Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Multiple-frame surveys for a multiple-data-source world

by Sharon L. Lohr

Release date: January 6, 2022

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Depository Services Program

• Inquiries line	1-800-635-7943
• Fax line	1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2021
Vol. 47, No. 2, pp. 229-263
Statistics Canada, Catalogue No. 12-001-X

Multiple-frame surveys for a multiple-data-source world

Sharon L. Lohr[1]

Abstract

Multiple-frame surveys, in which independent probability samples are selected from each of Q sampling frames, have long been used to improve coverage, to reduce costs, or to increase sample sizes for subpopulations of interest. Much of the theory has been developed assuming that (1) the union of the frames covers the population of interest, (2) a full-response probability sample is selected from each frame, (3) the variables of interest are measured in each sample with no measurement error, and (4) sufficient information exists to account for frame overlap when computing estimates. After reviewing design, estimation, and calibration for traditional multiple-frame surveys, I consider modifications of the assumptions that allow a multiple-frame structure to serve as an organizing principle for other data combination methods such as mass imputation, sample matching, small area estimation, and capture-recapture estimation. Finally, I discuss how results from multiple-frame survey research can be used when designing and evaluating data collection systems that integrate multiple sources of data.

Key Words: Combining data; Data integration; Dual-frame survey; Indirect sampling; Mass imputation; Misclassification; Survey design; Undercoverage.

1. Introduction

Throughout his 33-year career at the Census Bureau and subsequent 32-year career at Westat, Joe Waksberg repeatedly relied on multiple data sources to improve the quality of estimates while reducing costs. He used external data sources to evaluate coverage in the U.S. decennial census (Marks and Waksberg, 1966; Waksberg and Pritzker, 1969), to calibrate survey weights, and to improve efficiency or oversample rare populations when designing surveys (Hendricks, Igra and Waksberg, 1980; Cohen, DiGaetano and Waksberg, 1988; DiGaetano, Judkins and Waksberg, 1995; Waksberg, 1995; Waksberg, Judkins and Massey, 1997b).

On several occasions, Waksberg integrated data from two or more surveys directly in order to improve coverage or to obtain larger sample sizes for subpopulations (Waksberg, 1986; Burke, Mohadjer, Green, Waksberg, Kirsch and Kolstad, 1994; Waksberg, Brick, Shapiro, Flores-Cervantes and Bell, 1997a). In these multiple-frame surveys, independent samples were selected from sampling frames that together were thought to cover all, or almost all, of the target population. The data from the samples were combined to obtain estimates for the population as a whole and for subpopulations of interest. Waksberg approached the design of these multiple-frame surveys from the perspective of controlling both sampling and nonsampling errors, and found that using multiple frames met the challenges of producing reliable estimates in the face of increased data collection costs (with higher nonresponse for less expensive collection methods) and incomplete frame coverage.

1. Sharon Lohr is Professor Emerita at Arizona State University. E-Mail: sharon.lohr@asu.edu.

Statistical agencies and survey organizations today face the same types of challenges that Waksberg addressed – declining response rates and increasing costs of survey data collection – but at an intensified level. At the same time, the emergence of new data sources provides opportunities for obtaining information about parts of populations of interest – sometimes with amazing rapidity. Many organizations are now using or researching methods for integrating data from multiple sources to improve the accuracy or timeliness of population estimates.

I feel tremendously honored to be asked to give the Waksberg lecture, and in this paper I want to build on Waksberg's insights about multiple-frame surveys by discussing their use as an organizing principle for combining information from multiple sources. Traditionally, multiple-frame surveys have integrated data from Q probability samples $S_1, \ldots, S_Q$ that are selected independently from Q frames. But the general structure can be expanded to include frames that consist of administrative records or nonprobability samples. The structure can also be expanded to situations in which some data sources do not measure the variables of interest y but they measure covariates $\mathbf{x}$ that can be used to predict y.

A number of authors have reviewed methods for combining data from multiple sources; see, for example, Citro (2014), Lohr and Raghunathan (2017), National Academies of Sciences, Engineering, and Medicine (2017, 2018), Thompson (2019), Zhang and Chambers (2019), Beaumont (2020), Yang and Kim (2020), and Rao (2021). The sources include traditional probability samples, administrative data sets, sensor data, social network data, and general convenience samples.

Although the types of data (and the speed with which some types of data can be collected) have changed in recent years, the basic structure of the problem for combining data sources is unchanged from the earliest dual-frame surveys. Section 2 discusses the structure and assumptions for traditional multiple-frame surveys through the example of the National Survey of America's Families, a dual-frame survey that Waksberg worked on during the 1990s. Section 3 reviews methods for calculating estimates of population characteristics from traditional multiple-frame surveys where all assumptions are met, including the special case in which one sample is a census of a subset of the population. Section 4 then discusses how the multiple-frame structure incorporates many of the methods currently used for combining data, sometimes with relaxed assumptions. Section 5 addresses issues for designing data collection systems that control sampling and nonsampling errors, with a discussion of possible future directions for research.

2. Classical multiple-frame survey structure and assumptions

First, let's look at an example of what I shall call a "classical" multiple-frame survey – a survey that is designed to take probability samples from each of a fixed number of frames – and define the notation and assumptions that will be used to describe estimators and their properties.

2.1 National Survey of America's Families

The goal of the 1997 National Survey of America's Families (NSAF) was to provide information on social and economic characteristics of the U.S. civilian noninstitutional population under age 65, with emphasis on obtaining reliable estimates for persons and families – particularly families with children – below 200 percent of the poverty threshold. Estimates were desired for the nation as a whole; in addition, separate estimates were desired for 13 states that were purposively selected to vary by geographic region, dominant political party, size, and fiscal capacity.

To meet the precision requirements for estimates, it was desired to have an effective sample size of about 800 poor children in each state. This goal could have been met by taking a household sample from an area frame. Waksberg et al. (1997b) had determined that screening households for income and subsampling nonpoor households would be the most cost-effective way of achieving the desired sample sizes in an area-frame sample, but the cost would be high because only about one in eight families was expected to have children and be under 200 percent of the poverty threshold.

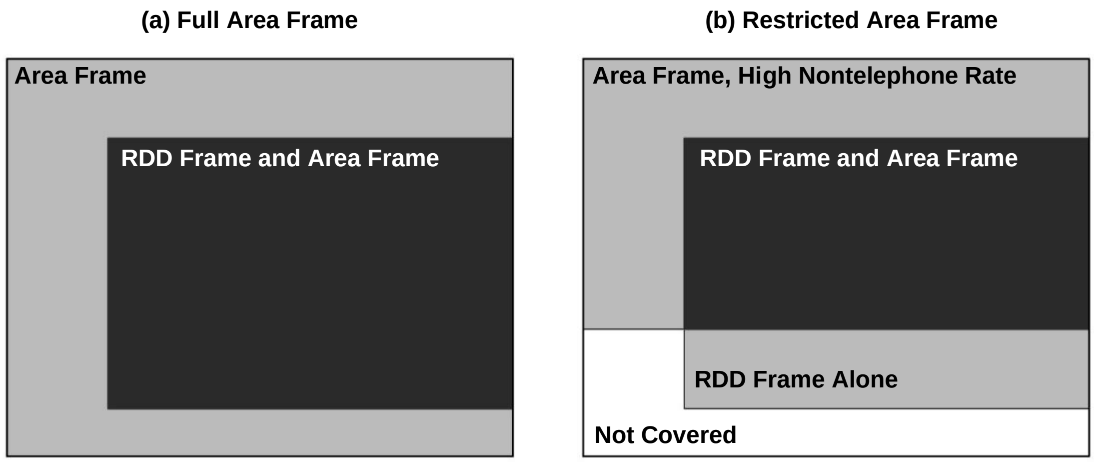
Screening costs would be greatly reduced if the survey could be conducted by telephone using random digit dialing (RDD). But Current Population Survey data indicated that about 20 percent of families living in poverty did not have telephones, so the RDD frame was expected to have substantial undercoverage of the target population. Moreover, households under 200 percent of poverty without telephones might have different income levels or health characteristics than households under 200 percent of poverty with telephones.

Thus, a sample from the area frame would provide high coverage but also come with unacceptably high costs. An RDD survey would have lower costs but would have substantial undercoverage of the population of interest. Waksberg et al. (1997a) used a dual-frame survey, with one sample from the area frame and a second sample chosen independently from the RDD frame, to take advantage of the lower costs of an RDD sample yet also cover nontelephone households. Figure 2.1(a) shows the structure of the two frames.

To further reduce costs, Waksberg et al. (1997a) excluded census block groups with few nontelephone households from the area frame; according to the 1990 census, the excluded areas accounted for less than ten percent of the nontelephone households in each state. With this exclusion, the area and RDD frames each contained households not found in the other frame, as shown in Figure 2.1(b).

Households with telephones that were in the non-excluded block groups were present in both frames. If a probability sample were taken from each frame, households in that overlap (the dark shaded area in Figure 2.1(b)) could be selected in both samples. The survey designers could either conduct the interview with all households in each sample and then deal with the multiplicity in the estimation (an overlap design), or screen out the households in one of the frames that were also in the other frame (a screening design).

