used for any reference to this work

Sampling Procedures for Coordinating Stratified Samples: Methods Based on Microstrata

Desislava Nedyalkova[1], Johan Pea[2] and Yves Tillé[1]
[1]Institute of Statistics, University of Neuchâtel, Pierre à Mazel 7, 2000 Neuchâtel, Switzerland
E-mail: Yves.Tille@unine.ch
[2]Swiss Federal Statistical Office, 2010 Neuchâtel, Switzerland
E-mail: Johan.Pea@admin.ch

Summary

The aim of sampling coordination is to maximize or minimize the overlap between several samples drawn successively in a population that changes over time. Therefore, the selection of a new sample will depend on the samples previously drawn. In order to obtain a larger (or smaller) overlap of the samples than the one obtained by independent selection of samples, a dependence between the samples must be introduced. This dependence will emphasize (or limit) the number of common units in the selected samples. Several methods for coordinating stratified samples, such as the Kish & Scott method, the Cotton & Hesse method, and the Rivière method, have already been developed. Using simulations, we compare the optimality of these methods and their quality of coordination. We present six new methods based on permanent random numbers (PRNs) and microstrata. These new methods have the advantage of allowing us to choose between positive or negative coordination with each of the previous samples. Simulations are run to test the validity of each of them.

Key words: Sample coordination; stratified samples; permanent random numbers; microstrata.

1 Introduction

The coordination problem has been a main topic of interest for many years. We distinguish two main types of coordination: negative and positive. In negative coordination, we want to minimize the number of common units between several samples drawn successively in a population that changes over time, whereas in positive coordination, we want to maximize this number. The first papers on coordination were written by Patterson (1950) and Keyfitz (1951). The first works on coordination present methods that are in general restricted to two successive samples or small sample sizes. At a later stage, Kish & Scott (1971) generalized the coordination problem in the context of a larger sample size.

The concept of coordination based on PRNs was introduced by Brewer et al. (1972). Most of the national bureaus of statistics use variations of methods based on PRN sampling. Ohlsson (1995) presented a summary of the methods used in different countries. Another approach that takes into account the concept of PRNs, called order sampling, was proposed by Rosén (1997a,b).

The coordination of stratified samples is a more complex problem. The main reason is that, over time, units usually change from one stratum to another. Several methods, the Kish & Scott method presented in Kish & Scott (1971), the Cotton & Hesse method presented in Cotton & Hesse (1992), the Dutch method (EDS) described in De Ree (1983), Van Huis et al. (1994a,b), Koeijers & Willeboordse (1995), and the Rivière method presented in Rivière (1998, 1999, 2001a,b), have already been developed in order to obtain maximal or minimal coverage between samples drawn on different occasions. However, the Dutch method is not of much interest to us because it does not allow strata to be changed.

The methods that we will introduce are based on the use of PRNs and microstrata. They allow us to choose between negative and positive coordination with the previous waves, which is a major advantage. To do positive coordination, we should just coordinate negatively with the complement of the sample. In some of the methods, the PRNs are permuted in a chronological manner, according to what happened at the previous stages, whereas in others the PRNs are permuted in a retrospective manner. To illustrate the advantages and drawbacks of each one of these methods, simulations have been run.

This paper is structured as follows: some basic notions and definitions are given in Sections 2 and 3. Section 4 introduces the Kish & Scott method. Section 5 presents the Cotton & Hesse method. A comparison of the two methods is given in Section 6. Section 7 presents the Rivière method. Section 8 is devoted to the new methods that we introduce. Section 9 presents the simulation results. Finally, in Section 10, a few concluding remarks are given.

2 Population, Sample, and Sampling Design

We define a finite population as a set of N units $\{u_1, \dots, u_k, \dots, u_N\}$. Each unit can be identified without ambiguity by a label. Let $U = \{1, \dots, k, \dots, N\}$ be the set of these labels. The size N of the population is not necessarily known. In the problems of sampling coordination, the population can change over time. Suppose that we are interested in studying a population at times $t = 1, 2, \dots, T-1, T$. Let U^t denote the population at time t. The set $U^t \backslash U^{t-1}$ contains the births at time t. The set $U^{t-1} \backslash U^t$ holds the deaths at time t. The population U contains all the units from time 1 to T

$$U = \bigcup_{t=1}^{T} U^t.$$

At time t, a sample without replacement is a subset of the population U^t. Since $U^t \subset U$, a sample is also a subset of U. The sample is denoted by a vector

$$\mathbf{s}^t = \left(s_1^t, \dots, s_k^t, \dots, s_N^t\right)' \in \{0, 1\}^N,$$

where

$$s_k^t = \begin{cases} 1 & \text{if, at time } t, \text{ unit } k \text{ is in the sample} \\ 0 & \text{if, at time } t, \text{ unit } k \text{ is not in the sample,} \end{cases}$$

for all $k \in U$.

The joint sampling design, $p(\mathbf{s}^1, \dots, \mathbf{s}^t, \dots, \mathbf{s}^T)$, is a probability distribution for all the occasions. Let $\mathbf{S}^1, \dots, \mathbf{S}^t, \dots, \mathbf{S}^T$ denote the random samples as follows:

$$\Pr\left(\mathbf{S}^1 = \mathbf{s}^1, \dots, \mathbf{S}^t = \mathbf{s}^t, \dots, \mathbf{S}^T = \mathbf{s}^T\right) = p\left(\mathbf{s}^1, \dots, \mathbf{s}^t, \dots, \mathbf{s}^T\right).$$

The size of $\mathbf{S}^t$ is denoted by n^t.

From the joint sampling design, one can derive the marginal design for the particular time t:

$$\sum_{\mathbf{s}^1,\ldots,\mathbf{s}^{t-1},\mathbf{s}^{t+1},\ldots,\mathbf{s}^T} p(\mathbf{s}^1,\ldots,\mathbf{s}^t,\ldots,\mathbf{s}^T) = p_t(\mathbf{s}^t).$$

At time t, the first-order inclusion probability and the joint inclusion probability are denoted, respectively, by

$$\pi_k^t = E(S_k^t) \quad \text{and} \quad \pi_{k\ell}^t = E(S_k^t S_\ell^t),$$

where $k, \ell \in U^t$, $t = 1,\ldots,T$. The longitudinal inclusion probability, for times t and u, is given by

$$\pi_k^{tu} = E(S_k^t S_k^u),\ k \in U^t \cap U^u, t, u = 1,\ldots,T.$$

Finally, the joint longitudinal probability has the form:

$$\pi_{k\ell}^{tu} = E(S_k^t S_\ell^u),\ k, \ell \in S^t \cup S^u, t, u = 1,\ldots,T.$$

Note that this probability is not symmetrical. Indeed, $\pi_{k\ell}^{tu} \neq \pi_{\ell k}^{tu}$ and $\pi_{k\ell}^{tu} \neq \pi_{\ell k}^{ut}$.

The following basic result gives bounds for the longitudinal inclusion probabilities.

