Component of Statistics Canada
Catalogue no. 12-001-X Business Survey Methods Division

Article

Alternative survey sample designs: Sampling with multiple overlapping frames

by Sharon L. Lohr

December 2011

Statistics Canada Statistique Canada

Canada

Survey Methodology, December 2011
Vol. 37, No. 2, pp. 197-213
Statistics Canada, Catalogue No. 12-001-X

Alternative survey sample designs: Sampling with multiple overlapping frames

Sharon L. Lohr [1]

Abstract

Designs and estimators for the single frame surveys currently used by U.S. government agencies were developed in response to practical problems. Federal household surveys now face challenges of decreasing response rates and frame coverage, higher data collection costs, and increasing demand for small area statistics. Multiple frame surveys, in which independent samples are drawn from separate frames, can be used to help meet some of these challenges. Examples include combining a list frame with an area frame or using two frames to sample landline telephone households and cellular telephone households. We review point estimators and weight adjustments that can be used to analyze multiple frame surveys with standard survey software, and summarize construction of replicate weights for variance estimation. Because of their increased complexity, multiple frame surveys face some challenges not found in single frame surveys. We investigate misclassification bias in multiple frame surveys, and propose a method for correcting for this bias when misclassification probabilities are known. Finally, we discuss research that is needed on nonsampling errors with multiple frame surveys.

Key Words: Bias correction; Dual frame survey; Misclassification; Mode effects; Sampling for rare events; Sampling weights; Small area estimation.

1. Uses of multiple frame surveys

In classical design-based sampling theory, a probability sample is taken from the (single) sampling frame, and the inclusion probabilities in the sampling design can be used to make inferences about the population. Let y_i be a measurement on unit i in the population of N units, let $\mathcal{S}$ denote the set of units in the sample, and let $\pi_i = P$ (unit i is included in the sample). Then the Horvitz-Thompson (1952) estimator of the population total $Y = \sum_{i=1}^{N} y_i$ is $\hat{Y} = \sum_{i \in \mathcal{S}} w_i y_i$, where $w_i = 1 / \pi_i$ is the sampling weight. If the sampling frame includes everyone in the target population, all sampled units respond, and there is no measurement error, then the Horvitz-Thompson estimator is unbiased for Y.

The practical challenges of sampling in the 1940s and 1950s drove the methodological developments of stratified multistage surveys and estimators such as the Horvitz-Thompson estimator. In-person surveys relied on unequal probability sampling to balance interviewer workloads and reduce variances. Response rates were high in many government surveys so that the assumptions for the Horvitz-Thompson estimator were reasonable. We now face new challenges in household surveys. Nonresponse rates are increasing, which means that survey estimates rely more on models. The ethnic and language diversity of a population can result in undercoverage and measurement error. Increasing technological diversity means that different residents may be best reached by different sampling modes; one must then be confident that different sampling modes measure the same quantities. Costs of collecting data have risen greatly, in part due to increasing nonresponse; at the same time, governmental and research demands for data have also risen greatly.

Multiple frame surveys can achieve better population coverage at lower cost. They can be used as part of a structure of modular survey design that relies on different sampling frames to help reduce costs and achieve better coverage. They can also use administrative data efficiently. In this paper, we describe different types of multiple frame surveys and discuss some of the research that is completed and research that may be needed for their use.

One of the earliest multiple frame surveys (aside from early capture-recapture methods) was performed by the Census Bureau in 1949 (Hansen, Hurwitz and Madow 1953). In the Sample Survey of Retail Stores, a probability sample of primary sampling units (psus) was chosen. Within each psu, a list of large retail firms was constructed from records of the Old Age and Survivors Insurance Bureau. All firms on the list were sampled, and an area sample of firms in the psu that were not on the list was taken. In this case, a *screening* dual frame design was employed within each selected psu; units in the list frame were screened out of the area frame before sampling. Thus, the estimator of total sales summed the two estimators within each psu. No new statistical methods were required to estimate total sales in this survey, since essentially a stratified sample was taken in each psu: the firms on the list in the psu formed one stratum, and the firms in the area frame but not on the list in the psu formed the second stratum. The survey resulted in cost savings because it was relatively inexpensive to sample from the firms on the list, yet full coverage was obtained by also using the area frame.

1. Sharon L. Lohr, School of Mathematical and Statistical Sciences, Arizona State University, Tempe AZ 85287-1804. E-mail: sharon.lohr@asu.edu.

Many agricultural surveys also have used a screening dual frame survey design (González-Villalobos and Wallace 1996). In such a design, farms belonging to the list frame are removed from the area frame before sampling commences. Considerable cost savings can be realized since often the list frame is much less expensive to sample and it contains the largest farms.

In many cases, however, it may not be possible or practical to remove list-frame units from the area frame before sampling. Instead, in an overlapping dual frame survey, independent probability samples are taken from frame A (the area frame) and frame B (the list frame); this is depicted in Figure 1. Rare populations can often be sampled more efficiently using a multiple frame sample (Kalton and Anderson 1986). In an epidemiology study, for example, frame A might be that used for a general population health survey, while frame B might be a list frame of clinics specializing in a certain disease. The sample from frame B is expected to yield a high percentage of persons with the disease of interest, so that sampling will be efficient; the sample from frame A, though more expensive, leads to complete coverage of the population.

In other situations, all frames are incomplete, as considered by Hartley (1962); for example, frame A in Figure 2 might be a frame of landline telephones and frame B might consist of cellular telephone numbers. There are three domains: domain a consists of units in frame A but not in frame B, domain b consists of units in frame B but not in frame A, and domain ab consists of units in both frames. In the telephone context, domain a contains individuals belonging to a landline-only household, domain b consists of individuals who have only a cellular telephone, and domain ab consists of individuals who have both cellular and landline telephones. It is unknown in advance whether a household member sampled using one frame also belongs to the other frame (Brick, Dipko, Presser, Tucker and Yuan 2006); typically, respondents are asked about their cellular and landline telephone usage to determine domain membership.

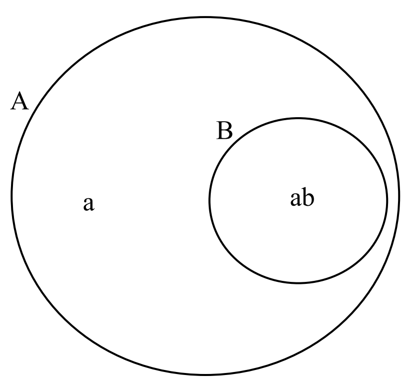

Figure 1 A dual frame design in which frame B is a subset of frame A

More than two frames can be employed as well, as illustrated in Figure 3 for a three-frame survey in which all frames are incomplete. In this situation, there are seven domains. Iachan and Dennis (1993) gave an example of a three-frame survey used to sample the homeless population, where frame A is a list of soup kitchens, frame B is a list of shelters, and frame C consists of street locations. Figure 4 displays a 3-frame survey in which frame A has complete coverage, while overlapping frames B and C are both incomplete but are less expensive to sample. This design has been used for the U.S. Scientists and Engineers Statistical Data System (SESTAT; National Science Foundation 2003) surveys. The same design might be used when A is the frame for a general population survey, B is a landline telephone survey, and C is a cell phone survey.

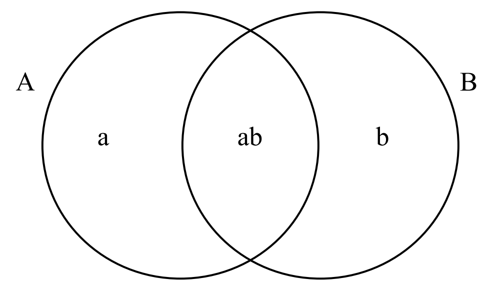

Figure 2 Frames A and B overlap, creating the three domains *a*, *b*, and *ab*

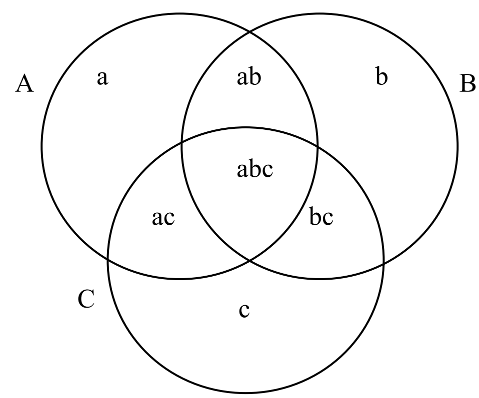

Figure 3 Frames A, B, and C are all incomplete and overlap

There is much potential for using multiple frame designs in household surveys, including:

1. Use of multiple list frames from administrative records.
2. Multiple mode sampling (for example, using independent samples from a cellular telephone frame and a landline telephone frame).

3. Future use of the internet for data collection. Although the internet presents many coverage and domain specification challenges, it is worthy of consideration because of the potential cost savings and ease of data collection and processing.
4. Improved small area estimation. A national survey is supplemented with smaller, localized surveys to obtain higher precision in those areas.
5. Improved estimation for rare populations. A general population survey may be supplemented by a survey from a frame with a high concentration of members of the rare population.
6. Modular survey design. A multiple frame approach can give more flexibility for design of continuing surveys. As particular frames become less expensive to sample, the relative allocation of sample size to the different frames can be modified. The modular approach also allows more flexibility in responding to changing needs for data.

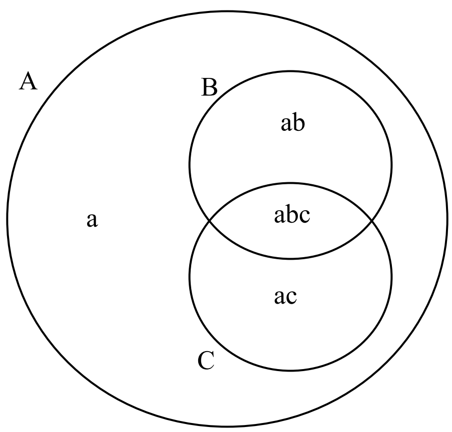

Figure 4 Frame A contains the entire population; frames B and C overlap and are both contained in frame A

The increased flexibility of multiple frame surveys comes at the cost of additional complexity, however. Information from the surveys must be combined to estimate population quantities, and there are many options for estimators. Section 2 summarizes estimators that have been developed for population totals and describes how these modify the sampling weights; Sections 3 and 4 discuss weight calibration and describe how to use survey software packages with multiple frame survey data. Nonsampling errors need to be considered in each frame singly, and in terms of their effect on estimates calculated from the combined information. Section 5 discusses effects of nonresponse and mode effects in multiple frame surveys.

In addition to the nonresponse, undercoverage, and measurement error problems that plague single frame surveys, multiple frame surveys may have domain misclassification. The weight modifications for the estimators in Section 2 depend on the domain membership of the observations. If some observations in domain a are likely to be mistakenly recorded as belonging to domain ab, estimators may have substantial bias. We study effects of domain misclassification in Section 6, and propose a new method for adjusting for misclassification bias when misclassification probabilities are known. Finally, Section 7 discusses design issues and Section 8 discusses the potential and challenges of multiple frame surveys.

2. Estimators in overlapping multiple frame surveys

In this section we review estimators for the population total Y from overlapping multiple frame surveys, along with the weight modifications induced by these estimators. For simplicity of notation, we concentrate on dual frame surveys in Section 2.1, and outline extensions to multiple frame surveys in Section 2.2. In a dual frame survey, we can write

$$Y = Y_a + Y_{ab} + Y_b,$$

where Y_a is the total of the population units in domain a, Y_{ab} is the total of the population units in domain ab, and Y_b is the total of the population units in domain b. A special case is estimating the population size $N = N_a + N_{ab} + N_b$, as discussed in Haines and Pollock (1998). We discuss estimating population quantities other than totals and means, and using data from multiple frame surveys in other analyses, in Section 4.