Figure 2.1 Frame coverage for the NSAF. The dark shaded area is in both frames.



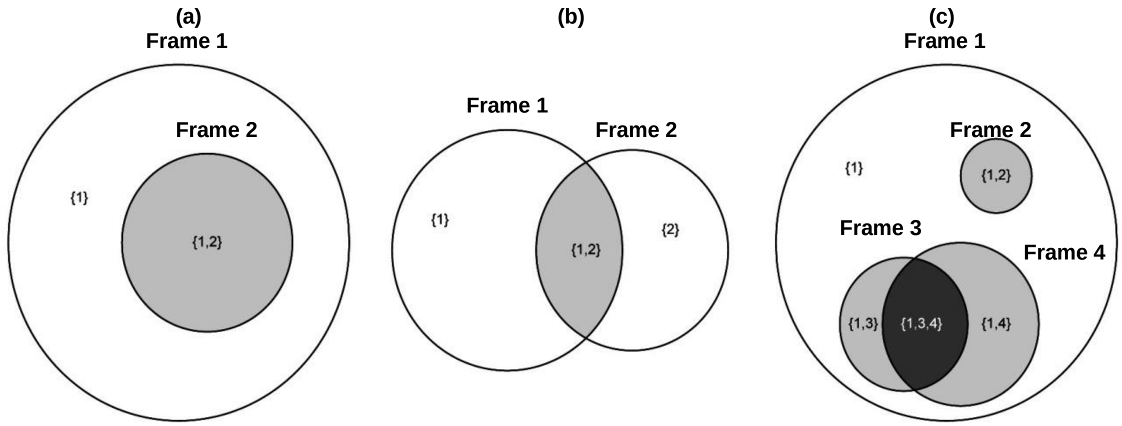
Waksberg and his colleagues chose to use screening. Households in the area sample were asked if they had a telephone, and only those without telephones were administered the detailed interview. The detailed interview was lengthy and expensive to conduct; screening out the telephone households during a short interview saved resources that could be used to increase the number of nontelephone households in the sample. Households with telephones were sampled only through the RDD frame; households in the RDD sample with no children and above 200 percent of the poverty line were subsampled. Because a screening survey was used, the combined sample from the two surveys was a stratified sample, and resources were allocated to the two samples using stratified sampling formulas that accounted for the higher cost of sampling from the area frame.

2.2 Notation and assumptions for multiple-frame surveys

In classical multiple-frame surveys such as the NSAF, a number of assumptions are needed to be able to obtain unbiased estimates of population characteristics along with confidence intervals having approximately correct coverage probabilities.

Suppose there are Q frames. A population domain d is defined by the intersections of the frames: domain {1, 3, 4}, for example, contains the population units that are in Frames 1, 3, and 4 but not in any of the other frames. Let D denote the set of possible domains; depending on the overlap of units, D can contain between 1 and $2^Q - 1$ domains. Figure 2.2 shows three examples of frame relationships. When Frame 1 is complete but Frame 2 is incomplete as in Figure 2.2(a), $D = \{\{1\}, \{1, 2\}\};$ any population unit in Frame 2 is also in Frame 1. For an overlapping dual-frame survey such as that in Figure 2.2(b), $D = \{\{1\}, \{2\}, \{1, 2\}\}.$

Figure 2.2 Three frame structures. (a) Frame 1 has complete coverage and Frame 2 is incomplete. (b) Frames 1 and 2 are both incomplete but overlap. (c) Frame 1 is complete; Frames 2, 3, and 4 are all incomplete but Frames 3 and 4 overlap.



Define $\delta_i(d) = 1$ if unit i is in domain d and 0 otherwise, and let $\delta_i^{(q)} = 1$ if unit i is in Frame q and 0 otherwise. Frame q has population size $N^{(q)}$ and domain d has population size N_d; these sizes may be known or unknown. The target population has a total of N units.

The following assumptions are typically made in order to draw inferences from classical multiple-frame surveys.

(A1) The union of the Q frames covers the target population.

(A2) The sample S_q taken from Frame q is a probability sample where unit i has probability $\pi_i^{(q)}$ of being in S_q. Let $w_i^{(q)}$ represent the final weight for unit i in S_q; options for $w_i^{(q)}$ include the design weight $1/\pi_i^{(q)}$, the Hájek weight $N^{(q)}/[\hat{N}^{(q)}\pi_i^{(q)}]$ with $\hat{N}^{(q)} = \sum_{j \in S_q} 1/\pi_j^{(q)}$, or a nonresponse-adjusted weight.

(A3) The samples $S_1, \ldots, S_Q$ are selected independently.

(A4) The domain membership of each unit i in S_q, $\{\delta_i(d), d \in D\}$, is known.

(A5) The estimator of the population total in domain d from S_q, $\hat{Y}_d^{(q)} = \sum_{i \in S_q} \delta_i(d) w_i^{(q)} y_i$, is approximately unbiased for $Y_d = \sum_{i=1}^{N} \delta_i(d) y_i$, for all Frames q containing domain d and for all variables y.

(A6) There is no measurement error. If unit i is in Frame q and Frame q', y_i will have the same value if measured in S_q as it will if measured in $S_{q'}$.

These are strong assumptions; some relaxation of individual assumptions is possible for specific estimators, as discussed in Section 3. But they are weaker than assumptions needed for some of the other possible data integration methods. Record linkage, for example, has an implicit assumption that unit i in Frame q can be matched with a specific unit in Frame q'. For multiple-frame surveys, one must know whether a unit sampled from Frame q is also in other frames, but does not need to identify the matched unit.

2.3 Were the assumptions met in the NSAF?

Survey assumptions are rarely met exactly in practice, and the NSAF was no exception. Assumption (A1) was not met because of the exclusion of block groups with high telephone ownership. The sample from the area frame yielded fewer nontelephone households than expected, perhaps because of measurement error in the 1990 census or population changes since 1990. In addition, post-survey investigations using data from the 1997 Current Population Survey indicated that the block groups excluded from the frame may have had more nontelephone households than anticipated (Waksberg, Brick, Shapiro, Flores-Cervantes, Bell and Ferraro, 1998).

Although independent probability samples were taken from each frame, each sample had nonresponse. The estimated response rates for children were 65 percent in the RDD sample and 84 percent in the area sample. The weighting procedure attempted to address potential bias from undercoverage and nonresponse. The weights of the nontelephone households in the area sample were ratio-adjusted to attempt to compensate for undercoverage from the block group exclusions. Nonresponse-adjusted weights were calculated separately for the area- and RDD-frame samples, and then the combined samples were poststratified to Census Bureau control totals (Brick, Shapiro, Flores-Cervantes, Ferraro and Strickler, 1999). Groves and Wissoker (1999) found little evidence of residual bias in their nonresponse bias analysis; one of the few differences they reported was that households in the RDD sample that required more calls for contact, and households in a subsample taken of nonrespondents, were slightly less likely to be receiving food assistance.

In the NSAF, the domain membership was determined by asking household respondents in the area sample if they had a working telephone. If that question was answered accurately, then Assumption (A4) was met. The investigators attempted to reduce measurement error for Assumption (A6) by having centralized telephone interviewers conduct all of the detailed interviews; households in the area frame were interviewed over a cellular telephone brought by the field representative. Because interviews in domain $\{1, 2\}$ were obtained only from the RDD sample, however, no data are available for evaluating possible measurement error or relative nonresponse bias for the two samples.

Waksberg had used dual-frame surveys several times prior to the NSAF, mostly to increase sample sizes when sampling rare populations, but he recommended using them only when a simpler design would not meet the survey objectives. He wrote: “The price is additional complexity in the sampling operations and the possibility of error if the matching of the two frames is not done carefully.... My instincts are that a more complex scheme should not be used unless there is a reasonably good pay-off” (Waksberg, 1986).

Was the extra complication and expense of the dual-frame design worth the effort in the NSAF? Because telephone households were screened out of the area sample, and because the yield of nontelephone households was less than anticipated, only 1,488 of the total of 44,461 interviewed households came from the area sample. But because of the high poverty rate of the nontelephone households, the estimated percentage of children in households under 200 percent of the poverty threshold was about 3.6 percentage points higher with the full sample than with the RDD sample alone. Even though for many variables there was only a small difference between the full-sample estimate and the RDD-sample estimate, that difference could not have been evaluated without the area sample.

3. Estimation in classical multiple-frame surveys

The main problem for inference in a classical multiple-frame survey – one that is designed so as to satisfy Assumptions (A1) to (A6) – is how to account for potential overlap among the samples. In the NSAF, telephone households were screened out of the area sample, but in many applications screening is infeasible or it is more cost-effective to obtain data from the full sample selected from each frame. When separate surveys or data sources are not designed with data combination in mind, the overlap depends on the coverage of the individual data sources.