Result 1. For times t and u, we have

$$\max\left(0, \pi_k^t + \pi_k^u - 1\right) \le \pi_k^{tu} \le \min\left(\pi_k^t, \pi_k^u\right).$$

Proof. By definition, we have

$$\pi_k^{tu} = \Pr\left(S_k^t = 1 \text{ and } S_k^u = 1\right) \le \min\left[\Pr\left(S_k^t = 1\right), \Pr\left(S_k^u = 1\right)\right] = \min\left(\pi_k^t, \pi_k^u\right).$$

Moreover,

$$\begin{aligned}\pi_k^t - \pi_k^{tu} &= \Pr\left(S_k^t = 1\right) - \Pr\left(S_k^t = 1 \text{ and } S_k^u = 1\right) = \Pr\left(S_k^t = 1 \text{ and } S_k^u = 0\right) \\ &\le \min\left[\Pr\left(S_k^t = 1\right), \Pr\left(S_k^u = 0\right)\right] = \min\left(\pi_k^t, 1 - \pi_k^u\right).\end{aligned}$$

Thus,

$$\pi_k^{tu} \ge \pi_k^t - \min\left(\pi_k^t, 1 - \pi_k^u\right) = \max\left(0, \pi_k^t + \pi_k^u - 1\right).$$

Consider a population U split into H parts U_h, called "strata", such that

$$\cup_{h=1}^H U_h = U \quad \text{and} \quad U_h \cap U_i = \emptyset,$$

for all (h, i) with $h \neq i$. A design is called stratified if a random sample S_h of fixed size n_h is selected in each stratum U_h, and if the sample selection in each stratum is taken independently of the selection done in all the other strata.

3 Sample Coordination, Overlap, and Burden

The overlap is the number of common units at two different times t and u:

$$n^{tu} = \sum_{k \in U} S_k^t S_k^u.$$

The overlap can be random. The expected overlap is

$$E(n^{tu}) = E\left(\sum_{k \in U} S_k^t S_k^u\right) = \sum_{k \in U} E(S_k^t S_k^u) = \sum_{k \in U} \pi_k^{tu}.$$

The overlap rate is defined by:

$$\upsilon^{tu} = \frac{2n^{tu}}{n^t + n^u}.$$

If n^t and n^u are fixed, then the expected overlap rate is given by:

$$\tau^{tu} = \frac{2E(n^{tu})}{n^t + n^u}.$$

Let $\text{ALB} = \sum_{k\in U} \max(0, \pi_k^t + \pi_k^u - 1)$ denote the absolute lower bound and $\text{AUB} = \sum_{k\in U} \min(\pi_k^t, \pi_k^u)$ denote the absolute upper bound (see Matei & Tillé, 2005). Then, from result 1, we can directly derive bounds for the expected overlap:

$$\text{ALB} \le E(n^{tu}) \le \text{AUB}.$$

Unfortunately, except for very particular cases like simple random sampling (SRS), the ALB and AUB cannot be reached.

If, at times 1 and 2, two samples are drawn independently without coordination, then, for all $k \in U$,

$$\pi_k^1 \pi_k^2 = \pi_k^{12}.$$

In positive coordination, for all $k \in U$, the longitudinal inclusion probability must satisfy the conditions

$$\pi_k^1 \pi_k^2 \le \pi_k^{12} \le \min\left(\pi_k^1, \pi_k^2\right).$$

In negative coordination, for all $k \in U$, the longitudinal inclusion probability must satisfy the conditions

$$\max\left(0, \pi_k^1 + \pi_k^2 - 1\right) \le \pi_k^{12} \le \pi_k^1 \pi_k^2.$$

Note that, in the last case, the longitudinal inclusion probability can be zero only if $\pi_k^1 + \pi_k^2 \le 1$.

The response burden of a survey is usually quantified in terms of the time needed to complete the questionnaire. However, other aspects of response burden exist: for example, how difficult it is to provide the information or how sensitive the question sent to the respondent is. Therefore, the response burden can vary from one survey to another.

At time t, a survey has a burden denoted by b^t, which can be proportional to the time needed to complete the form or can be simply equal to one. After T waves, the total burden of unit k is defined as the sum of the burdens of the surveys in which unit k has been included:

$$c_k^T = \sum_{t=1}^{T} b^t S_k^t.$$

We also define the cumulative burden from survey m to survey T, named (m, T)-cumulated burden, as:

$$c_k^{m,T} = \sum_{t=m}^{T} b^t S_k^t.$$

The quality of a procedure concerning coordination can be measured using four possible criteria:

1. the procedure provides a controllable degree of overlap;
2. the sampling design is respected in each selection;

3. for each unit, a fixed time out of sample is respected;
4. the procedure is computed easily.

4 The Kish & Scott Method

Kish & Scott (1971) have proposed a method of substitution for coordinating stratified samples, which allows changes in the definition of the strata. Although they had introduced this method for positive coordination, presented in Algorithm 1, it also allows us to do negative coordination. At times 1 and 2, i.e. waves 1 and 2, the definition of the strata can change. From this point forward, we will use the terms times and waves interchangeably.

In order to present a rigorous algorithm, it is necessary to formalize the notation. We also assume that there are no births and deaths in the population. Suppose that the population U is stratified at time 1 into H strata $U_1^1, \dots, U_h^1, \dots, U_H^1$, and at time 2 into G strata $U_1^2, \dots, U_g^2, \dots, U_G^2$ as follows:

$$U = \cup_{h=1}^{H} U_h^1 = \cup_{g=1}^{G} U_g^2.$$

Let N_h^1 be the size of U_h^1, N_g^2 the size of U_g^2, and N_{hg}^{12} the size of $U_{hg}^{12} = U_h^1 \cap U_g^2$. Suppose that two independent stratified samples $\mathbf{s}^1$ and $\mathbf{s}^2$ are drawn, at time 1 and time 2, respectively. Also consider the following notations:

- s_g^i the set of units of stratum U_g^2 that are selected in $\mathbf{s}^i$, for $i = 1, 2$, with $n_g^i = \text{card}\,(\mathbf{s}_g^i)$,
- s_{hg}^i the set of units of $U_h^1 \cap U_g^2$ that are selected in $\mathbf{s}^i$, for $i = 1, 2$, with $n_{hg}^i = \text{card}\,(\mathbf{s}_{hg}^i)$,
- $s_{hg}^{12} = s_{hg}^1 \cap s_{hg}^2$, with $n_{hg}^{12} = \text{card}\,(\mathbf{s}_{hg}^{12})$.

This method is correct because it provides two conditional stratified samples. Nevertheless, only two waves can be coordinated because the coordination of more than two samples becomes very complicated. Simulations show that the coordination is not very good. Generally, the Cotton & Hesse method performs better, which we will show in a simulation example in Section 6 of our paper.

Algorithm 1 Positive coordination using the Kish & Scott method.

At wave 1, draw a stratified sample from U, denoted by $\mathbf{s}^1$.
At wave 2, draw a stratified sample from U, denoted by $\mathbf{s}^2$ and
for Each possible intersection of strata U_{hg}^{12} **do**
  if $n_{hg}^2 - n_{hg}^{12} \geq n_{hg}^1 - n_{hg}^{12}$ **then**
    Replace $n_{hg}^1 - n_{hg}^{12}$ units from $s_{hg}^2 \backslash s_{hg}^{12}$ with units of $s_{hg}^1 \backslash s_{hg}^{12}$ by means of SRS.
  else
    Replace $n_{hg}^2 - n_{hg}^{12}$ units from $s_{hg}^2 \backslash s_{hg}^{12}$ with units of $s_{hg}^1 \backslash s_{hg}^{12}$ by means of SRS.
  end if
end for

---

5 The Cotton & Hesse Method

The Cotton & Hesse method from the Institut National de la Statistique et des Études Économiques (INSEE) of France is fully described in Cotton & Hesse (1992). This method works when the strata change over time and can be used to obtain negative coordination. The principle is as follows: Each unit of the population receives a PRN ω_k from a uniform distribution U[0, 1]. At the first wave, the sample is defined, in each stratum, as the set of units that have the smallest random numbers. After the sample has been selected, the PRNs are permuted in such a way that the units selected at the first wave receive the largest PRNs, and the non-selected units receive the smallest PRNs. Within the two subsets of selected and non-selected units, the order of the permuted PRNs must remain unchanged. Then the same procedure is applied for the subsequent waves. The procedure for negative coordination is presented in Algorithm 2. Note that the dead units lose their PRNs, while the new units receive a new PRN.