We first set out some desirable properties for estimators from multiple frame surveys.

1. An estimator should be approximately unbiased for the corresponding finite population quantity.
2. Estimators should be internally consistent: that is, if $\hat{Y}_1$ estimates the number of female engineers in the population, $\hat{Y}_2$ estimates the number of male engineers in the population, and $\hat{Y}_3$ estimates the total number of engineers in the population, then we should have $\hat{Y}_1 + \hat{Y}_2 = \hat{Y}_3$. Internal consistency preserves multivariate relationships in the data. In practical terms, internal consistency requires that one set of weights be used for all estimates.
3. An estimator should be efficient, with low mean squared error.
4. An estimator should be of a form that can be calculated with standard survey software such as SUDAAN or SAS PROC SURVEYMEANS. This allows analysts to work with the data without having to write and test new software. In practical

terms, one data file is created from the multiple frame survey. The file includes one column of weights to be used for calculating point estimates, and it contains either variables describing the survey designs for formula-based variance estimation, or columns of replicate weights for replication-based variance estimation.

5. An estimator should, if possible, be robust to non-sampling errors that might occur with multiple frame surveys.

2.1 Estimators and weight adjustments for dual frame surveys

Consider the overlapping dual frame survey depicted in Figure 2, where domain ab is nonempty. A probability sample $\mathcal{S}(A)$ of size n_A is drawn from the N_A units in frame A, and an independent probability sample $\mathcal{S}(B)$ of size n_B is drawn from the N_B units in frame B. Unit i in sample $\mathcal{S}(A)$ has probability of inclusion π_i^A and weight w_i^A, and unit i in sample $\mathcal{S}(B)$ has probability of inclusion π_i^B and weight w_i^B. The weights may be the inverses of the inclusion probabilities, or they may be poststratified to agree with population counts; it is assumed that estimators of population totals are approximately unbiased.

Then $E[\sum_{i\in\mathcal{S}(A)} w_i^A y_i] \approx Y_a + Y_{ab}$ and $E[\sum_{i\in\mathcal{S}(B)} w_i^B y_i] \approx Y_b + Y_{ab}$. Consequently, an estimator that combines the observations from both surveys with the original weights, $\sum_{i\in\mathcal{S}(A)} w_i^A y_i + \sum_{i\in\mathcal{S}(B)} w_i^B y_i$, is biased for the population total Y. If the domain means differ, the corresponding estimator of the population mean may also be biased.

The various estimators for the population total Y that have been proposed in the literature modify the weights so that the estimators are approximately unbiased. The modified weights, shown below for the different estimators, are of the form $\tilde{w}_i^A = m_i^A w_i^A$ and $\tilde{w}_i^B = m_i^B w_i^B$. The population total is then estimated by

$$\hat{Y} = \sum_{i\in\mathcal{S}(A)} \tilde{w}_i^A y_i + \sum_{i\in\mathcal{S}(B)} \tilde{w}_i^B y_i \tag{1}$$

and the population mean $\bar{Y}$ is estimated by $\hat{\bar{Y}} = \hat{Y} / \hat{N}$ where

$$\hat{N} = \sum_{i\in\mathcal{S}(A)} \tilde{w}_i^A + \sum_{i\in\mathcal{S}(B)} \tilde{w}_i^B.$$

The estimators will be approximately unbiased, then, if $m_i^A \approx 1$ for $i \in a$, $m_i^B \approx 1$ for $i \in b$, and $m_i^A + m_i^B \approx 1$ for $i \in ab$. All of the estimators reviewed in this section satisfy the criteria needed for approximate unbiasedness in the absence of nonsampling errors (see Lohr 2009).

Fixed weight adjustments. The simplest weight modification to preserve approximate unbiasedness, described by Hartley (1962), takes

$$m_{i,\theta}^A = \begin{cases} 1 & \text{if } i \in a \\ \theta & \text{if } i \in ab, \end{cases} \quad m_{i,\theta}^B = \begin{cases} 1 & \text{if } i \in b \\ 1-\theta & \text{if } i \in ab, \end{cases} \tag{2}$$

where $\theta \in [0, 1]$. Using the modified weights $\tilde{w}_i^A = m_{i,\theta}^A w_i^A$ and $\tilde{w}_i^B = m_{i,\theta}^B w_i^B$ in (1), the resulting estimator $\hat{Y}(\theta)$ can also be expressed using the estimated domain totals $\hat{Y}_a^A = \sum_{i\in\mathcal{S}(A), i\in a} w_i^A y_i$, $\hat{Y}_{ab}^A = \sum_{i\in\mathcal{S}(A), i\in ab} w_i^A y_i$, $\hat{Y}_{ab}^B = \sum_{i\in\mathcal{S}(B), i\in ab} w_i^B y_i$, and $\hat{Y}_b^B = \sum_{i\in\mathcal{S}(B), i\in b} w_i^B y_i$. The estimator

$$\begin{aligned}\hat{Y}(\theta) &= \sum_{i\in\mathcal{S}(A)} m_{i,\theta}^A w_i^A y_i + \sum_{i\in\mathcal{S}(B)} m_{i,\theta}^B w_i^B y_i \\ &= \hat{Y}_a^A + \theta \hat{Y}_{ab}^A + (1-\theta)\hat{Y}_{ab}^B + \hat{Y}_b^B \end{aligned} \tag{3}$$

thus estimates the domain total Y_{ab} by a weighted average of the frame A estimator, $\hat{Y}_{ab}^A$, and the frame B estimator, $\hat{Y}_{ab}^B$.

For a fixed value of θ, the estimator $\hat{Y}(\theta)$ gives internal consistency since the same set of adjusted weights is used for all variables. The estimator is simple to use and implement. The efficiency of the estimator depends on the value chosen for θ. Brick *et al.* (2006) used $\theta = 1/2$ in their study of a dual frame survey in which frame A was a landline telephone frame and frame B was a cellular telephone frame, and the value of $\theta = 1/2$ is frequently recommended (see, for example, Mecatti 2007). When $\theta =$ 0 or 1, the data in the overlap domain from one of the samples are discarded and the survey becomes a screening dual frame survey.

Optimal estimators. Hartley (1962, 1974) proposed choosing θ in (3) so that the variance of $\hat{Y}(\theta)$ would be minimized. The optimizing value of θ is

$$\theta_H = \frac{V(\hat{Y}_{ab}^B) + \text{Cov}(\hat{Y}_b^B, \hat{Y}_{ab}^B) - \text{Cov}(\hat{Y}_a^A, \hat{Y}_{ab}^A)}{V(\hat{Y}_{ab}^A) + V(\hat{Y}_{ab}^B)}.$$

Since the variances and covariances are generally unknown, they must be estimated from the data, giving

$$\hat{\theta}_H = \frac{\hat{V}(\hat{Y}_{ab}^B) + \widehat{\text{Cov}}(\hat{Y}_b^B, \hat{Y}_{ab}^B) - \widehat{\text{Cov}}(\hat{Y}_a^A, \hat{Y}_{ab}^A)}{\hat{V}(\hat{Y}_{ab}^A) + \hat{V}(\hat{Y}_{ab}^B)}.$$

Skinner and Rao (1996) showed that Hartley's estimator can be calculated using adjusted weights. The weight modifications for Hartley's estimator $\hat{Y}(\hat{\theta}_H)$ are given by (2), substituting $\hat{\theta}_H$ for θ. Since $\hat{\theta}_H$ is consistent for θ_H, Hartley's estimator is asymptotically optimal among all estimators of the form $\hat{Y}_a^A + \hat{Y}_b^B + \theta\hat{Y}_{ab}^A + (1-\theta)\hat{Y}_{ab}^B$. The modified weights $\tilde{w}_{i,H}^A$ and $\tilde{w}_{i,H}^B$ are functions of the variances and covariances of estimated domain totals, however. This has two consequences: (1) the modified weights are random variables, and their variability needs to be accounted for in standard errors of estimators, and (2) the optimal weight modifications will differ for different response variables, leading to internal inconsistency.

Fuller and Burmeister (1972) proposed modifying Hartley's estimator by using additional information about N_{ab}, giving

$$\hat{Y}_{\text{FB}}(\beta) = \hat{Y}_a^A + \hat{Y}_b^B + \beta_1 \hat{Y}_{ab}^A + (1 - \beta_1)\hat{Y}_{ab}^B + \beta_2(\hat{N}_{ab}^A - \hat{N}_{ab}^B).$$

As with Hartley's estimator, the optimal values $\beta_{1\text{opt}}$ and $\beta_{2\text{opt}}$ are chosen to minimize the variance of $\hat{Y}_{\text{FB}}(\boldsymbol{\beta})$, and are thus functions of the covariances of the domain totals. Substituting consistent estimators $\hat{\beta}_{1\text{opt}}$ and $\hat{\beta}_{2\text{opt}}$ gives the weight adjustments for w_i^A and w_i^B. Lohr and Rao (2000) showed that the Fuller-Burmeister estimator $\hat{Y}_{\text{FB}}$ has the smallest asymptotic variance among the estimators considered. As with the Hartley estimator, however, the modified weights are random variables that differ for different responses, and in complex sampling designs the Fuller-Burmeister estimator is also internally inconsistent.

Pseudo-maximum likelihood (*PML*) *estimators*. To achieve internal consistency Skinner and Rao (1996) proposed a pseudo-maximum likelihood (PML) estimator that uses the same weights for all variables. When N_{ab} is unknown, it is estimated by $\hat{N}_{ab}^{\text{PML}}(\theta)$, which is the smaller of the roots of the quadratic equation

$$\left[\frac{\theta}{N_B} + \frac{1-\theta}{N_A}\right] x^2 - \left[1 + \theta\,\frac{\hat{N}_{ab}^A}{N_B} + (1-\theta)\,\frac{\hat{N}_{ab}^B}{N_A}\right] x + \theta \hat{N}_{ab}^A + (1-\theta)\hat{N}_{ab}^B = 0.$$

Skinner and Rao (1996) suggested using the value θ_P for θ that minimizes the asymptotic variance of $\hat{N}_{ab}^{\text{PML}}(\theta)$:

$$\theta_P = \frac{N_a N_B V(\hat{N}_{ab}^B)}{N_a N_B V(\hat{N}_{ab}^B) + N_b N_A V(\hat{N}_{ab}^A)}. \quad (4)$$

Substituting an estimator $\hat{\theta}_P$ for θ_P, the weight adjustments are:

$$m_{i,P}^A = \begin{cases} \dfrac{N_A - \hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P)}{\hat{N}_a^A} & \text{if } i \in a \\[2ex] \dfrac{\hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P)}{\hat{\theta}_P \hat{N}_{ab}^A + (1-\hat{\theta}_P)\hat{N}_{ab}^B}\hat{\theta}_P & \text{if } i \in ab, \end{cases}$$

$$m_{i,P}^B = \begin{cases} \dfrac{N_B - \hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P)}{\hat{N}_b^B} & \text{if } i \in b \\[2ex] \dfrac{\hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P)}{\hat{\theta}_P \hat{N}_{ab}^A + (1-\hat{\theta}_P)\hat{N}_{ab}^B}(1-\hat{\theta}_P) & \text{if } i \in ab. \end{cases}$$

If the value of θ_P cannot be estimated, for example if the two sampling frames coincide or the design in Figure 1 is used, then one can use an average design effect from each survey in the adjustment, as described in Lohr and Rao (2006). The PML estimator is internally consistent; while not guaranteed to give the smallest mean squared error, it has high efficiency in many survey situations.