With an overlap design, units that are contained in more than one frame have multiple chances for being selected in the sample. An estimator constructed by summing the weighted observations from each of the Q samples,

$$\hat{Y}_{\text{concat}} = \sum_{q=1}^{Q} \sum_{i \in S_q} w_i^{(q)} y_i,$$

will be a biased estimator of $Y = \sum_{i=1}^{N} y_i$ because the individual sample weights do not reflect the multiple chances of selection for units in overlap domains. Methods for estimating population totals thus typically multiply the survey weights $w_i^{(q)}$ by a multiplicity adjustment $m_i^{(q)}$ that satisfies $\sum_{q=1}^{Q} \delta_i^{(q)} m_i^{(q)} \approx 1$ for each unit i, resulting in the estimator

$$\hat{Y} = \sum_{q=1}^{Q} \sum_{i \in S_q} w_i^{(q)} m_i^{(q)} y_i = \sum_{q=1}^{Q} \sum_{i \in S_q} \tilde{w}_i^{(q)} y_i, \tag{3.1}$$

where $\tilde{w}_i^{(q)} = w_i^{(q)} m_i^{(q)}$ is the multiplicity-adjusted weight.

3.1 Hartley's composite estimator

Hartley (1962) was the first author to present a rigorous theory of estimation in dual-frame surveys where units in the overlap domain {1, 2} might be sampled from both frames. This four-page paper made several important contributions. First, Hartley defined the problem in statistical terms. Second, he proposed an optimal estimator for combining the estimates from the two surveys. And third, he studied the design problem of allocating the resources to the different samples, with a joint consideration of the allocation and the estimator that minimize the variance of the estimated population total subject to a fixed cost.

Hartley (1962) estimated the population total $Y = \sum_{i=1}^{N} y_i$ by

$$\hat{Y}(\theta) = \hat{Y}_{\{1\}}^{(1)} + \hat{Y}_{\{2\}}^{(2)} + \theta \hat{Y}_{\{1,2\}}^{(1)} + (1-\theta)\,\hat{Y}_{\{1,2\}}^{(2)}. \tag{3.2}$$

He proposed choosing θ to minimize $V\left[\hat{Y}(\theta)\right]$. This resulted in the value

$$\theta_H = \frac{V\left(\hat{Y}_{\{1,2\}}^{(2)}\right) + \mathrm{Cov}\left(\hat{Y}_{\{2\}}^{(2)}, \hat{Y}_{\{1,2\}}^{(2)}\right) - \mathrm{Cov}\left(\hat{Y}_{\{1\}}^{(1)}, \hat{Y}_{\{1,2\}}^{(1)}\right)}{V\left(\hat{Y}_{\{1,2\}}^{(1)}\right) + V\left(\hat{Y}_{\{1,2\}}^{(2)}\right)}. \tag{3.3}$$

The estimator in (3.2) is of the form in (3.1) with multiplicity weight adjustments

$$m_i^{(1)} = \delta_i\left(\{1\}\right) + \delta_i\left(\{1,2\}\right)\theta, \quad m_i^{(2)} = \delta_i\left(\{2\}\right) + \delta_i\left(\{1,2\}\right)\left(1-\theta\right).$$

If it is desired to use the optimal compositing factor θ_H, estimators may be substituted for the unknown covariances in (3.3). Because θ_H depends on covariances involving y, however, the optimal multiplicity adjustment may differ for different variables, giving a different set of weights for each. In addition, θ_H can be less than 0 or greater than 1, possibly resulting in negative weights for some observations. These features carry over to the Q-frame generalization of Hartley's optimal estimator studied by Lohr and Rao (2006).

The estimator in (3.2), with fixed value of θ, is approximately unbiased for Y under Assumption (A5). If the estimated domain totals and the estimates of the covariances in (3.3) are consistent, then the estimator with $\hat{\theta}_H$ is consistent for Y. Saegusa (2019) studied Hartley's estimator from the perspective of empirical process theory, establishing a law of large numbers and a central limit theorem when S_1 and S_2 are both simple random samples.

Hartley's application was in agriculture, and many of the early applications of dual-frame surveys were for agriculture or business surveys (Kott and Vogel, 1995), where list frames existed that contained the larger business or agricultural operations. A dual-frame survey with a disproportionately larger sample from the list frame reduced costs because (1) obtaining data from an operation in the list frame was often less expensive than obtaining data from an operation in the area frame and (2) oversampling the list frame was analogous to oversampling high-variance strata in stratified sampling and thus resulted in greater efficiency.

Later, as cellular telephones became more prevalent, concern about bias from using landline telephone samples alone led to use of dual-frame telephone surveys, with one sample from a landline frame and a second sample from a cellular telephone frame. Here, both frames are incomplete but together cover the population of persons with telephones. For these surveys, an important consideration is how to deal with persons having both kinds of telephones. The next section reviews choices for the compositing.

3.2 Multiplicity weighting adjustments

Hartley's optimal estimator, with θ_H, uses a different set of weights for each response variable, which can lead to internal inconsistencies among estimators. Various authors have proposed estimators that use a single set of weights for all analyses. Here, I briefly list some of the multiplicity adjustment factors $m_i^{(q)}$ that result in one set of weights for the general estimator of the population total in (3.1). The resulting estimators are approximately unbiased for the population total Y under Assumptions (A1), (A4), and (A5). These and additional estimators are reviewed in detail by Lohr (2011), Lu, Peng and Sahr (2013), Ferraz and Vogel (2015), Arcos, Rueda, Trujillo and Molina (2015), and Baffour, Haynes, Western, Pennay, Misson and Martinez (2016).

- Screening estimator, with $m_i^{(1)} = 1, m_i^{(2)} = 1 - \delta_i^{(1)}, \ldots, m_i^{(Q)} = \prod_{q=1}^{Q-1} \left(1 - \delta_i^{(q)}\right).$ A unit sampled from Frame q is discarded if it is in any of Frames $1, \ldots, q-1.$ This estimator is automatically used with a screening design such as the NSAF; with an overlap design, its use means that some data observations are thrown away.
- Multiplicity estimator, with $m_i^{(q)} = 1/$ (number of frames containing unit i) $= 1 \big/ \sum_{q=1}^{Q} \delta_i^{(q)}$. In a dual-frame survey, this gives the estimator in (3.2) with $\theta = 1/2$. Mecatti (2007) noted that with the multiplicity estimator, Assumption (A4) can be replaced by the slightly less restrictive assumption that $\sum_{q=1}^{Q} \delta_i^{(q)}$ is known for each sampled unit i.

  The multiplicity estimator can also be viewed as a special case of the generalized weight share method (Deville and Lavallée, 2006) using the standardized link matrix, since the number of links to population unit i is the number of frames containing that unit.
- Single-frame estimator (Bankier, 1986; Kalton and Anderson, 1986), which considers the observations as if they had been sampled from a single frame. If inverse probability weights are used, with $w_i^{(q)} = 1 \big/ \pi_i^{(q)}$, then $m_i^{(q)} = \pi_i^{(q)} \big/ \sum_{f=1}^{Q} \delta_i^{(f)} \pi_i^{(f)}$. This estimator requires that the inclusion probability for unit i be known for all Q frames, including frames from which the unit was not sampled. The multiplicity adjustments consider the inclusion probabilities for the designs but not the relative variances, which are affected by clustering and stratification in the individual samples.
- Effective sample size (ESS) estimator (Chu, Brick and Kalton, 1999; O'Muircheartaigh and Pedlow, 2002), where the domain estimator from each frame is weighted by the relative effective sample size from that frame. Let $n^{(q)}$ be the sample size from Frame q and let $\text{deff}^{(q)}$ denote the design effect for a key variable or a smoothed design effect for multiple variables. The effective sample size for S_q is $\tilde{n}^{(q)} = n^{(q)} \big/ \text{deff}^{(q)}$ and the multiplicity adjustment for unit i is

  $$m_i^{(q)} = \frac{\tilde{n}^{(q)}}{\sum_{f=1}^{Q} \delta_i^{(f)} \tilde{n}^{(f)}}.$$

  This estimator considers the relative variances of estimators from different samples and is often more efficient than the screening, multiplicity, and single-frame estimators.

The pseudo-maximum-likelihood (PML) estimator of Skinner and Rao (1996) is of this type when the frame sizes $N^{(q)}$ and domain sizes N_d are unknown; Skinner and Rao (1996) recommended using the design effect for estimating $N_{\{1,2\}}$ to establish the effective sample size for the dual-frame case. The PML estimator is asymptotically equivalent to an ESS estimator that poststratifies to the domain sizes N_d when those are known; when the frame sizes $N^{(q)}$ are known but not $N_{\{1,2\}}$, the PML estimator is asymptotically equivalent to calibrating the ESS estimator to estimated domain sizes calculated from the pseudo-likelihood function.

Approximately unbiased estimates of the variances for all estimators considered in this section can be derived under Assumptions (A1) to (A6) and additional regularity conditions that ensure consistency of estimated totals and variance estimators from the Q samples. Skinner and Rao (1996) studied linearization variance estimators; Chauvet (2016) derived linearization variance estimators for the French housing survey that accounted for the variance reduction due to high sampling fractions from some of the frames. Lohr and Rao (2000) developed theory for using the jackknife with multiple frames, and Lohr (2007) and Aidara (2019) considered bootstrap variance estimators. These methods rely on Assumption (A3) of independent samples; Chauvet and de Marsac (2014) considered the situation in which the samples share primary sampling units but independent samples are taken at the second stage of the design.