Algorithm 2 Negative coordination using the Cotton & Hesse method.

Assign, independently, a PRN ω_k^1 to each unit $k \in U$ and construct $\boldsymbol{\omega}^1 = \{\omega_1^1, \ldots, \omega_N^1\}$.
for $t = 1, \ldots, T$ **do**
  Select the units that have the n_h^t smallest ω_k^t to obtain the sample $\mathbf{s}^t$.
  Assign the largest ω_k^t to the units that belong to $\mathbf{s}^t$.
  Assign the smallest ω_k^t to the units that belong to $U \setminus \mathbf{s}^t$.
  Construct $\boldsymbol{\omega}^{t+1}$ as a permutation of $\boldsymbol{\omega}^t$ so that the rank of $\boldsymbol{\omega}^t$ in $\mathbf{s}^t$ and in $U \setminus \mathbf{s}^t$ is respected.
end for

The major advantage of this method is that the strata can change over time. The method is correct, because after the permutation, the PRNs remain independent uniform random numbers; the method is thus very simple to apply. Another advantage of this method is that only the permuted PRNs must be retained from one wave to another. The drawback of this method is that it allows only one kind of coordination. Once you have decided to do negative coordination between two time periods, you cannot do positive coordination while drawing another sample. Moreover, the order of the surveys is fixed and cannot be changed.

6 Comparison of the Kish & Scott and Cotton & Hesse Methods

We will consider a very simple example in order to compare both methods. Since each problem of coordination can be also viewed as a problem of optimization, we can find the optimal solution and then compare this solution to the solutions of the Kish & Scott and Cotton & Hesse methods. We will see that neither method is optimal.

Suppose that the population $U = \{1, 2, 3, 4\}$. At time 1, the strata are $\{1, 2\}$, $\{3, 4\}$, and at time 2, $\{1, 3\}$, $\{2, 4\}$. At times 1 and 2, two stratified samples are selected with only one unit in each stratum. The aim is to obtain the best negative coordination. At time 1, the possible samples are given by:

$$\mathbf{s}_1^1 = (1\ 0\ 1\ 0)', \quad \mathbf{s}_2^1 = (1\ 0\ 0\ 1)', \quad \mathbf{s}_3^1 = (0\ 1\ 1\ 0)', \quad \mathbf{s}_4^1 = (0\ 1\ 0\ 1)'.$$

All samples are selected with probability $p(s_i^1) = 1/4, i = 1, \ldots, 4$ and $\pi_k^1 = 1/2$ for all $k \in U$.

At time 2, the possible samples are

$$\mathbf{s}_1^2 = (1\ 1\ 0\ 0)', \quad \mathbf{s}_2^2 = (1\ 0\ 0\ 1)', \quad \mathbf{s}_3^2 = (0\ 1\ 1\ 0)', \quad \mathbf{s}_4^2 = (0\ 0\ 1\ 1)'.$$

Here, again, all samples are selected with probability $p(s_i^2) = 1/4, i = 1, \ldots, 4$ and $\pi_k^2 = 1/2$ for all $k \in U$.

Table 1
Overlap between the possible samples at times 1 and 2.

	$\mathbf{s}_1^2$	$\mathbf{s}_2^2$	$\mathbf{s}_3^2$	$\mathbf{s}_4^2$
$\mathbf{s}_1^1$	1	1	1	1
$\mathbf{s}_2^1$	1	2	0	1
$\mathbf{s}_3^1$	1	0	2	1
$\mathbf{s}_4^1$	1	1	1	1

Table 2
Set of optimal solutions for $c \in [0, 1/4]$.

	$\mathbf{s}_1^2$	$\mathbf{s}_2^2$	$\mathbf{s}_3^2$	$\mathbf{s}_4^2$
$\mathbf{s}_1^1$	1/4-c	0	0	c
$\mathbf{s}_2^1$	0	0	1/4	0
$\mathbf{s}_3^1$	0	1/4	0	0
$\mathbf{s}_4^1$	c	0	0	1/4-c

Since $n^1 = 2$ and $n^2 = 2$, then

$$\mathrm{ALB} = \sum_{k \in U} \max\left(0, \pi_k^1 + \pi_k^2 - 1\right) = 0.$$

Nevertheless, we will see that the ALB cannot be reached due to the constraints of stratification. The overlap between the different samples is given in Table 1.

If $n_{ij}^{12} = \mathbf{s}_i^{1\prime} \mathbf{s}_j^2$ is the number of common units of samples $\mathbf{s}_i^1$ and $\mathbf{s}_j^2$, and $p_{ij} = \Pr(\mathbf{s}_i^1, \mathbf{s}_j^2)$, the optimal solution is obtained by solving

$$\arg\min_{p_{ij}} \sum_{i=1}^{4} \sum_{j=1}^{4} n_{ij}^{12} p_{ij} \quad \text{subject to} \quad \begin{cases} p_{ij} > 0, \\ \sum_{i=1}^{4} p_{ij} = 1/4, \text{ for } j = 1, \dots, 4, \\ \sum_{j=1}^{4} p_{ij} = 1/4, \text{ for } i = 1, \dots, 4. \end{cases}$$

This is a linear program that can be solved easily. The set of optimal solutions is given in Table 2.

If we take $c = 0$, we find one of the following optimal solutions:

- If, at time 1, $\mathbf{s}_1^1$ is selected, then select $\mathbf{s}_1^2$ at time 2.
- If, at time 1, $\mathbf{s}_2^1$ is selected, then select $\mathbf{s}_3^2$ at time 2.
- If, at time 1, $\mathbf{s}_3^1$ is selected, then select $\mathbf{s}_2^2$ at time 2.
- If, at time 1, $\mathbf{s}_4^1$ is selected, then select $\mathbf{s}_4^2$ at time 2.

The expected overlap is $E\left(n^{12}\right) = 0.5$, and the expected overlap rate is

$$\tau^{12} = \frac{2E(n^{12})}{n^1 + n^2} = \frac{2 \times 0.5}{2 + 2} = 0.25.$$

Table 3
Negative coordination with the Kish & Scott method.

Time 1	s_1^1	s_1^1	s_1^1	s_1^1	s_2^1	s_2^1	s_2^1	s_2^1	s_3^1	s_3^1	s_3^1	s_3^1	s_4^1	s_4^1	s_4^1	s_4^1
Time 2	s_1^2	s_2^2	s_3^2	s_4^2	s_1^2	s_2^2	s_3^2	s_4^2	s_1^2	s_2^2	s_3^2	s_4^2	s_1^2	s_2^2	s_3^2	s_4^2
Overlap	1	1	1	1	1	2	0	1	1	0	2	1	1	1	1	1

Table 4
Negative coordination with the Cotton & Hesse method.