Single frame estimators. Bankier (1986) and Kalton and Anderson (1986) proposed estimators of the form in (1) that treat all the observations as though they had been sampled from one frame, with adjusted weights in the intersection domain relying on the inclusion probabilities for each frame. The weight adjustments for the Kalton and Anderson (1986) single frame estimator are:

$$m_{i,S}^A = \begin{cases} 1 & \text{if } i \in a \\ w_i^B \,/\, (w_i^A + w_i^B) & \text{if } i \in ab, \end{cases}$$

$$m_{i,S}^B = \begin{cases} 1 & \text{if } i \in b \\ w_i^A \,/\, (w_i^A + w_i^B) & \text{if } i \in ab. \end{cases}$$

If $w_i^A = 1/\pi_i^A$ and $w_i^B = 1/\pi_i^B$, the single frame estimator uses $\tilde{w}_{i,S}^A = \tilde{w}_{i,S}^B = 1/(\pi_i^A + \pi_i^B)$ for units in ab. The weight adjustment in domain ab relies on both π_i^A and π_i^B. Thus if a disproportionate stratified random sample is taken from frame B, one must know the frame-B stratum membership for units sampled in $\mathcal{S}(A)$. The adjusted weights from the single frame estimator can be interpreted in terms of inclusion probabilities for sampled units. If the sampling fractions are small, $\tilde{w}_{i,S}^A$ is approximately $1/P$(unit i is included in one of the samples). If each of $\mathcal{S}(A)$ and $\mathcal{S}(B)$ is self-weighting, then the single frame estimator reduces to (3).

The single frame weight modifications are the same for all response variables, so estimators are internally consistent. For complex surveys, however, single frame estimators may not be as efficient as the optimal or PML estimators. Their performance may be improved by raking toward the frame population totals (Skinner 1991).

Pseudo-empirical likelihood (*PEL*) *estimators*. Rao and Wu (2010) proposed empirical likelihood estimators for dual frame surveys. Using $\theta = \theta_P$, the empirical log likelihood function is defined by

$$\ell(\mathbf{p}_a, \mathbf{p}_{ab}^A, \mathbf{p}_{ab}^B, \mathbf{p}_b) = \frac{n_A + n_B}{N}\left[\sum_{i\in\mathcal{S}(A),\, i\in a} \frac{N_a}{\hat{N}_a} w_i^A \log(p_{ai}) + \sum_{i\in\mathcal{S}(A),\, i\in ab} \frac{\theta_P N_{ab}}{\hat{N}_{ab}^A} w_i^A \log(p_{abi}^A) + \sum_{i\in\mathcal{S}(B), i\in b} \frac{N_b}{\hat{N}_b} w_i^B \log(p_{bi}) + \sum_{i\in\mathcal{S}(B), i\in ab} \frac{(1-\theta_P)N_{ab}}{\hat{N}_{ab}^B} w_i^B \log(p_{abi}^B)\right],$$

where θ_P is given in (4). An estimator $\hat{\theta}_P$ is substituted if θ_P is unknown. Then $\ell(\mathbf{p}_a, \mathbf{p}_{ab}^A, \mathbf{p}_{ab}^B, \mathbf{p}_b)$ is maximized subject to

$$\sum_{i\in\mathcal{S}(A), i\in a} p_{ai} = 1, \ \sum_{i\in\mathcal{S}(A), i\in ab} p_{abi}^A = 1,$$

$$\sum_{i\in\mathcal{S}(B), i\in b} p_{bi} = 1, \ \sum_{i\in\mathcal{S}(B), i\in ab} p_{abi}^B = 1,$$

and

$$\sum_{i\in\mathcal{S}(A), i\in ab} p_{abi}^A y_i = \sum_{i\in\mathcal{S}(B), i\in ab} p_{abi}^B y_i. \quad (5)$$

When N_{ab} is unknown, the PEL weight modifications are

$$m_{i,\text{PEL}}^A = \begin{cases} \dfrac{p_{ai}^A}{w_i^A}\left\{N_A - \hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P)\right\} & \text{if } i \in a \\ \hat{\theta}_P \dfrac{p_{abi}^A}{w_i^A}\hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P) & \text{if } i \in ab, \end{cases}$$

$$m_{i,\text{PEL}}^B = \begin{cases} \dfrac{p_{bi}^B}{w_i^B}\left\{N_B - \hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P)\right\} & \text{if } i \in b \\ (1-\hat{\theta}_P) \dfrac{p_{abi}^B}{w_i^B}\hat{N}_{ab}^{\text{PML}}(\hat{\theta}_P) & \text{if } i \in ab. \end{cases}$$

The constraint in (5) changes the weights in the overlap domain so that the estimator of Y_{ab} from $\mathcal{S}(A)$ is forced to equal the estimator of Y_{ab} from $\mathcal{S}(B)$. This constraint, however, results in a different set of weights for each response variable. The PEL estimator thus is not internally consistent. Rao and Wu (2010) presented an alternative multiplicity version in which the weight adjustments do not depend on y; in the absence of auxiliary information, this estimator is the same as $\hat{Y}(1/2)$ in (3).

2.2 Weight adjustments with three or more frames

In the general case, suppose there are Q frames, denoted $A_1, \ldots, A_Q$. Let $\mathcal{S}(A_q)$ denote the probability sample from frame A_q, for $q = 1, \ldots, Q$. Unit i in sample $\mathcal{S}(A_q)$ has probability of inclusion $\pi_i^{A_q}$ and weight $w_i^{A_q}$. There are a total of D distinct domains.

A multiple frame estimator generalizing (1) is of the form

$$\hat{Y} = \sum_{q=1}^{Q} \sum_{i\in\mathcal{S}(A_q)} m_i^{A_q} w_i^{A_q} y_i,$$

where $m_i^{A_q}$ is the weight adjustment for observation i in $\mathcal{S}(A_q)$. A fixed weight estimator sets weight adjustments $m^{(A_q,d)}$ for each frame and domain, with the constraints that $m^{(A_q,d)} \geq 0$ ($m^{(A_q,d)}$ is assumed to equal 0 if domain d is not part of frame A_q) and $\sum_{q=1}^{Q} m^{(A_q,d)} = 1$ for $d = 1, \ldots, D$. Then, $m_i^{A_q} = m^{(A_q,d)}$ when observation i from $\mathcal{S}(A_q)$ is in domain d. A simple choice, which generalizes the fixed weight dual frame estimator $\hat{Y}(1/2)$ in (3), takes $m^{(A_q,d)} = [1/\text{number of frames that contain domain } d]$; this is called the multiplicity estimator by Mecatti (2007). Other choices include setting $m^{(A_q,d)} = 1$ in exactly one frame and 0 for the other frames, resulting in screening estimators.

Many of the properties from the dual frame situation extend to the case of three or more frames; multiple frame versions of the estimators in Section 2.1 were studied by Hartley (1974), Lohr and Rao (2006), and Mecatti (2007). How do the multiple frame estimators satisfy the criteria set out at the beginning of this section? All of the estimators – fixed weight, optimal, PML, PEL, and single frame – are approximately unbiased for population totals when sufficiently large samples are taken in the frames. The fixed weight, PML, and single frame estimators are internally consistent; the optimal Hartley-type and Fuller-Burmeister-type estimators in Lohr and Rao (2006) and a multiple-frame extension of the PEL estimator of Rao and Wu (2010) are not internally consistent. While the optimal estimators are asymptotically efficient, they are often unstable in small or moderate samples with three or more frames because the optimal estimated weight modifications are functions of large estimated covariance matrices. The optimal and PEL estimators are ill suited for use with standard survey software because they require a different set of weights for each response variable.

We recommend that one of the internally consistent estimators – fixed weight, PML, or single frame – be used in practice. Lohr and Rao (2006) concluded that the PML estimator has small mean squared error in many survey circumstances, and thus is a good choice for a survey that is conducted only once. With repeated surveys, though, the simplicity and transparency of a fixed weight estimator may be preferred. Fixed weight adjustments may make year-to-year comparisons easier in an annual survey where the domain proportions are relatively constant over time. Fixed weight estimators are also more amenable to weight adjustments for nonresponse and domain misclassification (see Sections 5.1 and 6.1). If fixed weight adjustments can be chosen that are close to the optimal weight adjustments for important responses, perhaps by using estimated design effects from previous surveys, the fixed weight estimator will have mean squared error close to that of the optimal and PML estimators.

3. Postratification to population controls

All of the estimators in Section 2 modify the original sampling weights. As a result, some properties of the original weights may be lost. For example, if a stratified random sample is taken in frame A, the modified weights will not necessarily have the property that the sum of the weights in a stratum equals the stratum population size.

Bankier (1986), in the original development of single frame estimation methods, suggested raking the single frame weights, $\tilde{w}_{i,S}^{A}$ and $\tilde{w}_{i,S}^{B}$, to stratum totals so that the adjusted weights $\tilde{w}_{i,S}^{A,\text{adj}}$ and $\tilde{w}_{i,S}^{B,\text{adj}}$ satisfy

$$\sum\nolimits_{i \in S_{Ah}} (\tilde{w}_{i,S}^{A,\text{adj}} + \tilde{w}_{i,S}^{B,\text{adj}}) = N_{Ah},$$

where S_{Ah} represents the sampled units from either frame in stratum h of frame A, and N_{Ah} is the population size of that stratum. Bankier (1986) and Skinner (1991) used raking ratio estimation to calibrate single frame estimators to the frame population sizes N_A and N_B. Kott, Amrhein and Hicks (1998) proposed using the least squares calibration methods of Deville and Särndal (1992) for calibrating weights to population totals such as stratum sizes.

For the PML estimator, Lohr and Rao (2000) recommended combining the samples first and then using calibration methods to adjust to population as well as separate-frame population totals. When nonresponse is present and a fixed weight estimator is used, Brick, Cervantes, Lee and Norman (2011) concluded that it is preferable to poststratify the individual samples first, and then combine the samples. In some situations, it is most efficient to poststratify both before and after combining samples; in other situations, poststratification can increase bias (see Section 6). Decisions about poststratification need to be made based on the mean squared error, which includes effects of nonsampling errors, and not just on the sampling variance.

4. Analyzing multiple frame surveys with survey software

4.1 Point estimation with survey software

Only internally consistent weight adjustments are suitable for use with survey software when there are multiple responses of interest. Each of the internally consistent methods presented in Section 2.1 results in one vector of adjusted weights for each sample. These may then be concatenated to form one vector of weights: $\tilde{\mathbf{w}} = [\tilde{w}_i^{A_1}, i \in \mathcal{S}(A_1), ..., \tilde{w}_i^{A_Q}, i \in \mathcal{S}(A_Q)]$. Let $\mathbf{y}$ be the corresponding vector of observations, formed by concatenating the observations from samples $\mathcal{S}(A_1)$ through $\mathcal{S}(A_Q)$. Then $\hat{Y} = \tilde{\mathbf{w}}'\mathbf{y}$. From a user's perspective, once the modified weights are constructed, the procedure followed to find point estimates of population totals and means is the same as in a single frame survey.