Calculating linearization variance estimates requires special software that implements the partial derivative calculations for the multiple frames. Replication variance estimation methods such as jackknife and bootstrap, however, can be calculated in standard survey software by creating a single data set that contains all the concatenated observations and weights $\tilde{w}_i^{(q)}$ from the Q samples and creating replicate weights using standard methods for stratified multistage samples (Metcalf and Scott, 2009). The concatenated data set has $\sum_{q=1}^{Q} H_q$ strata, where H_q is the number of strata for S_q; observations from different samples are in different strata. The replicate weight methods also can include effects of calibration (see Section 3.3) on the variance.

Of course, many applications call for estimates of quantities other than population totals, and the multiple-frame theory applies to parameters that are smooth functions of domain totals. A different compositing factor may be desired when quantities other than population totals are of primary interest, however, and there may be special considerations for other types of analyses. Other types of statistical analyses that have been studied in the multiple-frame setting include linear (Lu, 2014b) and nonparametric (Lu, Fu and Zhang, 2021) regression, logistic regression with ordinal data (Rueda, Arcos, Molina and Ranalli, 2018), empirical distribution functions (Arcos, Martínez, Rueda and Martínez, 2017), gross flow estimation with missing data (Lu and Lohr, 2010), and chi-squared tests (Lu, 2014a).

Lu (2014b) noted that linear regression parameters estimated using the multiplicity-adjusted weights are the finite population regression coefficients $\mathbf{B}$ that minimize the sum of squares $\sum_{i=1}^{N} (y_i - \mathbf{x}_i^T \mathbf{B})^2$. However, one of the reasons for taking a multiple-frame survey, rather than using an incomplete frame, is a concern that population characteristics may differ across domains. Lu (2014b) suggested examining the residuals separately by domain and also fitting separate regression models by domain to assess the appropriateness of the regression model.

3.3 Calibration

The PML estimator is calibrated to population counts that are known for the frames and domains. In a dual-frame survey where $N^{(1)}$ and $N^{(2)}$ are known, $\sum_{q=1}^{2}\sum_{i \in S_q} w_i^{(q)} m_{i,\text{PML}}^{(q)} \delta_i^{(f)} = N^{(f)}$ for $f = 1, 2.$ If the overlap domain size $N_{\{1,2\}}$ is also known, the PML estimator is calibrated to all three domain sizes. Skinner (1991) used calibration with the single-frame estimator, raking the estimator to the population frame counts.

Ranalli, Arcos, Rueda and Teodoro (2016) studied general calibration theory for dual-frame surveys. They assumed that a vector of auxiliary information $\mathbf{x}$ is available with known population totals $\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i,$ and calculated multiple-frame generalized regression weights as

$$c_i^{(q)} = \tilde{w}_i^{(q)} \left[1 + (\mathbf{X} - \hat{\mathbf{X}})^T \left(\sum_{f=1}^{Q} \sum_{k \in S_f} \alpha_k \tilde{w}_k^{(f)} \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \alpha_i \mathbf{x}_i \right], \tag{3.4}$$

where α_k is an arbitrary constant and $\hat{\mathbf{X}} = \sum_{f=1}^{Q} \sum_{k \in S_f} \tilde{w}_k^{(f)} \mathbf{x}_k$ estimates $\mathbf{X}$ using the multiplicity-adjusted weights. Under regularity conditions, they showed that for the dual-frame estimator in (3.2) with fixed $\theta,$ the variance of the generalized regression estimator $\hat{Y}_{\text{GR}} = \sum_{q=1}^{2} \sum_{i \in S_q} c_i^{(q)} y_i$ is approximated by

$$V(\hat{Y}_{\text{GR}}) \approx V\left[\sum_{q=1}^{2} \sum_{i \in S_q} \tilde{w}_i^{(q)} (y_i - \mathbf{x}_i^T \mathbf{B}) \right], \tag{3.5}$$

where $\mathbf{B} = \left(\sum_{i=1}^{N} \alpha_i \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{i=1}^{N} \alpha_i \mathbf{x}_i y_i.$ The variance of the estimator depends on the residuals from the regression model just as in the single-frame case.

Särndal and Lundström (2005) distinguished among types of auxiliary information that can be used in calibration. InfoU is information available at the population level. A vector $\mathbf{x}^*$ can be considered as InfoU if the population total $\mathbf{X}^* = \sum_{i=1}^{N} \mathbf{x}_i^*$ is known and $\mathbf{x}^*$ is observed for every respondent in the sample. InfoS is information available at the level of the sample, but not at the population level. Vector $\mathbf{x}^o$ is InfoS if it is known for every member of the sample, both respondents and nonrespondents, but $\sum_{i=1}^{N} \mathbf{x}^o$ is unknown.

In a multiple-frame survey, the variables available for InfoU and InfoS may differ across frames. For the NSAF, little auxiliary information was known for nonrespondents in the RDD sample but address-related information (for example, characteristics of the block group) was known for all members of the area-frame sample. The reverse may be true for a dual-frame survey in which Frame 1 is an area frame and Frame 2 is a list frame. The list frame may have rich information that can be used for weighting class adjustments or calibration, while the auxiliary information for the area frame may be restricted to information measured in the survey for which population totals are known from an external source such as a census or population register.

Ranalli et al. (2016) allowed for differing InfoU information across the frames; some of the auxiliary variables may be known for units from all samples and for the full population, while other variables may be of the form $x_i^* = x_i \delta_i^{(q)}$ with total $X^* = \sum_{i=1}^N x_i \delta_i^{(q)},$ the total of variable x in Frame q. Calibration to frame counts $N^{(q)}$ is thus a special case of the general calibration theory.

But the differing amounts of information for the frames may also have a bearing on the multiplicity adjustments. Suppose that Frame 2 has rich auxiliary information for calibration while Frame 1 has little information. Calibrating the weights $w_i^{(2)}$ before compositing may increase the relative effective sample size from S_2 and thus increase the value of $\tilde{n}^{(2)} \big/ \left(\tilde{n}^{(1)} + \tilde{n}^{(2)}\right)$ that would be used for the ESS estimator.

Haziza and Lesage (2016) argued that a two-step weighting procedure offers several advantages for single-frame surveys with nonresponse. The first step divides the design weight for unit i by its estimated response propensity (often calculated from InfoS information) and the second step calibrates the nonresponse-adjusted weights to population control totals (available from InfoU information). When there is substantial nonresponse, weighting adjustment factors from step 1 are often much higher than those from step 2; if the response propensity model is correct, the weighting adjustments in step 2 converge to 1 as $n \to \infty.$ The two-step procedure is thus more robust toward misspecification of the calibration model.

The same considerations apply for multiple-frame surveys. A two-step procedure, where step 1 adjusts the samples separately for nonresponse and step 2 calibrates the combined samples, provides robustness to the calibration model. Suppose that S_1 has full response; S_2 has nonresponse but the response propensities can be predicted perfectly from variable x. Then, performing a separate nonresponse adjustment for each sample in step 1 removes the bias for S_2 so that Assumption (A5) is satisfied. If the data are combined first and then calibrated using (3.4), however, the calibration may change the weights for units in S_1 in order to meet the calibration constraints – introducing bias for the estimates from S_1 while not removing it for estimates from S_2. More research is needed on the ordering of steps for weight adjustments. It may be better to perform two steps of nonresponse adjustments and calibration on each sample separately, then adjust the weights for multiplicity, and then calibrate to population totals (including re-calibrating on the individual frame variables).

One consequence of using an overlap estimator for a multiple-frame survey is that the multiplicity adjustments may introduce more weight variation, with observations belonging to one frame having much larger weights than observations belonging to more than one frame. If, for example, a list frame (Frame 2 in Figure 2.2(a, b)) is disproportionately oversampled, then the sampling weights for observations in domain $\{1\},$ which are sampled only from Frame 1, may be large relative to the weights for the other domains. Wolter, Ganesh, Copeland, Singleton and Khare (2019) suggested using a shrinkage estimator, estimating $Y_{\{1\}}$ by $\kappa \hat{Y}_{\{1\}}^{(1)} + (1-\kappa)\, N_{\{1\}} \,(\hat{Y}_{\{2\}}^{(2)} + \hat{Y}_{\{1,\,2\}}) \big/ N^{(2)},$ but the shrinkage may introduce bias – after all, the reason for using a more complicated multiple-frame design instead of just sampling from Frame 2 is to avoid potential bias from omitting domain {1}. A better solution, if feasible, is to address the weight variation when designing the survey, as discussed in Section 5.

3.4 Probability sample combined with census of a population subset

Lohr (2014) and Kim and Tam (2021) noted that the situation in Figure 2.2(a) includes the special case in which a probability sample S_1 is taken from Frame 1 having full coverage, and the sample S_2 from Frame 2 is a census of domain {1, 2}. The overlap domain is thus defined to be the units in S_2, which may be from administrative records or a convenience sample. Although S_2, considered by itself, may have undercoverage bias, in the multiple-frame setting the bias is eliminated by the presence of a sample from Frame 1. The units in S_2 have $w_i^{(2)} = 1$ and represent themselves alone; they do not represent any units in other parts of the population. When $N^{(2)} \big/ N$ is small, say from a small convenience sample, S_2 will have little effect on dual-frame estimators – almost all of the population is in domain {1}. But when $N^{(2)} \big/ N$ is large, as may occur when Frame 2 consists of administrative records, the availability of those records may improve the precision of $\hat{Y}$ if Assumptions (A1) to (A6) are met.