Ranks	Time 1	Permuted Ranks	Time 2	Overlap
(1 2 3 4)	(1 0 1 0)	(2 1 4 3)	(1 1 0 0)	1
(1 2 4 3)	(1 0 0 1)	(2 1 3 4)	(1 1 0 0)	1
(1 3 2 4)	(1 0 1 0)	(3 1 4 2)	(1 1 0 0)	1
(1 3 4 2)	(1 0 0 1)	(3 1 2 4)	(0 1 1 0)	0
(1 4 2 3)	(1 0 1 0)	(4 1 3 2)	(0 1 1 0)	1
(1 4 3 2)	(1 0 0 1)	(4 1 2 3)	(0 1 1 0)	0
(2 1 3 4)	(0 1 1 0)	(1 2 4 3)	(1 1 0 0)	1
(2 1 4 3)	(0 1 0 1)	(1 2 3 4)	(1 1 0 0)	1
(2 3 1 4)	(1 0 1 0)	(3 2 4 1)	(1 0 0 1)	1
(2 3 4 1)	(1 0 0 1)	(3 2 1 4)	(0 1 1 0)	0
(2 4 1 3)	(1 0 1 0)	(4 2 3 1)	(0 0 1 1)	1
(2 4 3 1)	(1 0 0 1)	(4 2 1 3)	(0 1 1 0)	0
(3 1 2 4)	(0 1 1 0)	(1 3 4 2)	(1 0 0 1)	0
(3 1 4 2)	(0 1 0 1)	(1 3 2 4)	(1 1 0 0)	1
(3 2 1 4)	(0 1 1 0)	(2 3 4 1)	(1 0 0 1)	0
(3 2 4 1)	(0 1 0 1)	(2 3 1 4)	(0 1 1 0)	1
(3 4 1 2)	(1 0 1 0)	(4 3 2 1)	(0 0 1 1)	1
(3 4 2 1)	(1 0 0 1)	(4 3 1 2)	(0 0 1 1)	1
(4 1 2 3)	(0 1 1 0)	(1 4 3 2)	(1 0 0 1)	0
(4 1 3 2)	(0 1 0 1)	(1 4 2 3)	(1 0 0 1)	1
(4 2 1 3)	(0 1 1 0)	(2 4 3 1)	(1 0 0 1)	0
(4 2 3 1)	(0 1 0 1)	(2 4 1 3)	(0 0 1 1)	1
(4 3 1 2)	(0 1 1 0)	(3 4 2 1)	(0 0 1 1)	1
(4 3 2 1)	(0 1 0 1)	(3 4 1 2)	(0 0 1 1)	1

The relative lower bound (RLB) (see Matei & Tillé, 2005) is the value of the objective function when the linear problem is solved. The following relation holds:

$$\mathrm{RLB} = \arg\min_{p_{ij}} \sum_{i=1}^{m} \sum_{j=1}^{q} n_{ij}^{12} p_{ij} \geq \sum_{k \in U} \max\left(0, \pi_k^1 + \pi_k^2 - 1\right) = \mathrm{ALB}.$$

Here, RLB $= 0.5$ is strictly larger than ALB $= 0$.

If we apply a negative coordination using the Kish & Scott method to the strata we have defined above, then we obtain the results given in Table 3. Note that, the expected overlap is 1 and the expected overlap rate is

$$\tau^{12} = \frac{2E(n^{12})}{n^1 + n^2} = \frac{2 \times 1}{2+2} = \frac{1}{2}.$$

In this case, the quality of coordination is not better than for independent stratified samples.

If we apply a negative coordination using the Cotton & Hesse method, then we obtain the following results, presented in Table 4.

Note that, in this case, the expected overlap is 2/3 and the expected overlap rate is

$$\tau_{12} = \frac{2E(n^{12})}{n^1 + n^2} = \frac{2 \times 2/3}{2+2} = \frac{1}{3}.$$

Thus, the optimal solution, which has an expected overlap of 1/2, is not reached. The Cotton & Hesse method does not provide the best solution to the problem of coordination.

This small example is very interesting because it shows that the optimality is not reached by either the Cotton & Hesse method or the Kish & Scott method. However, the Cotton & Hesse method gives a slightly better solution than the Kish & Scott method. This result was also confirmed by a set of simulations. Therefore, we advocate the use of the Cotton & Hesse method rather than the Kish & Scott method provided that the drawbacks of the Cotton & Hesse method are not an issue.

7 The Rivière Method

This method, based on the use of microstrata, was proposed in a large set of publications of Rivière (1998, 1999, 2001a,b) under the framework of the 1996 SUPCOM project of Eurostat. As a result, two software applications were developed: SALOMON in 1998 (see Mészáros, 1999) and MICROSTRAT in 2001. The method is based on four basic ideas:

- the use of PRNs that are allocated to each statistical unit,
- the use of a measure of burden, which can be the number of times that a unit has already been selected for all the waves that one wants to coordinate,
- the use of microstrata constructed at each wave by intersecting all the strata of the waves that one wants to coordinate,
- the permutation of the PRNs in proportion to the measure of burden within the microstrata so that the units with the smallest measures of burden obtain the smallest random numbers.

As a preliminary to the algorithm, each unit receives a PRN from the uniform distribution on [0,1] and a response burden equal to 0. Also note that for the Rivière method, if t is the first wave that we want to coordinate, a microstratum, at wave T, is defined by the intersection of the strata of waves t to $T-1$. The permutations are done within each microstratum according to the cumulative burden. Note that, within the subsets of equal burden, the order of the permuted PRNs must remain unchanged. Then, one can apply Algorithm 3 to do coordination using the Rivière method. A proof of the validity of the method has been given by Bleuer (2002). Nevertheless, if the algorithm is carried out just once, the procedure has a main drawback: by crossing the strata of all the previous surveys, the microstrata become very small and thus the coordination is not very good. For this reason, and in order to have a good coordination with the last surveys, Rivière (1999, p. 5) advocated the use of only three sorts.

Algorithm 3 Negative coordination with the Rivière method.

Assign a PRN ω_k^1 to each unit $k \in U$, i.e. construct $\boldsymbol{\omega}^1 = \{\omega_1^1, \ldots, \omega_N^1\}$.
Assign a burden equal to 0 to each unit $k \in U$, i.e. $c_k^1 = 0$.
for $T = 2, \ldots,$ Number of Waves **do**
  Compute the burden $c_k^T = \sum_{t=1}^{T-1} b^t S_k^t$.
  Construct the microstrata by crossing the strata of waves 1 to $T-1$.
  Permute the ω_k^1, in each microstratum, so that the units are sorted by increasing burden and the ranks remain unchanged in the subsets of equal burden.
  Select the first n_h^T units in each stratum.
end for

Table 5
Definition of strata used for the simulations.

Units	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Strata^1	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
Strata^2	1	1	2	2	1	1	2	2	1	1	2	2	1	1	2	2
Strata^3	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
Strata^4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

We ran a large set of simulations, trying to invalidate the method. We were sceptical about the use of multiple sorting. Nevertheless, after four waves, the method seems to provide the right inclusion probabilities of order 1 and 2, as long as the permutations are done in a strictly sequential manner. For instance, at wave T, if we want to coordinate with respect to wave t, the microstrata must be obtained by crossing all the strata of all the waves between t and T − 1. It is not possible to skip a wave, otherwise the inclusion probabilities are not satisfied, as shown in the following example.

Suppose that we have four waves during the first year. The population is of size $N = 16$. At each wave, we have two strata of size $N_h^t = 8$. The burden is equal to 1 for each survey. The strata are defined in Table 5.