The modified weights from an internally consistent procedure can be used to estimate any population quantity. Let $F(y)$ be the cumulative distribution function for the population, with

$$F(y) = \sum_{i=1}^{N} I(y_i \le y) \Big/ N,$$

where $I(y_i \le y) = 1$ if $y_i \le y$ and 0 otherwise. In a single frame survey, $F(y)$ is estimated by the empirical cumulative distribution function

$$\hat{F}(y) = \sum\nolimits_{i \in \mathcal{S}} w_i I(y_i \le y) \Big/ \sum\nolimits_{i \in \mathcal{S}} w_i.$$

The modified weights may be used to estimate $F(y)$ in a multiple frame survey:

$$\hat{F}(y) = \sum_{q=1}^{Q} \sum_{i \in \mathcal{S}(A_q)} \tilde{w}_i^{A_q} I(y_i \le y) \Big/ \sum_{q=1}^{Q} \sum_{i \in \mathcal{S}(A_q)} \tilde{w}_i^{A_q}.$$

The denominator is approximately unbiased for N, and the numerator is approximately unbiased for $\sum_{i=1}^{N} I(y_i \le y)$. Any functional of the cumulative distribution function may then be estimated using $\hat{F}(y)$: the mean, $\int y dF(y)$, the median m satisfying $F(m) \approx 1/2$, or any other quantity.

Since the estimators with modified weights are approximately unbiased for population means and totals, they are also approximately unbiased for smooth functions of population means such as ratios and regression coefficients. Any population quantity that could be estimated using the weights from a single frame survey can be estimated analogously using the adjusted weight vector for the multiple frame survey.

4.2 Variance estimation with survey software

Knowledge of the survey designs is needed to calculate standard errors. Variance estimation is straightforward for the estimator in (3), where the weight adjustments do not depend on the data. In that situation,

$$V[\hat{Y}(\theta)] = V\left[\sum_{i \in \mathcal{S}(A)} \tilde{w}_i^A y_i\right] + V\left[\sum_{i \in \mathcal{S}(B)} \tilde{w}_i^B y_i\right],$$

where $\tilde{w}_i^A$ and $\tilde{w}_i^B$ are defined below (2). Create the data set by concatenating the observations from $\mathcal{S}(A)$ and $\mathcal{S}(B)$ as in Section 4.1, using $\tilde{w}_i^A$ and $\tilde{w}_i^B$ as the weights. Define the stratification variable for the combined sample as the combination of categories given by the frame indicator variable, the frame-A stratification variable, and the frame-B stratification variable. Define the first-stage clustering variable for the combined sample similarly as the combination of categories of the individual frame clustering variables. Then, standard survey software may be used to estimate population means and totals using the modified weights, and to estimate variances using the stratification and clustering variables from the combined samples.

Variance estimation is more complicated when the weight modifications m_i^A or m_i^B depend on quantities that are estimated from the sample, as in the PML estimator, or when the combined sample is poststratified or calibrated to population quantities. Linearization, jackknife, and bootstrap methods may then be used to estimate variances.

In the following, we summarize methods that can be used for variance estimation if the psus from the frames are selected independently. If samples from the different frames share psus, other methods must be used. If, for example, psus are selected from the population, and a dual frame design is used within each selected psu, point estimators for psu totals can be calculated using one of the methods described in Section 2. Then standard replication methods can be used to calculate a with-replacement variance estimator.

Under regularity conditions, the linearization and jackknife methods are consistent for estimating the variance of a population characteristic τ that can be written as $\tau = g(\mathbf{A}, \mathbf{B})$, where $\mathbf{A}$ is a vector of population totals from frame A, $\mathbf{B}$ is a vector of population totals from frame B, and g is a twice continuously differentiable function (Skinner and Rao 1996; Lohr and Rao 2000). The vector $\mathbf{A}$ is estimated from $\mathcal{S}(A)$ by $\hat{\mathbf{A}}$, with estimated covariance matrix $\hat{\Sigma}_A$; similarly, $\hat{\mathbf{B}}$ estimates $\mathbf{B}$ from $\mathcal{S}(B)$, with $\hat{V}(\hat{\mathbf{B}}) = \hat{\Sigma}_B$. The linearization estimator of the variance of $\hat{\tau} = g(\hat{\mathbf{A}}, \hat{\mathbf{B}})$ is

$$\hat{V}_L(\hat{\tau}) = g_A' \hat{\Sigma}_A g_A + g_B' \hat{\Sigma}_B g_B,$$

where g_A is the vector of partial derivatives of $g(\mathbf{A}, \mathbf{B})$ with respect to the components of $\mathbf{A}$ and g_B is the corresponding vector of partial derivatives for frame B. Demnati, Rao, Hidiroglou and Tambay (2007) derived linearization estimators of the variance for multiple frame surveys by taking derivatives of a function of the weights rather than of the means. Linearization methods require that the derivatives be calculated separately for each estimator that is considered, and these calculations can be cumbersome. For that reason, it may be preferred to use replication methods if multiple frame surveys are adopted.

Suppose a stratified multistage sample with H strata is taken from frame A, where stratum h has $\tilde{n}_h^A$ primary sampling units. An independent stratified multistage sample with L strata is taken from frame B, where stratum l has $\tilde{n}_l^B$ primary sampling units. The jackknife estimator of the variance can be calculated by creating a total of $\sum_{h=1}^H \tilde{n}_h^A + \sum_{l=1}^L \tilde{n}_l^B$ replicate weight columns (Lohr and Rao 2000). The replicate weights for the column corresponding to the deletion of psu i from stratum h in $\mathcal{S}_A$ are formed by:

$$\tilde{w}_{k(hi)}^A = \begin{cases} \dfrac{\tilde{n}_h^A}{\tilde{n}_h^A - 1} \tilde{w}_k^A & \text{if unit } k \text{ is in stratum } h \text{ but not in psu } i, \\ 0 & \text{if unit } k \text{ is in psu } i \text{ of stratum } h, \\ \tilde{w}_k^A & \text{if unit } k \text{ is in stratum } g \neq h. \end{cases}$$

The jackknife coefficient for this column is the multiplier $(\tilde{n}_h^A - 1) / \tilde{n}_h^A$. The column of replicate weights corresponding to the deletion of psu j from stratum l in $\mathcal{S}_B$ is formed similarly, with jackknife coefficient $(\tilde{n}_l^B - 1) / \tilde{n}_l^B$. With more than two frames, additional columns of replicate weights are added corresponding to the deleted psus from those samples. Weights for a bootstrap method of variance estimation (see Lohr 2007) can be defined similarly.

Multiple frame replication variance methods can be used with standard survey packages that allow replicate weights. If desired, each column in the replicate weights can be post-stratified to population and frame totals, so that the post-stratification is accounted for in the variance estimation.

One challenge with replication variance methods is that the number of columns of replicate weights needed may be very large if a simple random sample or stratified random sample is taken in one of the frames. For the bootstrap, we have found that for some surveys at least 500 bootstrap iterations are needed for variance estimates with dual frame surveys, which again may be excessive. It is possible that combined strata variance estimation, as discussed in Lu, Brick and Sitter (2006), may be used with multiple frame surveys to reduce the number of replicates needed.

5. Nonsampling errors

Multiple frame surveys often have better population coverage than a single frame surveys. When all frames are incomplete, as in Figure 3, any one of frames A, B, or C, if used as the sole sampling frame, would have severe undercoverage. The multiple frame survey design ensures that all units in the overlapping frames have a positive probability of inclusion.

Like all surveys, multiple frame surveys are subject to nonsampling errors. They have nonresponse, which may differ in the different frames. While the union of the frames may have better coverage than a single frame, there may still be undercoverage of the target population. Estimators for multiple frame surveys are also sensitive to domain misclassification and biases that might result from different administration methods or modes in the component surveys. We discuss nonresponse and mode effects in this section, and study effects of domain misclassification in Section 6.

5.1 Nonresponse

In any survey, nonresponse can result in biased estimates of population totals and other quantities. Different nonresponse rates in the samples from the two frames can affect the point estimates of the population total given in Section 2; additionally, nonresponse can affect the weight adjustments prescribed by some of the methods.

Kennedy (2007) discussed a problem that has occurred when frame A consists of landline telephone numbers and frame B has cellular telephone numbers: the units in the intersection domain ab who were interviewed by cell phone differed from those in ab who were interviewed on the landline phone. For example, it was estimated that 18% of the intersection units were aged 18-25 in the frame-B sample, while it was estimated that only 8% of the intersection units were aged 18-25 using the frame-A sample. The difference was ascribed to nonresponse: it was thought that persons who predominantly use cellular telephones (and thus are difficult to reach through a landline survey) tend to be younger. Kennedy (2007) suggested raking using estimated relative telephone usage (*i.e.*, whether most of calls are on landline or cellular telephone).

Brick *et al.* (2011) proposed two methods for nonresponse adjustment in dual frame cellular/landline telephone surveys with fixed weight estimators. They considered a setup in which the overlap domain has two groups: households that receive all or nearly all of their calls on cellular telephones (cell-mainly), and the remaining households in the overlap domain (landline-mainly). The first method, which does not require external estimates of control totals, sets the value of θ in the fixed weight adjustment estimator to reduce the nonresponse bias by using the response rates for the cell-mainly and landline-mainly households in each sample. The second method requires poststratification control totals for the cell-mainly and landline-mainly groups in the overlap domain, N_{1ab} and N_{2ab}, and estimates the population total in domain ab by

$$\sum_{g=1}^{2}\left[\theta_g \frac{N_{gab}}{\hat{N}^A_{gab}} \hat{Y}^A_{gab} + (1-\theta_g) \frac{N_{gab}}{\hat{N}^B_{gab}} \hat{Y}^B_{gab}\right],$$

where $\hat{Y}^A_{gab}$ represents the estimated total of group g in domain ab from $\mathcal{S}(A)$, the other totals are defined similarly, and $0 \le \theta_g \le 1$ for $g = 1, 2$.

5.2 Mode effects

In some cases, multiple frame may also mean multiple mode. De Leeuw (2008) compared the advantages and disadvantages of different sampling modes, and summarized empirical research on mode biases. Persons may give different responses when presented with questions in a visual form than when presented with questions in an auditory form, resulting in mode bias. Mode effects that occur in single frame surveys will also occur in multiple frame surveys. If different modes are used in different frames, it is challenging to separate mode effects from other nonsampling errors.

Many of the multiple frame survey estimators combine estimates from the overlap domains, and these methods assume that the estimators of Y_{ab} from the component surveys both estimate the same quantity. If, however, the frame A survey is conducted in person while the frame B survey is conducted by telephone, it is possible that a census of the domain ab from frame B would give a different domain total than a census from frame A.

One possibility to investigate mode effects is to conduct the frame B survey using a split sample, *e.g.*, partly in person and partly by telephone, but that would reduce the cost savings from the dual frames. Careful pretesting can mitigate the mode effects. Research is needed in this area; the same problem of mode effects, of course, occurs in single frame surveys such as the American Community Survey in which nonresponse follow-up is done by different mode than the original sample (see Citro and Kalton 2007). The methods presented in de Leeuw, Hox and Dillman (2008) for designing surveys for multiple modes also apply in the multiple frame setting.

Vannieuwenhuyze, Loosveldt and Molenberghs (2011) presented a method for distinguishing mode effects from selection effects when a supplemental single-mode survey is available. They noted, however, that the method requires the strong assumption that the coverage and nonresponse errors are equivalent for both surveys. If this assumption is met for a dual frame survey so that the samples in the overlap domain from frames A and B represent the same population, and if domain classification is correct, the mode effect can be estimated from the overlap domain as $D_{ab} = \hat{Y}^A_{ab} - \hat{Y}^B_{ab}$. A difference that is significantly different from 0 indicates presence of a mode effect if there are no other nonsampling errors. If other nonsampling errors are present, a large value of D_{ab} does not provide information about the cause of the difference; experimentation is needed to distinguish possible causes.

6. Domain misclassification and bias adjustment

The estimators discussed in Section 2 construct weights for the observations based on domain membership. Thus in the estimator $\hat{Y}(\theta)$ in (3), the weight multiplier of an observation from sampling frame A is 1 if the observation is in domain a, and is θ if the observation is in domain ab, in order to account for the multiplicity of sampling.