When S_2 is a census with no measurement error, $\hat{Y}_{\{1,2\}}^{(2)} = Y_{\{1,2\}}$. The estimator in (3.2) is

$$\hat{Y}(\theta) = \hat{Y}_{\{1\}}^{(1)} + \theta \hat{Y}_{\{1,2\}}^{(1)} + (1-\theta)\, Y_{\{1,2\}}; \tag{3.6}$$

taking $\theta = 0$ uses the known population total from Frame 2 and relies on Frame 1 only for estimation of the part of the population not in Frame 2.

Kim and Tam (2021) noted that since $Y_{\{1,2\}}$ is known, it can be used as an InfoU calibration total. They proposed two calibration estimators: a ratio estimator $\hat{Y}_{\text{ratio}} = \hat{Y}^{(1)} Y_{\{1,2\}} \big/ \hat{Y}_{\{1,2\}}^{(1)}$ and a generalized regression calibration estimator. For many designs, however, the ratio estimator will be less efficient than $\hat{Y}(0)$ from (3.6) because

$$V\left(\hat{Y}_{\text{ratio}}\right) \approx V\left(\hat{Y}_{\{1\}}^{(1)}\right) + \left(\frac{Y_{\{1\}}}{Y_{\{1,2\}}}\right)^2 V\left[\hat{Y}_{\{1,2\}}^{(1)}\right] - 2\,\frac{Y_{\{1\}}}{Y_{\{1,2\}}}\,\text{Cov}\left(\hat{Y}_{\{1\}}^{(1)}, \hat{Y}_{\{1,2\}}^{(1)}\right);$$

the ratio adjustment can introduce extra variability from $\hat{Y}_{\{1,2\}}^{(1)}$ that is excluded from $\hat{Y}(0)$.

Calibrating $\hat{Y}(\theta)$ to $Y_{\{1,2\}} = \sum_{i=1}^{N} x_i$, for $x_i = \delta_i^{(2)} y_i$, the generalized regression weights in (3.4) become

$$c_i^{(q)} = \tilde{w}_i^{(q)} \left[1 + \left(Y_{\{1,2\}} - \hat{Y}_{\{1,2\}}(\theta) \right) \left(\sum_{f=1}^{Q} \sum_{k \in S_f} \tilde{w}_k^{(f)} \delta_k^{(2)} y_k^2 \right)^{-1} \delta_i^{(2)} y_i \right], \tag{3.7}$$

resulting in $\hat{Y}_{\text{GR}} = \hat{Y}(0)$ from (3.6). Similarly, calibrating on the vector $\mathbf{x}_i = \left(1, \delta_i^{(2)}, \delta_i^{(2)} y_i\right)^T$ results in $\hat{Y}_{\text{GR}} = \hat{Y}_{\{1\}}^{(1)} N_{\{1\}} \big/ \hat{N}_{\{1\}}^{(1)} + Y_{\{1,2\}}$.

For some designs, the variance can be reduced even further. Montanari (1987, 1998) proposed using the regression coefficient $\boldsymbol{\beta} = \left[V(\hat{\mathbf{X}})\right]^{-1} \operatorname{Cov}(\hat{Y}, \hat{\mathbf{X}})$ for calibration, resulting in the estimator

$$\hat{Y}_{\text{opt}} = \hat{Y} + (\mathbf{X} - \hat{\mathbf{X}})^T \boldsymbol{\beta}. \tag{3.8}$$

Rao (1994) called (3.8) the optimal regression estimator and showed that $V(\hat{Y}_{\text{opt}}) \le V(\hat{Y}_{\text{GR}})$. For the dual-frame situation considered in this section, with $x_i = \delta_i^{(2)} y_i$,

$$\boldsymbol{\beta} = \frac{\operatorname{Cov}\left(\hat{Y}^{(1)}, \hat{Y}^{(1)}_{\{1,2\}}\right)}{V\left(\hat{Y}^{(1)}_{\{1,2\}}\right)} = 1 + \frac{\operatorname{Cov}\left(\hat{Y}^{(1)}_{\{1\}}, \hat{Y}^{(1)}_{\{1,2\}}\right)}{V\left(\hat{Y}^{(1)}_{\{1,2\}}\right)}$$

and

$$\begin{aligned}\hat{Y}_{\text{opt}} &= \hat{Y}^{(1)} + \left(Y_{\{1,2\}} - \hat{Y}^{(1)}_{\{1,2\}}\right)\left[1 + \frac{\operatorname{Cov}\left(\hat{Y}^{(1)}_{\{1\}}, \hat{Y}^{(1)}_{\{1,2\}}\right)}{V\left(\hat{Y}^{(1)}_{\{1,2\}}\right)}\right] \\ &= \hat{Y}^{(1)}_{\{1\}} + \theta_H \hat{Y}^{(1)}_{\{1,2\}} + (1 - \theta_H)\, Y_{\{1,2\}}, \end{aligned} \tag{3.9}$$

where $\theta_H = -\operatorname{Cov}\left(\hat{Y}^{(1)}_{\{1\}}, \hat{Y}^{(1)}_{\{1,2\}}\right) \Big/ V\left(\hat{Y}^{(1)}_{\{1,2\}}\right)$ is Hartley's optimal value for θ from (3.3).

Although we usually think of the compositing factor θ as being between 0 and 1, θ_H can be outside of this range. For a conceptual example, suppose that Frame 2 is a list of children receiving food assistance at school and the sample from Frame 1 is a cluster sample of households. Then households in which one or more children are receiving food assistance have some household members in domain $\{1, 2\}$ and other members in domain $\{1\}$. If y exhibits high intra-household correlation, then we would expect $\hat{Y}^{(1)}_{\{1\}}$ and $\hat{Y}^{(1)}_{\{1,2\}}$ to be positively correlated. In this case, Hartley's optimal estimator results in negative weights for units in domain $\{1, 2\}$ from the probability sample.

Even though $\hat{Y}_{\text{opt}}$ is more efficient for special situations such as the cluster sample described above, it depends in practice on an estimate of the covariance, is optimal only for this particular y variable, and may have negative weights. Negative weights can also occur if one does optimal calibration with auxiliary variable $\left(1, \delta_i^{(2)}, \delta_i^{(2)} y_i\right)$; in fact, that calibration results in the estimator proposed by Fuller and Burmeister (1972). These optimal regression estimators are sensitive to the model assumptions, and in general I do not recommend their use.

When the Frame-2 sample is a census and Assumptions (A1) to (A6) are met, the precision of population estimates depends entirely on the design of S_1. When the samples are not designed to be part of a multiple-frame survey (and sometimes even when they are), it is likely that one or more of the assumptions is violated. Assumptions (A4) and (A6) are particularly suspect when it is desired to combine data from surveys that were not designed with combination in mind. Even if both surveys measure unemployment, they may use different questions so that the unemployment statistics from S_2 measure a different concept than the statistics from S_1. Domain misclassification may also occur. A unit in the census S_2 is known to also be in complete Frame 1, but it may be difficult to tell whether a unit in S_1 is also in the administrative records or convenience sample that serves as S_2. These problems are discussed in the next section.

4. Multiple-frame surveys and data integration

Rao (2021) reviewed a number of data integration methods for combining information from a probability sample S_1, assumed to come from a frame with complete coverage, with information from a nonprobability sample S_2, often a census of part of the population as in Section 3.4. Rao considered two cases for making inferences about y: (1) y is observed in both samples, and (2) auxiliary information $\mathbf{x}$ is observed in both samples but y is observed only in S_2. In this section I examine various data integration methods from the perspective of the multiple-frame paradigm and the assumptions in Section 2.2.

4.1 Small area estimation

Small area estimation can be considered to be a special case of a dual-frame estimation problem in which Assumption (A6) is not met. Here, S_1 is a probability sample from Frame 1 and Frame 2 is often an administrative data source. Both frames are assumed to have complete coverage of the population, but the variable of interest y is measured only in S_1. Auxiliary information $\mathbf{x}$ used to predict y is measured in both samples. Beaumont and Rao (2021) discussed integrating probability and nonprobability samples through the use of the Fay-Herriot (1979) estimator with small area estimation techniques.

A composite small area estimator (Rao and Molina, 2015) of the population mean η_a in area a is of the form

$$\hat{\eta}_a = \theta_a \hat{\eta}_a^{(1)} + (1 - \theta_a)\, \hat{\eta}_a^{(2)},$$

where $\hat{\eta}_a^{(1)}$ is the direct estimator for the sample mean in area a from S_1 (which may have large variance or may not exist), $\hat{\eta}_a^{(2)} = \mathbf{x}_a^T \hat{\boldsymbol{\beta}}$ is a predicted value from a regression model, and θ_a is a compositing factor. For the Fay-Herriot estimator, θ_a depends on the relative precision of the two estimators under an assumed regression model whose parameters are estimated from S_1. For the estimator $\hat{\eta}_a$, the variable y is measured differently in the two frames – predicted values are used for Frame 2 – and different compositing factors are used in different areas.