The sample strata sizes are $n_1^1 = 3$, $n_2^1 = 5$, $n_1^2 = 6$, $n_2^2 = 2$, $n_1^3 = 4$, $n_2^3 = 4$, $n^4 = 6$. If we apply the procedure of Salomon, the samples are coordinated as follows.

- For the initialization, uniform random numbers ω_k are generated for each unit.
- At wave 1, the units that have the smallest ω_k in each stratum are selected.
- At wave 2, the ω_k are permuted according to the burden in the strata of wave 1 for obtaining ω_k^2. Next, the units that have the smallest ω_k^2 in each stratum are selected.
- At wave 3, the ω_k^2 are permuted according to the burden in the strata of wave 2 to obtain ω_k^{3a}. Next, the ω_k^{3a} are permuted according to the burden in the crossing of the strata of wave 1 and 2, to obtain ω_k^{3b}. Finally, the units that have the smallest ω_k^{3b} in each stratum are selected.
- At wave 4, the ω_k^{3b} are permuted in function of the burden in the strata of wave 3 for obtaining ω_k^{4a}. Next, the ω_k^{4a} are permuted according to the burden in the crossing of the strata of wave 1 and 3 (**wave 2 is skipped**) to obtain ω_k^{4b}. Finally, the units that have the smallest ω_k^{4b} in each stratum are selected.

However, this procedure is not recommended by Rivière. The permutation is done in relation to the previous survey and in relation to all the surveys since the beginning of the year.

After 10 000 simulations, we estimated the inclusion probabilities by $\widetilde{\pi}_k^t$. Then we computed

$$z_k^4 = \frac{\widetilde{\pi}_k^4 - \pi_k^4}{\sqrt{\pi_k^4(1 - \pi_k^4)/\mathrm{sim}}},$$

where π_k^4 are the inclusion probabilities that we want to obtain, and sim is the number of simulations. If the method provides good inclusion probabilities, then the z_k should have a Normal distribution. We obtained the following vector of z_k^4:

$$(1.033, -6.383, -1.611, 8.138, 1.012, -8.076, -2.520, 7.333,$$
$$2.520, -7.415, -1.384, 6.403, 1.632, -8.882, -0.909, 9.109).$$

Several z_k^4 are larger than 1.96 in absolute value. We must therefore reject the hypothesis that the inclusion probabilities are correct.

This example does not show that the method is false. It only shows that the permutation must be done in a strictly sequential manner and that a wave should never be skipped. We ran a set of simulations in the same population as in Example 1 in order to compare the quality of the coordination of the Rivière and Cotton & Hesse methods. The results, which we do not present here, clearly showed that both methods give almost equivalent results.

8 Other Methods Using Microstrata

In this section, we introduce several new methods based on the idea of microstrata. A method of coordination must be evaluated through several waves, and it is a complicated matter to theoretically prove that a method works or not when we have a large number of waves. Moreover, on a large number of waves and large population and sample sizes, the methods seem to give equivalent results. So, in order to invalidate our methods and point out the differences between them, we decided to run simulations on four waves. Obviously, if we cannot prove by simulation that a method is false, it does not imply that the method works. From this point forward, we will refer to a method as sim-false if it was invalidated by simulations, and sim-correct, if the simulations fail to invalidate it.

Before coming to the idea of applying the microstrata technique for coordination, we had a very simple idea. First, for each wave T, we generate a new random number. Next, these random numbers are permuted within each of the strata of the previous waves. These permutations are done chronologically from wave 1 to wave $T - 1$. The method is described precisely in Algorithm 4.

Algorithm 4 First method of chronological permutations (sim-false).

At wave 1, assign a uniform random number, ω_k^1, to each unit $k \in U$.
Select the units that have the n_h^1 smallest ω_k^1 to obtain the sample $\mathbf{s}^1$.
for $T = 2, \ldots,$ Number of Waves **do**
  Generate and assign a new uniform random number, ω_k^t, to each unit $k \in U$.
  for $t = 1, \ldots, T - 1$
    Permute the ω_k^t, within the strata, so that the units that are selected receive the largest random numbers, the units that are not selected receive the smallest random numbers and the ranks remain unchanged in the subsets of selected and non-selected units.
  end for
  Select the units that have the n_h^T smallest ω_k^t, in each stratum, to obtain the sample $\mathbf{s}^T$.
end for

This method seems interesting but is sim-false in the sense that it does not provide the sim-correct inclusion probabilities. Simulations on at least 3 waves were needed to detect the problem. Our explanation is the following: at wave 3, the random numbers are permuted according to the first wave, and next according to the second wave. Nevertheless, the selection of the units of the second wave depends on the permuted random numbers from the first wave. This correlation implies that, after the permutation, the random numbers are not independent and uniform anymore. From this sim-false method, we concluded that, in order to coordinate a sample at time T, if a permutation of the random numbers is done in the strata of time t, the method will be false. The permutations must be done in the crossing of all the strata (the microstrata) from time t to $T - 1$.

One of the main differences between the methods we introduce is the order in which the permutations are done. We differentiate two types of order: chronological and retrospective.

Chronological means always starting with the first wave and going on to the next ones. As an example of chronological permutations at wave 4, we have: first, the permutations are done in the crossing of the strata of waves 1, 2, and 3, after that in the crossing of the strata of waves 2 and 3, and finally, in the crossing of the strata of wave 3. On the other side, a retrospective order means that the permutations are first done in the crossing of the strata of the latest wave and then going backwards to the first wave. As an example of retrospective permutations at wave 4, we have: first, the permutations are done in the crossing of the strata of wave 3, after that in the crossing of the strata of waves 2 and 3, and finally, in the crossing of the strata of waves 1, 2, and 3. However, the retrospective order has the small disadvantage that it takes more time to compute the permutations than if they are done chronologically.

In order to overcome the problem posed by Algorithm 4, the permutations could be done in the microstrata as described in Algorithm 5. Note, that the permutations are done in chronological order. We can modify the method described in Algorithm 5 by using PRNs instead of generating a new random number at each wave. This method is described in Algorithm 6. The simulations invalidated both methods, so they are both sim-false. We thus concluded that the use of burden in microstrata does not work if the permutations are done in a chronological order.

Algorithm 5 Second method of chronological permutations (sim-false).

At wave 1, assign a uniform random number to each unit $k \in U$.
Select the units that have the n_h^1 smallest ω_k^1 to obtain the sample $\mathbf{s}^1$.
for $T = 2, \ldots,$ Number of Waves **do**
  Assign a new uniform random number to each unit $k \in U$.
  for $t = 1, \ldots, T-1$ **do**
    Compute the $(t, T-1)$-cumulated burden, i.e. $c_k^{t,T-1} = \sum_{u=t}^{T-1} b^u S_k^u$.
    Construct the microstrata by crossing the strata of waves t to $T-1$.
    Permute the ω_k^t, in each microstratum, such that the units are sorted by increasing burden and the ranks remain unchanged in the subsets of equal burden.
  end for
  Select the units that have the n_h^T smallest ω_k^T in each stratum to obtain the sample $\mathbf{s}^T$.
end for

Algorithm 6 Third method of chronological permutations (sim-false).