In practice, domain membership may not be clear. For the situation in Figure 1, it may be unknown whether a respondent in an area frame also belongs to the list frame. If frame A is an area frame and frame B is an internet frame, for example, the only way to determine whether an individual sampled from frame A is also in frame B may be to ask the person about internet access, and the person might not give the correct response.

If matching or record linkage is used to determine frame membership, imperfect matching can also misclassify observations. Lesser and Kalsbeek (1999) discussed nonsampling errors that occur in dual frame surveys that have been conducted by the U.S. National Agricultural Statistics Service. Domain misclassification can occur if a farm sampled in the area frame is incorrectly classified with respect to its list frame membership. In landline/cellular dual frame telephone surveys, it is challenging to determine whether a person in one frame is also in the other frame (Kennedy 2007). A person reached in a landline telephone sample may also have a cell phone, but rarely take calls on the cell phone. While technically in the overlap domain, that person is virtually unreachable in the cell phone survey. Some landline/cellular surveys ask respondents about the relative amounts of cellular or landline telephone usage, but misclassification can occur.

In practice, we expect domain misclassification to be related to responses of interest; we also expect that in many situations, misclassification is more likely to occur in certain directions. In longitudinal dual frame surveys, domain misclassification can have greater effects than in cross-sectional surveys (Lu and Lohr 2010). In some situations, the domain indicator can be missing or unavailable. Clark, Winglee and Liu (2007) investigated logistic regression and record-linkage methods for predicting the domain of an observation with missing domain information.

6.1 Misclassification bias adjustments

If domain misclassification is severe, each method for modifying the survey weights to adjust for multiplicity can result in biased estimates of population quantities. In this section we derive a correction for the domain misclassification bias of the fixed weight estimator of Section 2.2 when misclassification probabilities are known. Let the D-vector $\boldsymbol{\delta}_i^{A_q}$ denote the true domain membership for observation i of frame A_q, containing a 1 in position d if observation i is in domain d, and 0 elsewhere. Let $\mathbf{Y} = (Y_1, \ldots, Y_D)'$ denote the vector of population totals for the D domains. For an overlapping dual frame survey, $\mathbf{Y} = (Y_a, Y_{ab}, Y_b)'$; for a three-frame survey, $\mathbf{Y} = (Y_a, Y_{ab}, Y_{ac}, Y_{abc}, Y_b, Y_{bc}, Y_c)'$. If there is no domain misclassification,

$$\hat{\mathbf{Y}}^{A_q} = \sum_{i \in \mathcal{S}(A_q)} \boldsymbol{\delta}_i^{A_q} w_i^{A_q} y_i$$

is the corresponding estimator of $\mathbf{Y}$ from $\mathcal{S}(A_q)$. For fixed weight adjustment vector $\mathbf{m}^{A_q} = (m^{(A_q,1)}, \ldots, m^{(A_q,D)})'$ in frame A_q, satisfying $\sum_{q=1}^{Q} \mathbf{m}^{(A_q,d)} = 1$, then $E[\sum_{q=1}^{Q} (\mathbf{m}^{A_q})' \hat{\mathbf{Y}}^{A_q}] = Y$.

Now suppose there is misclassification. Let $\boldsymbol{\eta}_i^{A_q}$ denote the observed classification for observation i in $\mathcal{S}$. We can write $\boldsymbol{\eta}_i^{A_q} = (\mathbf{M}_i^{A_q})' \boldsymbol{\delta}_i^{A_q}$, where $\mathbf{M}_i^{A_q}$ is a $D \times D$ matrix containing a 1 in position (d, e) if observation i in true domain d is (mis)classified to domain e, and 0 elsewhere.

To allow for differential misclassification within domains, we posit a structure in which the misclassification probabilities can differ for subpopulations in a frame. In a landline/cellular survey, for example, some age groups may be known to have higher misclassification probabilities than others. Chambers, Chipperfield, Davis and Kovačević (2008) used a similar grouping approach to correct for record linkage errors. Suppose the population can be divided into G groups, $g = 1, \ldots, G$, in which the misclassification probabilities are known for each frame A_q. Let $\phi_g^{A_q}(d, e)$ denote the probability that an observation in group g with true domain d is classified into domain e in sample $\mathcal{S}(A_q)$, and let $\boldsymbol{\Phi}_g^{A_q}$ be the $D \times D$ matrix with entries $\phi_g^{A_q}(d, e)$. For observation i belonging to group g and true domain d, assume that row d of $\mathbf{M}_i^{A_q}$ is generated as a multinomial random variable of size 1 with probabilities in row d of the expected misclassification matrix $\boldsymbol{\Phi}_g^{A_q}$, and that all $\mathbf{M}_i^{A_q}$ are independent of each other and of the sample inclusion variables. We thus have G matrices of misclassification probabilities for frame A_q, $\boldsymbol{\Phi}_1^{A_q}, \ldots, \boldsymbol{\Phi}_G^{A_q}$. Denote the vector of population totals for group g by $\mathbf{Y}(g) = \sum_{i=1}^{N} \boldsymbol{\delta}_i^{A_q} \chi_i(g) y_i$, where $\chi_i(g) = 1$ if observation i is in group g and 0 otherwise.

With the observed domain classifications $\boldsymbol{\eta}_i^{A_q}$, the design-weighted estimator of the vector of domain totals in group g is

$$\begin{aligned}\hat{\mathbf{Y}}^{A_q}(\text{mis}, g) &= \sum_{i \in \mathcal{S}(A_q)} \boldsymbol{\eta}_i^{A_q} \chi_i(g) w_i^{A_q} y_i \\ &= \sum_{i \in \mathcal{S}(A_q)} (\mathbf{M}_i^{A_q})' \boldsymbol{\delta}_i^{A_q} \chi_i(g) w_i^{A_q} y_i,\end{aligned}$$

so that $E[\hat{\mathbf{Y}}^{A_q}(\text{mis}, g)] = (\boldsymbol{\Phi}_g^{A_q})' \mathbf{Y}(g)$.

Now consider a new vector of weight adjustments $\tilde{\mathbf{m}}_g^{A_q} = (\tilde{m}_g^{(A_q,1)}, \ldots, \tilde{m}_g^{(A_q,D)})'$ for group g in frame A_q. Then

$$E\left[\sum_{g=1}^{G} \sum_{q=1}^{Q} (\tilde{\mathbf{m}}_g^{A_q})' \hat{\mathbf{Y}}^{A_q}(\text{mis}, g)\right] = \sum_{g=1}^{G} \sum_{q=1}^{Q} (\boldsymbol{\Phi}_g^{A_q} \tilde{\mathbf{m}}_g^{A_q})' \mathbf{Y}(g).$$

Since $\sum_{g=1}^{G} \sum_{q=1}^{Q} (\mathbf{m}^{A_q})' \mathbf{Y}(g) = Y$, the bias will be eliminated under this model when

$$\tilde{\mathbf{m}}_g^{A_q} = (\boldsymbol{\Phi}_g^{A_q})^{+} \mathbf{m}^{A_q}, \tag{6}$$

where $(\boldsymbol{\Phi}_g^{A_q})^{+}$ is the Moore-Penrose inverse of $\boldsymbol{\Phi}_g^{A_q}$, obtained by taking the inverse of the nonzero rows and columns of $\boldsymbol{\Phi}_g^{A_q}$.

Replacing weight adjustments $\mathbf{m}^{A_q}$ by $\tilde{\mathbf{m}}_g^{A_q}$ eliminates the bias under the multinomial misclassification model but inflates the variance. For frame A_q,

$$V\left[\sum_{g=1}^{G} (\tilde{\mathbf{m}}_g^{A_q})' \, \hat{\mathbf{Y}}^{A_q}(\text{mis}, g)\right]$$
$$= E\Bigg[V\Bigg(\sum_{g=1}^{G} \sum_{i\in\mathcal{S}(A_q)} \{(\boldsymbol{\Phi}_g^{A_q})^{+}\mathbf{m}^{A_q}\}' (\mathbf{M}_i^{A_q})' \, \delta_i^{A_q} \chi_i(g) w_i^{A_q} y_i \mid \mathcal{S}(A_1), ..., \mathcal{S}(A_Q)\Bigg)\Bigg]$$
$$+ V\Bigg[E\Bigg(\sum_{g=1}^{G} \sum_{i\in\mathcal{S}(A_q)} \{(\boldsymbol{\Phi}_g^{A_q})^{+}\mathbf{m}^{A_q}\}' (\mathbf{M}_i^{A_q})' \, \delta_i^{A_q} \chi_i(g) w_i^{A_q} y_i \mid \mathcal{S}(A_1), ..., \mathcal{S}(A_Q)\Bigg)\Bigg]$$
$$= \sum_{g=1}^{G} [(\boldsymbol{\Phi}_g^{A_q})^{+}\mathbf{m}^{A_q}]' \, E\Bigg[\sum_{i\in\mathcal{S}(A_q)} \chi_i(g)(w_i^{A_q} y_i)^2 \, \{\text{diag}[(\boldsymbol{\Phi}_g^{A_q})'\delta_i^{A_q}] - (\boldsymbol{\Phi}_g^{A_q})'\delta_i^{A_q}(\delta_i^{A_q})'\boldsymbol{\Phi}_g^{A_q}\}\Bigg] (\boldsymbol{\Phi}_g^{A_q})^{+}\mathbf{m}^{A_q}$$
$$+ V\left[\sum_{i\in\mathcal{S}(A_q)} \{\mathbf{m}^{A_q}\}' \, \delta_i^{A_q} w_i^{A_q} y_i\right].$$

The second term is the variance of the contribution from frame A_q when the units are classified correctly. The first term is zero only when $\boldsymbol{\Phi}_g^{A_q}$ is diagonal for all g, *i.e.*, there is no misclassification.

The weight adjustments in (6) may be extended to the case in which the original fixed weights $\mathbf{m}^{A_q}$ vary for the groups, as long as $\sum_{q=1}^{Q} m_g^{(A_q,\, d)} = 1$ for each domain. Note that the bias correction method in this section is proposed only for the fixed weight estimators, and not for the PML, PEL, or optimal estimators where the multiplicity weights depend on the data. The bias correction depends on the correct specification of the misclassification probabilities. If the misclassification probabilities are estimated from another survey, the operational methods of the surveys must be similar.

6.2 Simulation study

Lohr and Rao (2006) found in simulation studies that the PML estimator has smaller mean squared error than the other estimators when random misclassification is present, but this is due largely to the smaller variance of that estimator. To study sensitivity of estimators to other forms of domain misclassification, we performed a simulation study for two- and three-frame surveys. The population for domain d was generated using the model $y_{ij} = \mu_d + \alpha_i + \varepsilon_{ij}$ for $i = 1, ..., N_d$ and $j = 1, ..., 5$, with $\alpha_i \sim N(0, 1)$ and $\varepsilon_{ij} \sim N(0, 1)$ generated independently, and then probability samples were drawn from this population.