4.2 Mass imputation and sample matching

Suppose that S_1 is a full-response probability sample from Frame 1, but the variable of interest y is not measured in S_1. However, y is measured in S_2 from Frame 2, and auxiliary variables $\mathbf{x}$ are measured in both samples. Let $\tilde{y}_i$ be the predicted value of y_i from an imputation model, relating y_i to

$\mathbf{x}_i$, that is developed on S_2 and let $\tilde{Y}^{(1)} = \sum_{i \in S_1} w_i^{(1)} \tilde{y}_i$ and $\tilde{Y}_d^{(1)} = \sum_{i \in S_1} w_i^{(1)} \delta_i(d) \tilde{y}_i$ be the estimated population and domain-d totals from S_1 using the imputed values.

Similarly to small area estimation, mass imputation fits into the dual-frame context by relaxing Assumption (A6) of no measurement error. Kim and Rao (2012) and Chipperfield, Chessman and Lim (2012) considered the situation where both frames are complete and S_1 and S_2 are both probability samples. The frames can differ – Frame 1, for example, might be an area frame and Frame 2 might be a population register – but both are assumed to have full coverage. Chipperfield et al. (2012) used a composite estimator

$$\hat{Y}_{\text{imp}} = \theta \tilde{Y}^{(1)} + (1 - \theta) \hat{Y}^{(2)}, \tag{4.1}$$

where the optimal value of the compositing factor θ minimizes the variance (considering both the sampling and imputation variability). Kim and Rao (2012) proposed adding a correction for bias with the estimator

$$\tilde{Y}^{(1)} + \sum_{i \in S_2} w_i^{(2)} (y_i - \tilde{y}_i);$$

this estimator is of the same form as (4.1) with $\theta = 1$ if the estimated parameters in the imputation model are required to satisfy $\sum_{i \in S_2} w_i^{(2)} (y_i - \tilde{y}_i) = 0.$

If the imputation model produces unbiased and accurate predictions for y_i, combining the samples augments the effective sample size for calculating estimates. When both samples are probability samples with full coverage, it is possible to perform model diagnostics on S_2. Chipperfield et al. (2012) suggested several diagnostics, including testing the imputation model on small areas, investigating whether it is possible to predict survey membership from the value of y_i (for S_2) or $\tilde{y}_i$ (for S_1), and studying the sensitivity of the mean squared error to different levels of bias in $\tilde{Y}^{(1)}$. The sensitivity of the diagnostics, however, depends on the quality and size of S_2. If S_2 is small relative to S_1, S_1 may contain subpopulations that are not well represented in S_2 and are poorly fit by the imputation model.

The situation becomes more complicated when Frame 2 is incomplete or when S_2 has selection bias. When domain $\{1\}$ is nonempty as in Figure 2.2(a), then the composite estimator with imputed values becomes

$$\hat{Y}_{\text{imp}} = \tilde{Y}_{\{1\}}^{(1)} + \theta \tilde{Y}_{\{1,2\}}^{(1)} + (1 - \theta) \hat{Y}_{\{1,2\}}^{(2)}. \tag{4.2}$$

The properties of the estimator in (4.2) depend on how well the imputation model predicts the values of y_i in S_1. Several imputation methods have been proposed. With sample matching (Rivers, 2007), $\tilde{y}_i$ for observation i in S_i is set equal to the value of y_i of the observation's nearest neighbor (with respect to the values of $\mathbf{x}$) in S_2. Rivers (2007), considering the situation in which S_2 is a convenience sample, took $\theta = 1$ in (4.2) and used the information in S_2 for the sole purpose of finding the imputed values $\tilde{y}_i$ for S_1. Yang, Kim and Hwang (2021) studied theoretical properties of mass-imputed estimators that employ nearest neighbor methods.

Chen, Li and Wu (2020), building on the work of Lee (2006), Lee and Valliant (2009), and Valliant and Dever (2011) on using propensity score weighting to estimate population characteristics from a nonprobability sample, proposed a "doubly-robust" estimator for the situation where $\mathbf{x}_i$ is measured in both surveys but y_i is measured only in nonprobability sample S_2. Let $R_i^{(2)} = 1$ if population unit i is in S_2 and 0 otherwise. Under strong assumptions that (1) $R_i^{(2)}$ and y_i are independent given covariates $\mathbf{x}_i$, (2) $\pi_i^{(2)} = P\left(R_i^{(2)} = 1\right) > 0$ for all population units i, and (3) $R_i^{(2)}$ and $R_j^{(2)}$ are conditionally independent given $\mathbf{x}$, they estimated $\pi_i^{(2)}$ as a function of $\mathbf{x}_i$, using information in S_1, and proposed the estimator

$$\hat{Y}_{\text{DR}} = \sum_{i \in S_1} w_i^{(1)} \tilde{y}_i + \sum_{i \in S_2} \frac{1}{\hat{\pi}_i^{(2)}} (y_i - \tilde{y}_i),$$

where $\tilde{y}_i$ is an imputation prediction for the unknown values of y in S_1 (developed using the information in S_2). The estimator $\hat{Y}_{\text{DR}}$ is approximately unbiased for Y if either the imputation model or the model predicting $\pi_i^{(2)}$ is correct. If the imputation model is correct, then the first term of $\hat{Y}_{\text{DR}}$ is approximately unbiased for Y and the second term has expected value 0. If the model predicting $\pi_i^{(2)}$ is correct, then $\sum_{i \in S_2} y_i \big/ \hat{\pi}_i^{(2)}$ is approximately unbiased for Y and $E\left[\sum_{i \in S_1} w_i^{(1)} \tilde{y}_i - \sum_{i \in S_2} \tilde{y}_i \big/ \hat{\pi}_i^{(2)}\right] \approx 0.$ If neither model is correct, however, $\hat{Y}_{\text{DR}}$ may have large bias.

Kim and Tam (2021) considered an extension of the situation in Section 3.4 in which y_i is not measured in S_1, or is measured differently than in S_2, and proposed substituting an imputed value $\tilde{y}_i$ for y_i in the estimators from S_1 in (3.6), obtaining the estimator in (4.2) with $\theta = 0$; they calibrated this estimator to the known domain size $N_{\{1,2\}}$.

4.3 Imputation and the NSAF

The estimators in Section 4.2 impute a predicted value $\tilde{y}_i$ for the unknown value of y_i in S_1. All have the strong assumption that the imputation model developed on S_2 applies to the units in domain $\{1\}$. As Lu (2014b) noted when studying regression for dual-frame surveys, relationships between $\mathbf{x}$ and y may differ across domains. Thus, an imputation model developed on a sample from an incomplete frame, or on a sample with selection bias, may provide poor predictions for y in other parts of the population. Moreover, without data on y in the part of the population that is imputed, it may not be possible to assess the quality of the predictions.

A dual-frame survey was taken for the NSAF because of concern that characteristics of interest might differ for telephone and nontelephone households. Let $y_i = 1$ if child i is in a household that is below 200 percent of the poverty threshold, and 0 otherwise. Using the full sample from both frames (Urban Institute and Child Trends, 2007) an estimated 42.2 percent of children lived in households below 200 percent of the poverty threshold, with standard error 0.5 percent. The estimated percentage from the RDD sample was 38.6 percent and the estimated percentage from the area sample was 93.4 percent. Children in the nontelephone households, sampled from the area frame, were much more likely to be living in poverty.

Now suppose that the NSAF had not measured poverty and income variables in the area sample, and y_i was imputed using regression relationships developed in the RDD sample. In many surveys, the only information available for developing an imputation model is demographic variables. Fitting a logistic regression model to the RDD sample that predicts y from race (with categories white, black, and other), and assigning each child in the area sample to the category with highest predicted probability, results in an estimate of 30.5 percent of children in the area sample living in poverty – a lower value than in the RDD sample. Adding an indicator variable for living in a single-parent household to the model, the estimated percentage for the area sample goes up to 51.9 percent. Both of these estimates, and estimates calculated using cell-mean imputations, are far below the percentage of 93.4 percent from the real data.

The problem, of course, is that the auxiliary information is not rich enough to provide a good prediction of poverty in the area sample. The key feature of the data, and the reason that Waksberg and his colleagues used a dual-frame survey, is that being without a telephone is highly associated with poverty. That association cannot be estimated from the RDD sample where all households have telephones. It might be possible to develop an imputation model using information from other surveys such as the Current Population Survey, where both telephone and non-telephone households are sampled, but I could not find an imputation model predicting y from non-income variables in the RDD sample that provided good predictions.

The nontelephone households were a small part of the population for the NSAF, but the differences between the multivariate relationships in the telephone and nontelephone households were so great that the imputation only slightly reduced bias. If poverty had not been measured for the nontelephone sample, however, and the published statistics had relied only on the imputations, there would have been no way to detect the bias.

4.4 Domain misclassification

One major challenge for combining data using a multiple-frame approach is identifying the domain membership (or multiplicity) of units in the data sources. This is challenging even for surveys that are designed to make use of multiple frames.