Assign a PRN ω_k^1 to each unit $k \in U$, i.e. construct $\boldsymbol{\omega}^1 = \{\omega_1^1, \ldots, \omega_N^1\}$.
Select the units that have the n_h^1 smallest ω_k^1 to obtain the sample $\mathbf{s}^1$.
for $T = 2, \ldots,$ Number of Waves **do**
  for $t = 1, \ldots, T-1$ **do**
    Compute the $(t, T-1)$-cumulated burden, i.e. $c_k^{t,T-1} = \sum_{u=t}^{T-1} b^u S_k^u$.
    Construct the microstrata by crossing the strata of waves t to $T-1$.
    Permute the ω_k^t, in each microstratum, so that the units are sorted by increasing burden and the ranks remain unchanged in the subsets of equal burden.
  end for
  Select the units that have the n_h^T smallest ω_k^T in each stratum to obtain the sample $\mathbf{s}^T$.
end for

These conclusions led us to create six other methods that could not be invalidated by simulation. We have abandoned the methods based on generating a new random number at each wave due to the simulation results in which these methods do not perform any better than the PRNs methods. Thus, we will present only the three methods based on PRNs. Method 7 is based on the idea of microstrata, cumulative burden, and multiple permutations, which are, this time, done in a retrospective way. This method can be considered as a modification of the Rivière method. The difference between our method and the Rivière method is that, at a given time T, it uses multiple permutations, whereas the Rivière method uses only one permutation done in the microstrata, constructed by crossing the strata from waves 1 to $T-1$. Based on the simulation results, this method is sim-correct.

Algorithm 7 First method of retrospective permutations (sim-correct).

Assign a PRN ω_k^1 to each unit $k \in U$, i.e. construct $\boldsymbol{\omega}^1 = \{\omega_1^1, \ldots, \omega_N^1\}$.
Select the units that have the n_h^1 smallest ω_k^1 to obtain the sample $\mathbf{s}^1$.
for $T = 2, \ldots,$ Number of Waves **do**
  for $t = T-1, \ldots, 1$ **do**
    Compute the $(t, T-1)$-cumulated burden, i.e. $c_k^{t,T-1} = \sum_{u=t}^{T-1} b^u S_k^u$.
    Construct the microstrata by crossing the strata of waves t to $T-1$.
    Permute the ω_k^t, in each microstratum, so that the units are sorted by increasing burden and the ranks remain unchanged in the subsets of equal burden.
  end for
  Select the units that have the n_h^T smallest ω_k^T in each stratum to obtain the sample $\mathbf{s}^T$.
end for

The last two methods are presented in Algorithms 8 and 9 and are based on the idea of multiple permutations done in microstrata, which are the crossing of the strata of the previous waves and the subsets defined as $\mathbf{s}^T$. In these methods, the cumulative burden is not taken into account. The permutations are done according to the sample indicator variables and not according to the cumulative burden. In Algorithm 8, the permutations are done in a retrospective order, while in Algorithm 9 they are done in a chronological order. Based on the simulation results, these methods are sim-correct.

Algorithm 8 Second method of retrospective permutations (sim-correct).

Assign a PRN ω_k^1 to each unit $k \in U$, i.e. construct $\boldsymbol{\omega}^1 = \{\omega_1^1, \ldots, \omega_N^1\}$.
Select the units that have the n_h^1 smallest ω_k^1 to obtain the sample $\mathbf{s}^1$.
for $T = 2, \ldots,$ Number of Waves **do**
  for $t = T-1, \ldots, 1$ **do**
    Construct the microstrata as the intersection of the crossing of the strata $t, \ldots, T-1$ and the crossing of the subsets defined by $\mathbf{s}^{t+1}, \ldots, \mathbf{s}^{T-1}$.
    Construct $\boldsymbol{\omega}^{t+1}$ by permuting $\boldsymbol{\omega}^t$, within each microstratum, so that the units that are selected receive the largest random numbers, the units that are not selected receive the smallest random numbers and the ranks remain unchanged in the subsets of selected and non-selected units.
  end for
  Select the units that have the n_h^T smallest ω_k^T in each stratum to obtain the sample $\mathbf{s}^T$.
end for

---

9 Simulation Study and Results

In this section, we will test the new methods presented in Section 8. We should note that simulations were done in larger sample (population) sizes but in this case all methods were performing well. Thus, in order to find which methods result in false inclusion probabilities, we decided to use small sample (population) sizes. We simulated 500 000 drawings of stratified simple random samples in a population of $N = 16$ units. Four waves were taken into account. The strata are defined in Table 5. The sample strata sizes are

$$n_1^1 = 3,\ n_2^1 = 5,\ n_1^2 = 6,\ n_2^2 = 2,\ n_1^3 = 4,\ n_2^3 = 4,\ n^4 = 6.$$

Algorithm 9 Fourth method of chronological permutations (sim-correct).

Assign a PRN ω_k^1 to each unit $k \in U$, i.e. construct $\boldsymbol{\omega}^1 = \{\omega_1^1, \ldots, \omega_N^1\}$.
Select the units that have the n_h^1 smallest ω_k^1 to obtain the sample $\mathbf{s}^1$.
for $T = 2, \ldots,$ Number of Waves **do**
  for $t = 1, \ldots, T-1$ **do**
    Construct the microstrata as the intersection of the crossing of the strata $t, \ldots, T-1$ and the crossing of the subsets defined by $\mathbf{s}^{t+1}, \ldots, \mathbf{s}^{T-1}$.
    Construct $\boldsymbol{\omega}^{t+1}$ by permuting $\boldsymbol{\omega}^t$, within each microstratum, so that the units that are selected receive the largest random numbers, the units that are not selected receive the smallest random numbers and the ranks remain unchanged in the subsets of selected and non-selected units.
  end for
  Select the units that have the n_h^T smallest ω_k^T in each stratum to obtain the sample $\mathbf{s}^T$.
end for

To compare the results of the simulations, we decided to analyze three different simulation outputs:

1. The first-order inclusion probabilities.
2. The second-order inclusion probabilities.
3. The quality of the coordination.

To analyze the first- and second-order inclusion probabilities, which we denote, respectively, by π_k^{sim} and $\pi_{k\ell}^{sim}$, we calculated a kind of "z-value", which enables us to do a Normal test on the value obtained by simulation. This "z-value" was obtained using the following formula:

For the first-order inclusion probabilities:

$$z_{\pi_k} = \sqrt{sim} \cdot \frac{\pi_k^{sim} - \left(\frac{n_h}{N_h}\right)}{\sqrt{\frac{n_h}{N_h} \cdot \left(1 - \frac{n_h}{N_h}\right)}}$$

For the second-order inclusion probabilities:

$$z_{\pi_{k\ell}} = \sqrt{sim} \cdot \frac{\pi_{k\ell}^{sim} - \left(\frac{n_h^k \cdot n_h^\ell}{N_h^k \cdot N_h^\ell}\right)}{\sqrt{\frac{n_h^k \cdot n_h^\ell}{N_h^k \cdot N_h^\ell} \cdot \left(1 - \frac{n_h^k \cdot n_h^\ell}{N_h^k \cdot N_h^\ell}\right)}}$$

Table 6
Wave number for each plot.

1	2
3	4

The obtained values are the "z-values" for the centered inclusion probabilities. An acceptable value of the centered π_k and $\pi_{k\ell}$ should lie in the interval $[-2,2]$ (95% confidence interval).

On each graph, there are four plots corresponding to each of the four waves, as shown in Table 6.

On each plot we can see:

- on the horizontal axis - the units of the population,
- on the vertical axis - the centered "z-value",
- the limit values of the confidence interval, given by dashed lines,
- the acceptable values - the circles between the dashed lines,
- the unacceptable values - the circles outside of the confidence interval.