For the two-frame study, the domain means are $\mu_a = -1$, $\mu_{ab} = 0$, $\mu_b = 2$ and factors in the simulation are:

1. Sample size: 100 or 200 from each frame.
2. Cluster sample or simple random sample drawn from frame A. A cluster sample was drawn by selecting a simple random sample of $n_A/5$ of the groups used in generating the population.
3. Misclassification probabilities for frame A (all probabilities not listed are 0):
   a. $\phi_{aa}^A = 1$, $\phi_{ab,ab}^A = 1$ (no misclassification);
   b. $\phi_{aa}^A = 0.9$, $\phi_{a,ab}^A = 0.1$, $\phi_{ab,ab}^A = 1$;
   c. $\phi_{aa}^A = 0.9$, $\phi_{a,ab}^A = 0.1$, $\phi_{ab,ab}^A = 0.9$, $\phi_{ab,a}^A = 0.1$;
   d. $\phi_{aa}^A = 1$, $\phi_{ab,ab}^A = 0.9$, $\phi_{ab,a}^A = 0.1$.
4. Misclassification probabilities for frame B:
   a. $\phi_{bb}^B = 1$, $\phi_{ab,ab}^B = 1$ (no misclassification);
   b. $\phi_{bb}^B = 0.8$, $\phi_{b,ab}^B = 0.2$, $\phi_{ab,ab}^B = 1$;
   c. $\phi_{bb}^B = 0.8$, $\phi_{b,ab}^B = 0.2$, $\phi_{ab,ab}^B = 0.9$, $\phi_{ab,b}^B = 0.1$;
   d. $\phi_{bb}^B = 1$, $\phi_{ab,ab}^B = 0.8$, $\phi_{ab,b}^B = 0.2$.
5. Population sizes: $N_a = N_b = N_{ab} = 25{,}000$; $N_a = N_b = 10{,}000$, $N_{ab} = 55{,}000$; $N_a = 25{,}000$, $N_{ab} = 40{,}000$, $N_b = 10{,}000$.

Ten thousand replicates were run for each combination of the factors, giving the Monte Carlo estimate of bias a standard error of approximately 100. We studied all estimators in Section 2, including $\hat{Y}(1/2)$, $\hat{Y}(2/3)$, and $\hat{Y}(1)$ from (3). We also examined poststratified estimators that could be employed when the domain population counts N_a, N_{ab}, and N_b are known: estimators with subscript "post1" apply poststratification to the two samples first and then combine the samples, and estimators with subscript "post2" combine the samples first and then poststratify to the domain population counts. The bias corrected estimators $\hat{Y}(1/2)_{bc}$ and $\hat{Y}(2/3)_{bc}$ modify the initial fixed weights corresponding to $\theta = 1/2$ and $\theta = 2/3$ using (6). With misclassification pattern (b) in frame A, for example, the bias-corrected weight adjustments for $\hat{Y}(1/2)_{bc}$ are $\tilde{m}_i^A = 19/18$ for i classified in a and $\tilde{m}_i^A = 1/2$ for i classified in ab; for pattern (c), the bias-corrected weight adjustments are 17/16 and 7/16, respectively. The single frame estimator is omitted from these tables since it is the same as either $\hat{Y}(1/2)$ or $\hat{Y}(2/3)$; the single frame estimator raked to the population totals N_A and N_B is denoted by $\hat{Y}_{\text{SF, rake}}$. Tables 1 and 2 display results for $n_A = 100$, $n_B = 100$, $N_a = N_{ab} = N_b = 25{,}000$, and a simple random sample from frame A; Tables 3 and 4 give results for $n_A = 200$, $n_B = 100$, $N_a = N_{ab} = N_b = 25{,}000$, and a cluster sample from frame A. The general patterns of results are similar for the other simulations and are not shown here.

First, consider the fixed weight estimators. The bias-corrected estimators reduce the bias as expected; in all cases studied with misclassification, the empirical bias from the bias-corrected estimators was less than 200 in absolute value, which is within the margin of error. Although the standard deviation for the bias-corrected estimators is higher than for the uncorrected estimators, in most cases the mean squared errors are comparable.

The screening estimator $\hat{Y}(1)$, which discards units from frame B in domain ab, exhibits no misclassification bias when frame B is correctly classified. It also exhibits no bias in Tables 1 and 3 with frame-B misclassification pattern (d) because the observations misclassified from domain ab to domain b have mean 0; for different sets of domain means, pattern (d) does create bias. For the other cases, the screening estimator has the highest bias. For every misclassification pattern, the screening estimator has high mean squared error because data are thrown away. If the domain means are similar, then the misclassification might not result in appreciable bias but discarding observations from domain ab in $\mathcal{S}(B)$ would greatly increase the mean squared error.

Poststratifying to the domain totals when there is misclassification often increases the bias instead of decreasing it. Consider line 4 of Table 1, where 20% of the $\mathcal{S}(B)$ observations in ab are mistakenly classified into domain b. The weights of the observations that are really in domain b, with mean 2, are reduced from 500 to approximately 417, which causes the poststratified versions of $\hat{Y}(1/2)$ to be biased. The effect of poststratification on the mean squared error is mixed, and depends on whether the variance reduction achieved by poststratifying exceeds the additional bias that can be introduced. Raking to the frame totals N_A and N_B, in $\hat{Y}_{\text{SF, rake}}$, has similar effect on misclassification bias as poststratification.

For the simple random samples in Tables 1 and 2, the PML and PEL estimators often exhibit much more bias than the uncorrected fixed weight estimators. The relative contributions from the two frames for these methods depend on the estimated variances of $\hat{N}_{ab}^{A}$ and $\hat{N}_{ab}^{B}$, the domain weights depend on $\hat{N}_{ab}^{\text{PML}}$, and these two factors interact in complex ways depending on the misclassification structure. For misclassification pattern (d) in either frame, $\hat{N}_{ab}^{\text{PML}}$ is too small because observations in domain ab are misclassified; consequently, the weights for the observations in the nonoverlapping domains are too large. A poststratified version of the PML estimator shared the bias problems of the fixed weight poststratified estimators. The PEL estimator, by forcing the estimators of Y_{ab} to be equal, can worsen the bias. For example, in the simulation in line 3 of Table 1, with correct classification for frame A and pattern (c) for frame B, the PEL bias is 50% larger than the PML bias. In this case, the PEL estimator pulls the unbiased estimator $\hat{Y}_{ab}^{A}$ from $\mathcal{S}(A)$ toward the biased estimator from frame B. The optimal estimators also exhibit high bias.

Table 1
Estimated bias for dual frame misclassification, with $n_A = n_B = 100$ and a simple random sample taken from each frame. MPA and MPB refer to the misclassification patterns for frames A and B

MPA	MPB	$\hat{Y}(1/2)$	$\hat{Y}(1/2)_{\text{post1}}$	$\hat{Y}(1/2)_{\text{post2}}$	$\hat{Y}(1/2)_{bc}$	$\hat{Y}(2/3)$	$\hat{Y}(2/3)_{bc}$	$\hat{Y}(1)$	$\hat{Y}_H$	$\hat{Y}_{\text{FB}}$	$\hat{Y}_{\text{PML}}$	$\hat{Y}_{\text{PEL}}$	$\hat{Y}_{\text{SF, rake}}$
a	a	-194	-87	-87	-194	-215	-215	-258	-68	10	-121	-119	-163
a	b	-5,015	4,145	4,529	5	-6,678	17	-10,002	-5,417	1,248	2,486	1,542	2,361
a	c	-5,142	-1,118	-898	-128	-6,823	-138	-10,185	-5,413	-2,583	-1,650	-2,482	-1,690
a	d	-57	-8,430	-8,431	-47	-69	-55	-92	30	-6,576	-6,723	-6,725	-6,795
b	a	1,163	-1,238	-1,290	-82	748	-82	-82	1,355	-2,376	-2,631	-2,551	-2,704
b	b	-3,724	3,040	3,264	43	-5,784	65	-9,905	-3,967	-920	-30	-850	-100
b	c	-3,882	-2,192	-2,187	-124	-5,977	-136	-10,167	-3,954	-4,319	-3,821	-4,477	-3,853
b	d	1,322	-9,445	-9,621	92	917	104	108	1,600	-8,219	-8,720	-8,531	-8,879
c	a	1,366	1,315	1,312	123	969	140	174	1,530	1,529	1,325	1,355	1,276
c	b	-3,729	5,456	5,948	51	-5,801	64	-9,945	-4,216	2,096	3,500	2,391	3,355
c	c	-3,797	235	512	-15	-5,868	2	-10,011	-4,089	-1,377	-417	-1,318	-466
c	d	1,285	-7,072	-7,212	56	873	60	48	1,535	-4,665	-5,131	-4,976	-5,222
d	a	-120	2,134	2,134	-111	-132	-126	-155	32	3,710	3,535	3,538	3,470
d	b	-4,979	6,497	7,086	34	-6,620	65	-9,901	-5,599	4,339	5,928	4,788	5,697
d	c	-5,152	1,174	1,644	-137	-6,835	-152	-10,200	-5,622	310	1,626	578	1,540
d	d	90	-5,999	-5,998	107	98	119	114	193	-2,964	-3,116	-3,120	-3,155

Table 2
Estimated $\sqrt{\text{MSE}}$ for dual frame misclassification, with $n_A = n_B = 100$ and a simple random sample taken from each frame. MPA and MPB refer to the misclassification patterns for frames A and B

MPA	MPB	$\hat{Y}(1/2)$	$\hat{Y}(1/2)_{\text{post1}}$	$\hat{Y}(1/2)_{\text{post2}}$	$\hat{Y}(1/2)_{bc}$	$\hat{Y}(2/3)$	$\hat{Y}(2/3)_{bc}$	$\hat{Y}(1)$	$\hat{Y}_H$	$\hat{Y}_{\text{FB}}$	$\hat{Y}_{\text{PML}}$	$\hat{Y}_{\text{PEL}}$	$\hat{Y}_{\text{SF, rake}}$
a	a	9,646	7,917	7,910	9,646	9,729	9,729	10,304	9,677	8,151	8,081	8,115	8,075
a	b	10,602	9,351	9,531	9,926	11,531	10,197	14,181	11,157	8,212	8,377	8,198	8,311
a	c	10,779	8,622	8,603	10,071	11,715	10,402	14,376	11,243	8,817	8,514	8,720	8,508
a	d	9,789	11,719	11,704	9,674	9,884	9,795	10,432	9,819	10,979	10,978	11,003	11,007
b	a	9,623	8,182	8,185	9,718	9,686	9,766	10,307	9,780	8,446	8,447	8,444	8,459
b	b	9,955	9,054	9,137	9,995	10,949	10,212	14,069	10,489	8,074	7,913	7,995	7,898
b	c	10,146	9,014	9,014	10,160	11,197	10,489	14,404	10,616	9,443	9,108	9,448	9,114
b	d	9,868	12,600	12,716	9,826	9,952	9,927	10,567	10,023	12,063	12,284	12,188	12,371
c	a	9,843	8,185	8,180	9,887	9,853	9,877	10,341	9,991	8,516	8,417	8,442	8,402
c	b	10,049	10,113	10,396	10,039	11,029	10,229	14,127	10,662	8,520	8,863	8,529	8,778
c	c	10,247	8,701	8,718	10,254	11,233	10,534	14,306	10,799	8,762	8,527	8,669	8,516
c	d	10,021	10,861	10,936	9,966	10,068	10,016	10,579	10,177	10,113	10,211	10,168	10,240
d	a	9,795	8,127	8,121	9,734	9,845	9,788	10,343	9,829	9,158	9,024	9,042	8,991
d	b	10,718	10,601	10,970	10,001	11,602	10,258	14,149	11,358	9,461	10,157	9,595	9,986
d	c	10,847	8,558	8,650	10,099	11,769	10,426	14,387	11,424	8,674	8,707	8,608	8,664
d	d	9,945	10,070	10,057	9,778	10,019	9,885	10,510	9,986	9,458	9,412	9,449	9,417

When a cluster sample is taken from frame A, as in Tables 3 and 4, the bias patterns are similar. When there is no misclassification, the MSEs of the optimal and PML estimators are smaller than that of $\hat{Y}(2/3)$ because they account for the survey design. With misclassification, though, the MSE advantage is reduced because of the increased bias.