The NSAF was designed as a screening survey where telephone households were excluded from the area sample. All households sampled from Frame 2, the RDD frame, were correctly classified since they were contacted by telephone. The more difficult part was obtaining the correct domain classification for households in the area-frame sample. Initial prescreening questions asked whether the household had any working telephones; those that answered no were transferred to the telephone interviewer who conducted the detailed interview. The telephone interviewer administered another brief screening interview and asked again about telephone service. An additional 7 percent of households were excluded after answering the more detailed questions about telephone ownership. Some had told the in-person interviewer that they did not have a telephone because they thought the interviewer wanted to borrow it. Others had misunderstood the question about telephone ownership – one respondent, answering the prescreening questions in the living room, thought they applied only to telephones in the living room and did not mention the telephone in the bedroom (Cunningham, Shapiro and Brick, 1999). Although the second screening interview may have corrected for misclassification from respondents who mistakenly said they did not have a telephone during the prescreening, there was no remedy for potential misclassification from respondents who responded in prescreening that they had a telephone when in fact they did not. Misclassification in this direction may have been part of the reason the investigators had a smaller sample size of nontelephone households than they had anticipated.

In dual-frame telephone surveys, the domain for Figure 2.2(b) (cell only, landline only, or both) is usually determined by asking the respondent about other available telephones and, sometimes, the relative amount each type of telephone is used. Brick, Flores-Cervantes, Lee and Norman (2011) found that their landline samples and cell samples both had smaller estimated proportions of dual users than expected from statistics collected on telephone ownership in the National Health Interview Survey. They conjectured that this was because of persons who had access to both types of telephones but rarely used one of them.

Domain membership may be unknown or difficult to estimate when combining existing data sources. In some cases, as when administrative lists are combined, it may be possible to link records, or the data files may contain information that indicates whether the unit is in other frames. In others, there may be little or no information available on domain membership. How can one know whether a participant in an opt-in panel survey is also in a frame of Medicare recipients if no questions about Medicare are asked in the survey?

Lohr (2011) found that even a small amount of domain misclassification could create large biases in dual-frame estimators; moreover, calibration to domain counts that were based on misclassifications could worsen the bias. She proposed a method for adjusting for bias due to domain misclassification, assuming that misclassification probabilities P (observation classified in domain $d' \mid$ observation actually in domain d) are known or can be accurately estimated for different population subgroups. Lin, Liu and Stokes (2019) studied a similar method using misclassification probabilities P (observation actually in domain $d \mid$ observation classified in domain d').

It may be possible to use multiple-frame methods when domain membership is unknown if the probability that unit i is in domain d can be estimated from auxiliary information $\mathbf{x}_i$ known for all sampled units. Kim and Tam (2021) proposed substituting an estimator for the unknown domain membership for the situation in Section 3.4 where S_2 is a census of a subset of the population. They set $\tilde{\delta}_i(\{1, 2\}) = 1$ if the predicted probability that unit $i \in S_1$ was in domain $\{1, 2\}$, $\hat{P}\left[\delta_i(\{1, 2\}) = 1 \mid \mathbf{x}_i\right]$, exceeded 1/2, and estimated the population total for domain $\{1\}$ as $\sum_{i \in S_1} w_i^{(1)} \left[1 - \tilde{\delta}_i(\{1, 2\})\right] y_i$.

When domain membership is imputed, the mean squared error depends on the accuracy of the domain imputations as well as design features and nonresponse bias in S_1. More research is needed to establish statistical properties of estimators when domain membership is estimated. It may also be desired to study alternative estimators that use the predicted probabilities directly to estimate the total in domain $\{1\}$ as $\sum_{i \in S_1} w_i^{(1)} \hat{P}\left[\delta_i(\{1, 2\}) = 0 \mid \mathbf{x}_i\right] y_i$.

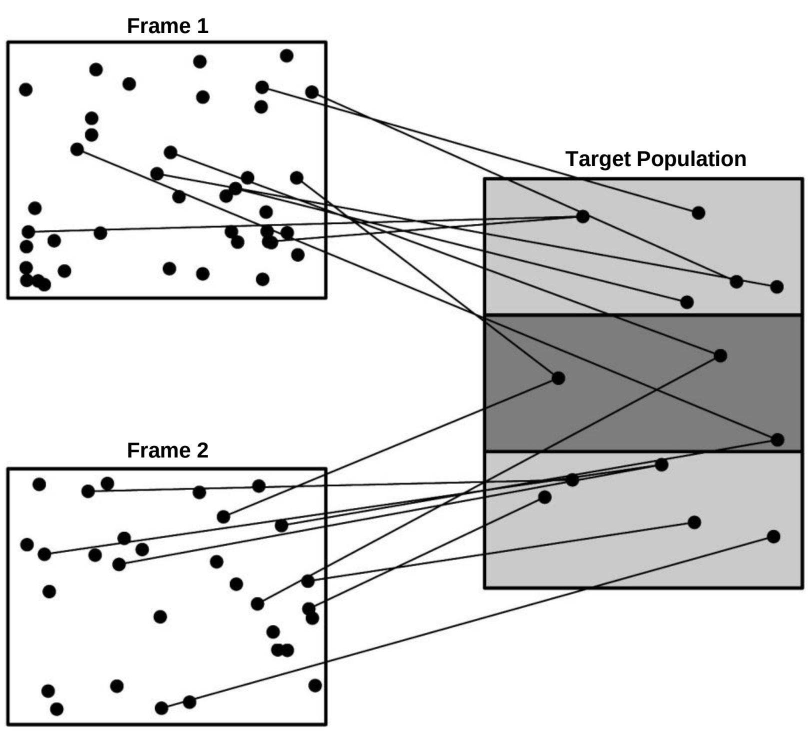
Dever (2018) used sample matching to evaluate the frame overlap for a probability sample S_1, taken from an address-based sampling frame, and a nonprobability sample S_2 recruited from social media sites. She investigated the percentage of respondents in S_1 who had no close match in S_2. Although this procedure does not provide an unbiased estimate of the size of domain {1}, a large percentage of unmatched cases for large samples can indicate that S_2 represents a different population than S_1.

4.5 Indirect sampling and capture-recapture estimation

Sections 4.2 to 4.4 looked at extensions of multiple-frame estimators that relaxed Assumptions (A2), (A4), and (A6). All of these, though, assumed that at least one of the frames, or their union, had full coverage. Let's now look at an example where Assumption (A1) of full coverage is relaxed, and the multiple frames are used to estimate the population size.

In indirect sampling, the target population consists of units that are linked to units in the sampling frame but are not necessarily in the frame (Lavallée, 2007) – units in the target population are sampled indirectly through the links to the sampling units in the frame. Lavallée and Rivest (2012) extended the idea to multiple-frame sampling. As an example, suppose the target population consists of home care workers, who provide paid care for elderly, ill, or disabled persons in their homes. Frame 1 might be a list of persons receiving Medicare benefits, and Frame 2 might be a list of home health care aides from employment or licensing agencies. Persons in the Frame-1 sample are asked to identify workers who provide them with home care, who are then interviewed. A sample of workers from Frame 2 is also interviewed. The home care workers identified from the Frame-1 sample may have links to multiple persons in Frame 1 and may also be in Frame 2. Similarly, persons in the Frame-2 sample may also have links to units in Frame 1. An example of linkage structure is shown in Figure 4.1.

Figure 4.1 Indirect sampling with two frames linked to the target population. Units in the dark shaded area have links to both frames.



With indirect sampling, the Q frames can contain different types of units (the situation with different types of units was also considered by Hartley, 1974). We are not interested in overlap of the sampling frames (shown as nonoverlapping in Figure 4.1 because they contain different types of units) but in the overlap for the units in the target population. Sampled units in the target population have multiple chances of selection if they are linked to multiple units in one or both sampling frames.

Let $l_{j,k}^{(q)} = 1$ if unit j from Frame q is linked to unit k in the target population, and let $L_k^{(q)}$ be the total number of links between unit k in the target population and Frame q (assumed to be knowable from asking unit k). Then an estimator $\hat{Y}^{(q)}$ can be found for each frame using the links as

$$\hat{Y}^{(q)} = \sum_{j \in S^{(q)}} w_j^{(q)} \sum_k \frac{l_{j,k}^{(q)}}{L_k^{(q)}} y_k = \sum_k u_k^{(q)} y_k,$$

where

$$u_k^{(q)} = \sum_{j \in S^{(q)}} w_j^{(q)} \frac{l_{j,k}^{(q)}}{L_k^{(q)}}.$$

In the context of our example, person j in S_1 would say they receive paid home care from provider k, resulting in $l_{j,k}^{(1)} = 1.$ Then the linked home care provider would be asked about how many other persons they work for who receive Medicare (assume they would know this or it could be determined from other sources), giving the value $L_k^{(1)}$. The quantity $u_k^{(q)}$ sums the weights of the units in S_q with links to unit k, adjusting for the multiplicity of the links to that frame. If $w_j^{(q)} = 1 / \pi_j^{(q)}$, then

$$E[u_k^{(q)}] = E\left[\sum_{j \in S^{(q)}} w_j^{(q)} \frac{l_{j,k}^{(q)}}{L_k^{(q)}}\right] = a_k^{(q)},$$

where $a_k^{(q)} = 1$ if target population member k is linked to at least one unit in Frame q and 0 otherwise.