The third output used to compare the methods is the quality of the coordination which is simply given by the average number of common units in the samples.

To interpret the results, we shall analyze each of the graphs. It is very important to note that, for all methods, the results are always correct for the first two waves. Thus, we will be considering only the last two waves.

On Figures 1 and 2, the first- and second-order inclusion probabilities, respectively, for Algorithms 4, 5, 6 and Algorithms 7, 8, 9 are plotted. We can see that:

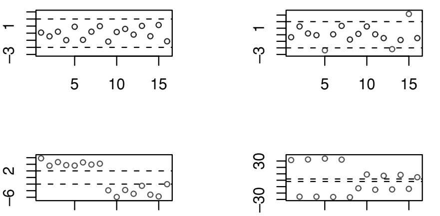
- Most or all of the first-order inclusion probabilities lie outside the confidence interval.
- Most of the second-order inclusion probabilities lie outside the confidence interval.

In conclusion, we can say that according to the graphs, the methods given by Algorithms 4, 5, and 6 are sim-false methods.

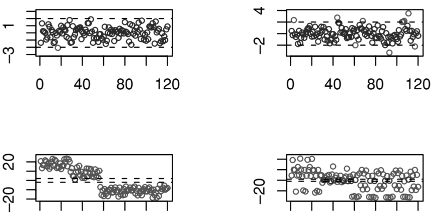
The quality of the coordination is given in Table 7. As we are coordinating negatively, the aim is to minimize the expected overlap, i.e. the number of common units in the samples. First, we compare the overlap between the samples of waves 3 and 4. If it is approximately the same, then we compare the overlap between the samples of waves 2 and 4. On this basis, we can conclude that the coordination works equally good for the Algorithms 7, 8, and 9.

10 Conclusions

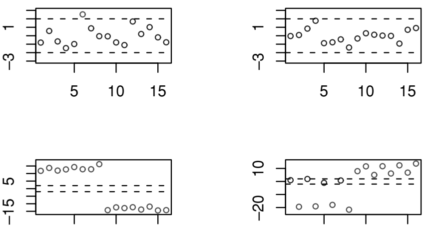
The Kish & Scott, Cotton & Hesse, and Rivière methods allow the definition of the strata to be changed, which enables us to create a dynamic system of coordination. In the case of negative coordination, two bounds can be used as benchmarks for comparing the quality of the coordination: the absolute lower bound (ALB), which is rarely reached, and the relative lower bound (RLB), defined as the solution of the linear program. A simple counter-example shows that neither the Kish & Scott nor the Cotton & Hesse method allows us to reach the RLB. Nevertheless, no other solution has been proposed to avoid the enumeration of all possible samples.



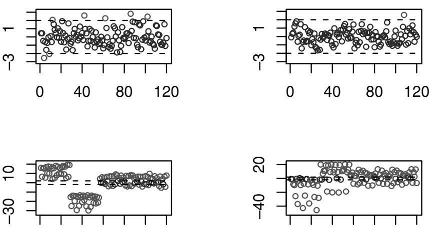
(a) First-order inclusion probabilities for Algorithm 4



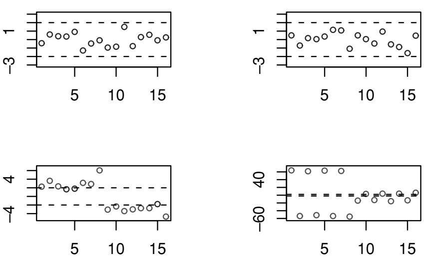
(b) Second-order inclusion probabilities for Algorithm 4



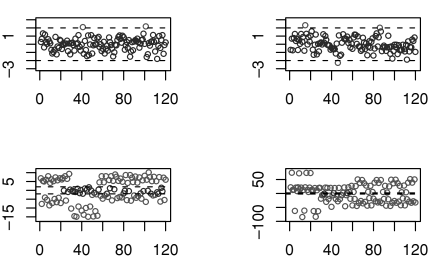
(c) First-order inclusion probabilities for Algorithm 5



(d) Second-order inclusion probabilities for Algorithm 5



(e) First-order inclusion probabilities for Algorithm 6

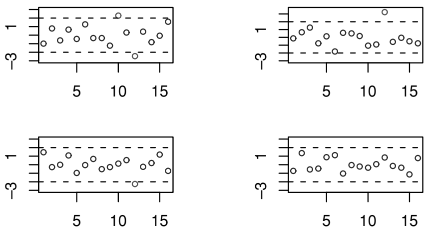


(f) Second-order inclusion probabilities for Algorithm 6

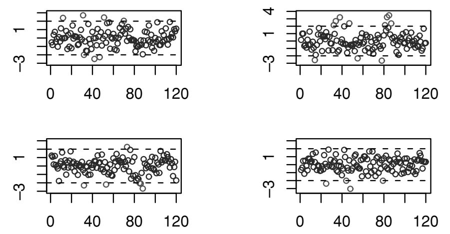
Figure 1. First- and second-order inclusion probabilities for Algorithms 4, 5, 6.

Based on the simulation results, we can see that the quality of coordination of the Kish & Scott method is worse than that of the Cotton & Hesse method. Moreover, the Kish & Scott method does not allow more than two samples to be coordinated. For these reasons, the Cotton & Hesse method should be preferred to the Kish & Scott method.

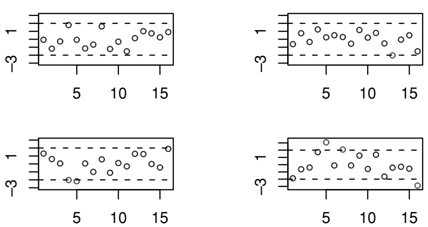
The Rivière method is slightly more complex to implement but, at present, the capacity of computation does not constitute a barrier to rapid implementation. We show that the condition for validity of the method is that the crossing of the strata is done for all the waves since the first survey that we want to coordinate. This limits our capacity to coordinate with really old surveys. The problem that can occur is that there are no more or very few units left in the microstrata. In a simulation based on four waves, the quality of the coordination of the Cotton & Hesse method is



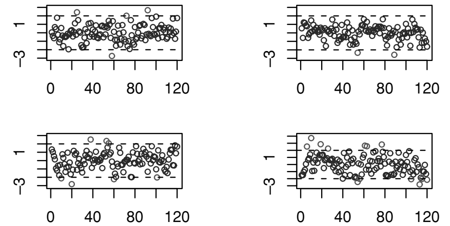
(a) First-order inclusion probabilities for Algorithm 7



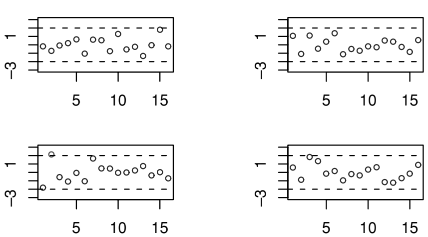
(b) Second-order inclusion probabilities for Algorithm 7



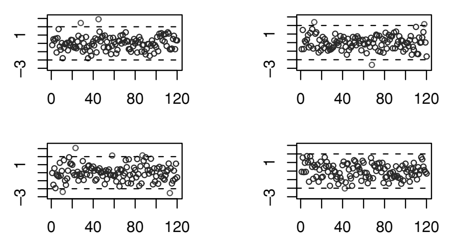
(c) First-order inclusion probabilities for Algorithm 8



(d) Second-order inclusion probabilities for Algorithm 8



(e) First-order inclusion probabilities for Algorithm 9



(f) Second-order inclusion probabilities for Algorithm 9

Figure 2. First- and second-order inclusion probabilities for Algorithms 7, 8, 9.

the same as that of the Rivière method. The concordance is so accurate that we could conjecture that both methods provide the same joint sampling design.