To study misclassification with a three-frame survey, we selected simple random samples from each frame, and had correct classifications for frames B and C. Table 5 shows results for a simulation with three frames and a simple random sample of size 200 from each frame. The population was generated with $N_d = 10{,}000$ in each domain and domain means $\mu_a = 1$, $\mu_{ab} = 2$, $\mu_{ac} = 3$, $\mu_{abc} = 4$, $\mu_b = 5$, $\mu_{bc} = 6$, $\mu_c = 7$. In this simulation, frames B and C are correctly classified, and the misclassification patterns for frame A are given in the table. We also studied other domain means, population domain sizes, and sample sizes using a factorial design; results for the other settings showed a similar pattern and are not shown here. The multiplicity estimator $\hat{Y}_{\text{ave}}$, with $m_i = 1$ for $i \in \{a, b, c\}$, $m_i = 1/2$ for $i \in \{ab, ac, bc\}$, and $m_i = 1/3$ for $i \in abc$, is optimal when there is no misclassification, and it equals the unraked single frame estimator. The other fixed weight estimators studied are $\hat{Y}_{2/3}$, with $m^{(A,a)} = m^{(B,b)} = m^{(C,c)} = 1$, $m^{(A,ab)} = m^{(A,ac)} = m^{(A,abc)} = 2/3$, $m^{(B,ab)} = m^{(C,ac)} = 1/3$, and $m^{(B,abc)} = m^{(C,abc)} = 1/6$, and the screening estimator $\hat{Y}_{\text{scr}}$, with $m^{(A,a)} = m^{(B,b)} = m^{(C,c)} = m^{(A,ab)} = m^{(A,ac)} = m^{(A,abc)} = m^{(B,bc)} = 1$.

As with the two-frame study, the bias-corrected estimators are approximately unbiased. The screening estimator is also approximately unbiased since only $\mathcal{S}(A)$ is misclassified. The other estimators all exhibit substantial bias with at least some of the misclassification patterns. For the simulation settings in Table 5, the poststratified, single frame raking, Hartley, and PML estimators exhibit large bias but nevertheless have smaller mean squared error than the fixed weight and bias-corrected estimators; this MSE ordering does not hold in some of the other simulation settings.

Mecatti (2007) and Rao and Wu (2010) argued that the fixed weight multiplicity estimator $\hat{Y}_{\text{ave}}$ is unbiased if the only misclassification is among domains that belong to the same number of frames. Misclassifying observations from domain ab to domain ac (pattern c) results in no bias because the weight adjustment in both domains is $1/2$. In practice, though, one would expect pattern (c), with two errors in domain membership (not reporting membership in frame B and erroneously reporting membership in frame C), to be less likely to occur in practice than misclassifying an observation in ab as either a or abc; $\hat{Y}_{\text{ave}}$ can be very sensitive to the latter forms of misclassification. Although a fixed weight estimator is insensitive to misclassification among domains in which the weight adjustments are equal, in these simulations every fixed weight estimator exhibits significant bias for at least some misclassification patterns.

Tables 1 to 5 show that each estimator from Section 2 can exhibit severe bias from domain misclassification. We recommend that the possible extent of domain misclassification be studied during the survey pretesting phase, so that this information can be used in the survey design. If misclassification probabilities are known accurately, then it may be possible to choose a fixed weight estimator that is insensitive to the presumed form of misclassification. When a misclassification-robust estimator cannot be found or when it is inefficient, the fixed weight estimators can be adjusted to reduce the bias. It should be noted that the bias-corrected weights proposed in Section 6.1 are sensitive to the input misclassification probabilities. They also do not account for other nonsampling errors such as nonresponse; applying the misclassification weight adjustments in Section 6.1 followed by the nonresponse weight adjustments described in Brick *et al.* (2011) may result in final weights that correct neither for misclassification nor for nonresponse. If domain misclassification and nonresponse are both present, weight adjustments are needed that deal with both problems simultaneously.

Table 3
Estimated bias for dual frame misclassification, with $n_A = 200$, $n_B = 100$, a cluster sample taken from frame A and a simple random sample taken from frame B. MPA and MPB refer to the misclassification patterns for frames A and B

MPA	MPB	$\hat{Y}(1/2)$	$\hat{Y}(1/2)_{\text{post1}}$	$\hat{Y}(1/2)_{\text{post2}}$	$\hat{Y}(1/2)_{bc}$	$\hat{Y}(2/3)$	$\hat{Y}(2/3)_{bc}$	$\hat{Y}(1)$	$\hat{Y}_H$	$\hat{Y}_{\text{FB}}$	$\hat{Y}_{\text{PML}}$	$\hat{Y}_{\text{PEL}}$	$\hat{Y}_{\text{SF, rake}}$
a	a	-148	-142	-139	-148	-155	-155	-170	-312	63	-119	-172	-184
a	b	-5,090	4,199	4,599	-72	-6,774	-84	-10,144	-4,976	1,210	3,615	2,181	1,025
a	c	-5,069	-1,088	-851	-72	-6,759	-96	-10,139	-4,800	-1,994	177	-1,136	-3,216
a	d	-39	-8,379	-8,383	-35	-63	-58	-111	-237	-5,757	-5,909	-5,961	-6,996
b	a	1,168	-1,221	-1,258	-79	768	-63	-32	1,395	-1,690	-1,663	-2,514	-3,170
b	b	-3,716	2,979	3,236	60	-5,784	79	-9,918	-2,815	-86	1,776	-346	-2,087
b	c	-3,704	-2,108	-2,074	73	-5,771	92	-9,905	-2,561	-2,970	-1,410	-3,267	-5,814
b	d	1,317	-9455	-9,610	95	926	123	144	1,609	-7,285	-7,317	-7,938	-9,498
c	a	1,179	1,281	1,304	-66	772	-58	-41	1,486	1,831	1,652	943	840
c	b	-3,879	5,545	6,087	-118	-5,971	-126	-10,156	-2,972	3,532	4,597	2,405	1,683
c	c	-3,811	318	636	-44	-5,893	-42	-10,058	-2,671	110	1,128	-784	-2,328
c	d	1,423	-6,858	-6,973	191	1,022	206	220	1,824	-4,328	-4,014	-4,516	-5,624
d	a	-33	2,282	2,290	-28	-35	-32	-40	-148	3,627	3,138	3,103	3,728
d	b	-4,974	6,514	7,123	46	-6,660	30	-10,033	-4,863	4,768	6,274	4,742	4,549
d	c	-4,951	1,412	1,883	80	-6,621	84	-9,961	-4,682	1,357	2,863	1,451	388
d	d	42	-5,987	-5,991	53	40	52	37	-126	-2,899	-2,780	-2,791	-3,317

Table 4
Estimated $\sqrt{\text{MSE}}$ for dual frame misclassification, with $n_A = 200$, $n_B = 100$, a cluster sample taken from frame A and a simple random sample taken from frame B. MPA and MPB refer to the misclassification patterns for frames A and B

MPA	MPB	$\hat{Y}(1/2)$	$\hat{Y}(1/2)_{\text{post1}}$	$\hat{Y}(1/2)_{\text{post2}}$	$\hat{Y}(1/2)_{bc}$	$\hat{Y}(2/3)$	$\hat{Y}(2/3)_{bc}$	$\hat{Y}(1)$	$\hat{Y}_H$	$\hat{Y}_{\text{FB}}$	$\hat{Y}_{\text{PML}}$	$\hat{Y}_{\text{PEL}}$	$\hat{Y}_{\text{SF, rake}}$
a	a	10,916	8,912	8,899	10,916	11,092	11,092	11,879	10,975	9,250	9,155	10,109	9,418
a	b	11,786	10,186	10,324	11,157	12,743	11,503	15,463	12,253	8,906	9,391	10,123	9,231
a	c	11,983	9,575	9,537	11,409	12,922	11,814	15,600	12,395	9,575	9,279	10,391	10,039
a	d	11,042	12,357	12,375	10,941	11,250	11,173	12,056	11,051	11,591	11,605	12,229	12,053
b	a	10,698	9,133	9,154	10,872	10,921	11,049	11,875	10,823	9,255	9,151	10,195	9,766
b	b	10,957	9,803	9,867	11,071	12,033	11,413	15,262	11,215	8,681	8,748	9,610	9,182
b	c	11,115	9,860	9,846	11,272	12,172	11,675	15,361	11,306	9,721	9,252	10,558	10,970
b	d	10,988	13,269	13,408	11,046	11,222	11,262	12,143	11,084	12,484	12,347	13,279	13,598
c	a	10,995	9,090	9,073	11,106	11,187	11,254	12,028	11,125	9,309	9,190	9,798	9,389
c	b	11,104	10,779	11,015	11,090	12,162	11,380	15,348	11,430	9,450	9,724	9,754	9,144
c	c	11,155	9,425	9,400	11,189	12,234	11,600	15,424	11,389	9,219	9,064	9,868	9,658
c	d	10,922	11,328	11,421	10,896	11,121	11,091	11,929	11,017	10,759	10,456	11,151	11,182
d	a	11,011	9,080	9,045	10,920	11,181	11,103	11,913	11,041	9,873	9,579	10,375	10,135
d	b	11,838	11,357	11,669	11,164	12,723	11,453	15,299	12,337	10,258	10,848	11,009	10,403
d	c	11,804	9,334	9,371	11,159	12,707	11,548	15,298	12,224	9,349	9,507	10,102	9,442
d	d	11,179	10,839	10,854	10,989	11,355	11,199	12,059	11,195	10,440	10,302	10,916	10,519

Table 5
Estimated bias and $\sqrt{\text{MSE}}$ for misclassification in a 3-frame survey, with $n_A = n_B = n_C = 200$ and a simple random sample taken from each frame. MPA refers to the misclassification patterns for frame A. Pattern (a) has no misclassification; (b) $\phi^A_{aa} = 0.8$, $\phi^A_{a,ab} = 0.1$, $\phi_{a,abc} = 0.1$, $\phi^A_{ab,ab} = 1$, $\phi^A_{ac,ac} = 1$, $\phi^A_{abc,abc} = 1$; (c) $\phi^A_{aa} = 1$, $\phi^A_{ab,ab} = 0.9$, $\phi^A_{ab,ac} = 0.1$, $\phi^A_{ac,ac} = 1$, $\phi^A_{abc,abc} = 1$; (d) $\phi^A_{aa} = 1$, $\phi^A_{ab,ab} = 0.9$, $\phi^A_{ab,abc} = 0.1$, $\phi^A_{ac,ac} = 1$, $\phi^A_{abc,abc} = 1$; (e) $\phi^A_{aa} = 1$, $\phi^A_{ab,ab} = 0.8$, $\phi^A_{ab,a} = 0.1$, $\phi_{ab,abc} = 0.1$, $\phi^A_{ac,ac} = 1$, $\phi^A_{abc,abc} = 1$

	MPA	$\hat{Y}_{\text{ave}}$	$\hat{Y}_{\text{ave, post1}}$	$\hat{Y}_{\text{ave, post2}}$	$\hat{Y}_{\text{ave}, bc}$	$\hat{Y}_{2/3}$	$\hat{Y}_{2/3, bc}$	$\hat{Y}_{\text{scr}}$	$\hat{Y}_H$	$\hat{Y}_{\text{PML}}$	$\hat{Y}_{\text{SF, rake}}$
bias	a	-8	31	28	-8	5	5	20	9	-26	-208
	b	-938	-1,409	-1,478	57	-586	77	107	-2,039	-5,676	-5,624
	c	-26	-485	-508	-26	6	6	6	-324	-825	-957
	d	-231	-514	-557	104	108	108	85	-326	-1,321	-1,438
	e	704	287	247	34	697	27	-4	1,488	1,420	1,193
$\sqrt{\text{MSE}}$	a	9,003	4,419	4,410	9,003	10,013	10,013	13,108	7,990	7,281	7,293
	b	8,961	4,711	4,730	8,955	9,952	9,953	13,092	8,085	9,107	9,074
	c	9,119	4,432	4,422	9,119	10,140	10,140	13,238	8,112	7,396	7,422
	d	8,894	4,405	4,405	8,893	9,874	9,874	12,919	7,957	7,414	7,433
	e	9,088	4,438	4,424	9,059	10,071	10,046	13,180	8,254	7,621	7,581

7. Design issues

As discussed in Section 1, multiple frame designs can give better coverage and precision than a single frame survey with equivalent cost. The design problem is more complex than with a single frame survey, though, since a design that is optimal for frame A and frame B separately may not be optimal for the combined sample. Similarly, a design that is optimal when estimator $\hat{Y}(1/2)$ is used may not be optimal for $\hat{Y}_{\text{PML}}$.

Hartley (1962, 1974) derived optimal designs for the estimator $\hat{Y}(\hat{\theta}_H)$ when a simple random sample is taken in each frame. The optimal sample sizes n_A and n_B depend on the relative costs of sampling from the two frames, and on the means and variances of the response variable within the domains. Cochran (1977, pages 144-145) described the dual frame survey in Figure 1 in his chapter on stratified sampling. In this situation, N_a and N_{ab} may be known, especially if frame B is a list frame. Domains a and ab are treated as strata; there is one sample from stratum a and two independent samples from stratum ab. The design problem may be approached as a stratified sample design.

In general, the optimal design is a function of sampling variances and nonsampling errors in each frame, as well as of the estimator chosen. Biemer (1984) and Lepkowski and Groves (1986) discussed designs for the situation in Figure 1 when a stratified multistage sample is taken from each frame, using the Hartley estimator $\hat{Y}(\hat{\theta}_H)$. Lepkowski and Groves (1986) considered interviewer variability and mode bias as well as sampling error when assessing the precision of various designs; frames with less mode bias are allotted higher sample sizes. Brick (2010) derived optimal allocations in the presence of nonresponse, and found that considering the nonresponse when allocating resources to the two frames can greatly increase efficiency in both screening and overlap dual frame surveys.

One of the advantages of a multiple frame design is its flexibility; it is well suited for a modular approach to survey design. In some situations, it may be practical to take an initial sample from the general population (frame A in Figure 4). The design of the samples from frames B and C, corresponding to subpopulations of interest, can then be determined using information in the frame-A sample. For example, if the frame-A sample yields too few engineers, the sample size from an engineering society membership list frame can be correspondingly increased.

Rao (2003) suggested using multiple frame surveys to improve the accuracy of small area estimates in subgroups of interest. In this application, supplemental surveys can be taken in frames with high concentrations of subgroups of interest. As research needs change, resources can be reallocated among the supplemental surveys without changing the main survey design. A crime victimization survey that uses a national area frame may be supplemented by local victimization surveys; as victimization patterns change, the local surveys can have different sample sizes or be moved to other geographic regions.

Most survey designs attempt to achieve efficiency for the important responses, but in some situations a design that is efficient for one response is inefficient for others. For a survey in which each of four responses of interest was highly correlated with one of the possible stratification variables (but not necessarily correlated with the other stratification variables), Skinner, Holmes and Holt (1994) used a multiple frame survey with four independent stratified samples drawn from a common sampling frame. Each sample was stratified using the stratification variable that was correlated with one of the responses of interest, and so was highly efficient for that response. In estimation, information from all four samples was combined.

Multiple frame surveys can also be used in conjunction with sequential or adaptive sampling methods to improve yield of a rare or hard-to-reach population such as recent immigrants. For example, a stratified multistage sampling design might be employed for frame A, while an adaptive cluster sampling design (Thompson 2002) might be used for frame B. Domain estimates can be calculated separately for the two designs, and then combined using methods in Section 2. In this situation, frames A and B may completely overlap, so that domain misclassification will not be an issue.

8. Conclusions

In this paper, we have summarized some of the issues involved in using multiple frame methods for U.S. household surveys. Multiple frame designs have great potential for improving efficiency of data collection in household surveys. They can improve coverage by combining incomplete frames, improve the accuracy of estimates for subgroups or rare populations, and increase the flexibility and responsiveness of federal data collection. Multiple frame surveys can facilitate sampling hard-to-reach populations such as recent immigrants or households with infants; a general population survey can be combined with an adaptive sample design or a list frame of births.

In many cases, multiple frame surveys can provide more accurate estimates of population quantities without increasing data collection costs, but the design and estimator must be chosen carefully to realize these savings. A multiple frame survey, like other surveys, may have nonresponse, mode effects, and measurement errors. In addition, unless all of the frames consist of the entire population, multiple frame survey estimators can be sensitive to domain misclassification. One correction for misclassification was given in Section 6, but more research is needed on these challenges. Effects of domain misclassification, nonresponse, and mode bias may be confounded. A designed experiment may help disentangle these effects. We are currently studying the relation among these three types of nonsampling errors. Each form of nonsampling error affects the accuracy of multiple frame estimators, and anticipated nonsampling errors need to be incorporated in an optimal design.

Acknowledgements

This research was partially supported by the National Science Foundation under grants SES-0604373 and DRL-0909630. Some of the material in this paper was presented at the 2010 annual meeting of the Italian Statistical Society and published in the proceedings for that conference. The author is grateful to the reviewers for their helpful comments.

References

Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. *Journal of the American Statistical Association*, 81, 1074-1079.

Biemer, P.P. (1984). Methodology for optimal dual frame sample design. Bureau of the Census SRD Research Report CENSUS/SRD/RR-84/07.

Brick, J.M. (2010). Dual frame landline and cell phone surveys. Paper presented at the annual meeting of the Statistical Society of Canada, Québec City.

Brick, J.M., Cervantes, I.F., Lee, S. and Norman, G. (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology*, 37, 1, 1-12.

Brick, J.M., Dipko, S., Presser, S., Tucker, C. and Yuan, Y. (2006). Nonresponse bias in a dual frame survey of cell and landline numbers. *Public Opinion Quarterly*, 70, 780-793.

Chambers, R., Chipperfield, J., Davis, W. and Kovačević, M. (2008). Inference based on estimating equations and probability-linked data. University of Wollongong Centre for Statistical & Survey Methodology Working Paper 18-09.

Citro, C.F., and Kalton, G., Eds. (2007). *Using the American Community Survey: Benefits and Challenges.* Washington, D.C.: National Academies Press.

Clark, J., Winglee, M. and Liu, B. (2007). Handling imperfect overlap determination in a dual-frame survey. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 3233-3238.

Cochran, W.G. (1977). *Sampling Techniques*, 3rd Ed. New York: John Wiley & Sons, Inc.

de Leeuw, E. (2008). Choosing the method of data collection. In *International Handbook of Survey Methodology*, (Eds., E. de Leeuw, J. Hox and D. Dillman). New York: Lawrence Erlbaum, 113-135.

de Leeuw, E., Hox, J. and Dillman, D. (2008). Mixed-mode surveys: When and why. In *International Handbook of Survey Methodology*, (Eds., E. de Leeuw, J. Hox and D. Dillman). New York: Lawrence Erlbaum, 299-316.

Demnati, A., Rao, J.N.K., Hidiroglou, M.A. and Tambay, J.-L. (2007). On the allocation and estimation for dual frame survey data. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 2938-2945.

Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

Fuller, W.A., and Burmeister, L.F. (1972). Estimators for samples selected from two overlapping frames. *Proceedings of the Social Statistics Section*, American Statistical Association, 245-249.

González-Villalobos, A., and Wallace, M.A. (1996). *Multiple Frame Agriculture Surveys*, Rome: Food and Agriculture Organization of the United Nations. Vols. 1 and 2.

Haines, D.E., and Pollock, K.H. (1998). Combining multiple frames to estimate population size and totals. *Survey Methodology*, 24, 79-88.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). *Sample Survey Methods and Theory*. New York: John Wiley & Sons, Inc. Volume 1.

Hartley, H.O. (1962). Multiple frame surveys. *Proceedings of the Social Statistics Section*, American Statistical Association, 203-206.

Hartley, H.O. (1974). Multiple frame methodology and selected applications. *Sankhyā*, Series C, 36, 99-118.

Horvitz, D.G., and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, 663-685.

Iachan, R., and Dennis, M.L. (1993). A multiple frame approach to sampling the homeless and transient population. *Journal of Official Statistics*, 9, 747-764.

Kalton, G., and Anderson, D.W. (1986). Sampling rare populations. *Journal of the Royal Statistical Society*, Series A, 149, 65-82.

Kennedy, C. (2007). Evaluating the effects of screening for telephone service in dual frame RDD surveys. *Public Opinion Quarterly*, 71, 750-771.

Kott, P.S., Amrhein, J.F. and Hicks, S.D. (1998). Sampling and estimation from multiple list frames. *Survey Methodology*, 24, 3-9.

Lepkowski, J.M., and Groves, R.M. (1986). A mean squared error model for multiple frame, mixed mode survey design. *Journal of the American Statistical Association*, 81, 930-937.

Lesser, V.M., and Kalsbeek, W.D. (1999). Nonsampling errors in environmental surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 473-488.

Lohr, S.L. (2007). Recent developments in multiple frame surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 3257-3264.

Lohr, S.L. (2009). Multiple frame surveys. In *Handbook of Statistics*, *Sample Surveys: Design, Methods and Applications*, (Eds., D. Pfeffermann and C.R. Rao). Amsterdam: North Holland, Vol. 29A, 71-88.

Lohr, S.L., and Rao, J.N.K. (2000). Inference in dual frame surveys. *Journal of the American Statistical Association*, 95, 271-280.

Lohr, S.L., and Rao, J.N.K. (2006). Estimation in multiple-frame surveys. *Journal of the American Statistical Association*, 101, 1019-1030.

Lu, W., Brick, J.M. and Sitter, R. (2006). Algorithms for constructing combined strata variance estimators. *Journal of the American Statistical Association*, 101, 1680-1692.

Lu, Y., and Lohr, S.L. (2010). Gross flow estimation in dual frame surveys. *Survey Methodology*, 36, 13-22.

Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, 33, 151-157.

National Science Foundation (2003). *SESTAT: Design and Methodology*, http://srsstats.sbe.nsf.gov/docs/techinfo.html.

Rao, J.N.K. (2003). *Small Area Estimation*. New York: John Wiley & Sons, Inc.

Rao, J.N.K., and Wu, C. (2010). Pseudo-empirical likelihood inference for dual frame surveys. *Journal of the American Statistical Association*, 105, 1494-1503.

Skinner, C.J. (1991). On the efficiency of raking ratio estimation for multiple frame surveys. *Journal of the American Statistical Association*, 86, 779-784.

Skinner, C.J., Holmes, D.J. and Holt, D. (1994). Multiple frame sampling for multivariate stratification. *International Statistical Review*, 62, 333-347.

Skinner, C.J., and Rao, J.N.K. (1996). Estimation in dual frame surveys with complex designs. *Journal of the American Statistical Association*, 91, 349-356.

Thompson, S.K. (2002). *Sampling Techniques*, 2nd Ed. New York: John Wiley & Sons, Inc.

Vannieuwenhuyze, J., Loosveldt, G. and Molenberghs, G. (2011). A method for evaluating mode effects in mixed-mode surveys. *Public Opinion Quarterly*, 74, 1027-1045.