However, the Rivière method is more flexible, the use of burden allows us to give more importance to chosen surveys. The burden can also change over time, for instance, if we want to have a positive coordination. Nevertheless, we do not understand why, in the implementation of MICROSTRAT and SALOMON software, only three sorts are done. We should rather advocate the use of one sort per wave, which seems to be possible with current computers, even with several decades of waves.

We have seen that it is easy to construct new methods because the ideas on which they are based are simple and intuitive. The method presented in Algorithm 4 is a non-PRN method where, at each wave, a new random number is generated for all the units of the population. The methods given in Algorithms 5 and 6 can be seen as modifications of the Rivière method. Because of their simplicity, it is difficult to understand why these methods do not work. After

Table 7
Expected overlaps.

	Algorithm 7			
	wave 1	wave 2	wave 3	wave 4
wave 1	8.000	2.159	4.027	3.085
wave 2	2.159	8.000	2.073	3.941
wave 3	4.027	2.073	8.000	0.126
wave 4	3.085	3.941	0.126	6.000
	Algorithm 8			
	wave 1	wave 2	wave 3	wave 4
wave 1	8.000	2.161	4.028	3.086
wave 2	2.161	8.000	2.071	3.945
wave 3	4.028	2.071	8.000	0.125
wave 4	3.086	3.945	0.125	6.000
	Algorithm 9			
	wave 1	wave 2	wave 3	wave 4
wave 1	8.000	2.163	4.024	3.088
wave 2	2.163	8.000	2.073	3.943
wave 3	4.024	2.073	8.000	0.125
wave 4	3.088	3.943	0.125	6.000

Table 8
Summary table.

Algorithm	Burden	Sample	Retrospective	Chronological	Sim-false
3	Yes	No	Yes	No	No
4	No	No	No	Yes	Yes
5	Yes	No	No	Yes	Yes
6	Yes	No	No	Yes	Yes
7	Yes	No	Yes	No	No
8	No	Yes	Yes	No	No
9	No	Yes	No	Yes	No

running simulations on only four waves, we showed that they are sim-false. A summary of the tested methods is given in Table 8.

We have also seen that there is a way to construct modifications of the Cotton & Hesse method while permuting the random numbers in the crossing of all the strata. These methods have proven to be sim-correct.

Although it is difficult to give preference to one or another of the newly introduced methods, we believe that the method presented in Algorithm 8, which proved to be sim-correct, can be a good solution to the sample coordination problem. Like the Rivière method, it is based on the use of PRNs and is retrospective. Like the Cotton & Hesse method, the permutations are done while respecting the ranks in the vector of random numbers. Its innovation comes from the way in which the microstrata are constructed.

The general conclusion is that using methods based on microstrata makes coordination with old surveys very difficult because of the need to cross all the intermediate strata, which finally results in a very small sample size in the microstrata. Unfortunately, this seems to be a constraint of the coordination problem that cannot be ignored.

Acknowledgements

The authors wish to thank the Editor and anonymous referees for their helpful suggestions and comments that helped improve this article. The authors are also grateful to the Swiss National

Science Foundation (grant FN205121-105187/1) and the Swiss Federal Statistical Office for financial support. The views expressed herein are those of the authors and not necessarily those of the Swiss Federal Statistical Office or of the Swiss National Science Foundation.

References

Bleuer, S. R. (2002). Report on Rivière's random permutations method of sampling co-ordination. Report 2002-07-26. Statistics Canada.

Brewer, K., Early, L. & Joyce, S. (1972). Selecting several samples from a single population. Aust. J. Stat., **3**, 231–239.

Cotton, F. & Hesse, C. (1992). Tirages coordonnes d'échantillons. Document de travail de la Direction des Statistiques Économiques E9206. Tech. rep., INSEE, Paris.

De Ree, S. (1983). A system of co-ordinated sampling to spread response burden of enterprises, Contributed paper, 44th Session of the ISI Madrid, pp. 673–676.

Keyfitz, N. (1951). Sampling with probabilities proportional to size: adjustmentfor changes in the probabilities. J. Amer Statist. Assoc., **46**, 105–109.

Kish, L. & Scott, A. (1971). Retaining units after changing strata and probabilities. J. Amer Statist. Assoc., **66**, 461–470.

Koeijers, E. & Willeboordse, A. (1995). Reference manual on design and implementation of business surveys. Tech. rep., Statistics Netherlands.

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

Mészáros, P. (1999). A program for sample co-ordination: Salomon. In Proceedings of the International Seminar on Exchange of Technology and Knowhow, pp. 125–130. Prague.

Ohlsson, E. (1995). Business Survey Methods. Vol. 1, chap. 9, "Coordination of samples using permanent random numbers", pp. 153–169. New York: Wiley, Inc.

Patterson, H. (1950). Sampling on successive occasions with partial replacement of units. J. Roy. Statist. Soc., **B12**, 241–255.

Rivière, P. (1998). Description of the chosen method: Deliverable 2 of supcom 1996 project (part "Co-ordination of samples"), pp. 11–33. Report, EUROSTAT.

Rivière, P. (1999). Coordination of samples: the microstrata methodology. 13th International Roundtable on Business Survey Frames, Paris: INSEE.

Rivière, P. (2001a). Coordinating samples using the microstrata methodology. In Proceedings of Statistics Canada Symposium2001.

Rivière, P. (2001b). Random permutations of random vectors as a way of co-ordinating samples. Report, University of Southampton, INSEE.

Rosén, B. (1997a). Asymptotic theory for order sampling. J. Statist. Plan. Infer., **62**, 135–158.

Rosén, B. (1997b). On sampling with probability proportional to size. J. Statist. Plan. Infer., **62**, 159–191.

Van Huis, L., Koeijers, C. & de Ree, S. (1994a). EDS, sampling system for the central business register at Statistics Netherlands. Tech. rep., Statistics Netherlands.

Van Huis, L., Koeijers, C. & de Ree, S.J.M. (1994b). Response burden and co-ordinated sampling for economic surveys. Tech. rep., Statistics Netherlands, Volume 9.

Résumé

Le but de la coordination d'échantillons est de maximiser ou minimiser le recouvrement de plusieurs échantillons à l'intérieur d'une population qui évolue au fil des temps. Pour effectuer une coordination, la sélection d'un nouvel échantillon dépendra donc des échantillons précédemment tirés. Afin d'obtenir un recouvrement plus fort ou plus faible que celui fourni par des tirages indépendants, une dépendance entre les échantillons doit être introduite. Cette dépendance va augmenter ou limiter le nombre d'unités communes à tous les échantillons sélectionnés. Plusieurs méthodes pour coordonner des échantillons stratifiés ont déjà été développées. Parmi eux les méthodes de Kish and Scott, de Cotton and Hesse, et de Rivière sont présentées en détail. En utilisant des simulations, on compare l'optimalité et la qualité de la coordination pour chacune de ces trois méthodes. On présente six nouvelles méthodes basées sur l'utilisation de nombres aléatoires permanents et des microstrates et on essaye de les valider à l'aide des simulations.