Multiple-frame methods may then be used to estimate characteristics of the population of home care providers, assuming that unit k linked from S_q provides accurate information on (1) the number of links to members of Frame $q\,(L_k^{(q)})$, needed for multiplicity adjustments with Frame q, and (2) whether they are also linked to the other frame(s) $(a_k^{(f)}$ for $f \neq q)$, needed to adjust for the multiplicity of linkage from different frames.

Lavallée and Rivest (2012) noted that if the union of the two frames has incomplete coverage – Assumption (A1) is violated – the samples from the two frames can be used to estimate the size of the target population. Let $\hat{T}^{(q)} = \sum_k u_k^{(q)}$ for $k = 1, 2.$ Then $E\left[\hat{T}^{(q)}\right]$ is the number of target population members that can be linked from Frame q. Each sample also provides an estimate of the number of target population units that can be linked from both frames: $\hat{T}_{\{1,2\}}^{(1)} = \sum_k u_k^{(1)} a_k^{(2)}$ and $\hat{T}_{\{1,2\}}^{(2)} = \sum_k u_k^{(2)} a_k^{(1)}$. These can be composited to obtain an estimator $\hat{T}_{\{1,2\}}$ of the number of persons in the target population who can be captured from both frames.

The Lincoln-Petersen capture-recapture estimator of population size can then be used. Under the strong assumption that being captured by Frame 1 is independent of being captured by Frame 2, the total number of home care providers can be estimated by $\hat{T}^{(1)}\hat{T}^{(2)}/\hat{T}_{\{1,2\}}$. In some cases where the independence assumption is not met for the entire population, it may be approximately met on subpopulations whose estimated numbers can be summed. If there are more than two frames, loglinear models may be used to explore associations among the frames (Lohr, 2022, Chapter 14); Zhang (2019) presented a model for the situation in which frames may contain misclassified units.

Alleva, Arbia, Falorsi, Nardelli and Zuliani (2020) proposed using indirect multiple-frame sampling to estimate the number of people infected by SARS-CoV-2 during the early stages of the COVID-19 pandemic in 2020 – information needed for estimating transmissibility and infection parameters in epidemiologic models. In this application, Frame 1 consists of persons with verified infections (perhaps obtained from hospitals, quarantine centers, or clinics), and Frame 2 consists of other persons; the persons in S_2 are administered a test for SARS-CoV-2. The linked sample consists of persons who had contact within the past 14 days with anyone in S_1 or with a member of S_2 who tested positive.

5. Design of data collection systems

Section 4 discussed how estimators for integrated data can be thought of within a multiple-frame survey structure. This structure can also be used when designing data collection systems that make use of multiple sources. Hartley (1962) derived the values of $n^{(1)}$, $n^{(2)}$, and θ that minimize the variance of $\hat{Y}(\theta)$ in (3.2) when S_1 and S_2 are both simple random samples. His basic method can be extended to explore effects of sample design choices for other situations by considering mean squared errors under a range of potential bias assumptions.

There has been a substantial amount of work on optimal design and effects of nonresponse for dual-frame cellular/landline telephone surveys. Brick, Dipko, Presser, Tucker and Yuan (2006) and Brick et al. (2011) investigated nonsampling errors; Lu, Sahr, Iachan, Denker, Duffy and Weston (2013) performed a simulation study to calculate the anticipated mean squared error under various cost models and potential biases. Lohr and Brick (2014), studying allocation of resources in dual-frame telephone surveys with nonresponse, found that for some cost structures a screening survey, in which respondents with landlines are screened out of the cell phone sample, was more cost-efficient than an overlap survey. Levine and Harter (2015) presented graphical results to provide allocation guidance, considering the variance inflation from weight variation. Chen, Stubblefield and Stoner (2021) considered the design problem of oversampling minority populations in dual-frame telephone surveys, using optimal allocation methods from stratified sampling. Most of these articles focus on minimizing the variance of estimates for a fixed cost, and do not consider the effects of potential bias.

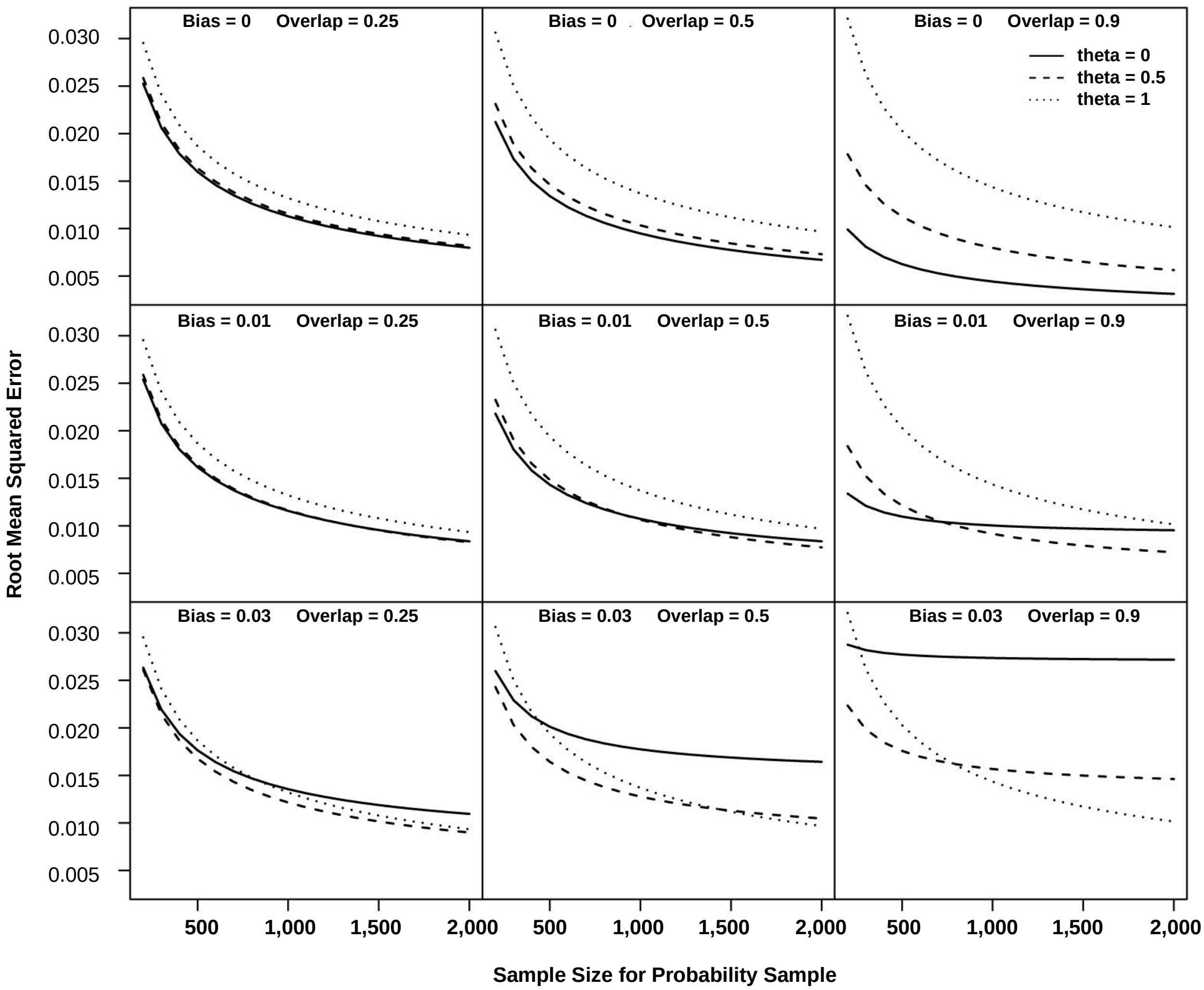
A number of papers in the 1980s studied error structures and designs for dual-frame surveys, typically supplementing a sample from an RDD frame with a sample from an area frame that was assumed to have full coverage. Biemer (1984) and Choudhry (1989) explored optimal designs theoretically and through simulation studies. Groves and Lepkowski (1985, 1986), and Traugott, Groves and Lepkowski (1987) investigated dual-frame designs with a view to minimizing mean squared error when estimates from the RDD frame may be biased. Lepkowski and Groves (1986) found that as the amount of bias increased in the RDD sample, its optimal allocation decreased, reaching an allocation of zero when the bias was 9 percent of the anticipated estimated percentage.

A small amount of bias can have similar effect for the situation considered in Section 3.4, where a census is taken from incomplete Frame 2 and a high-quality probability sample is taken from complete Frame 1. The plots in Figure 5.1 show the root mean squared error (RMSE) for an estimated proportion when S_1 is a simple random sample of size n and S_2 is a census of domain $\{1, 2\}$, for combinations of overlap size $N_{\{1,2\}} / N$ in $\{0.25, 0.5, 0.9\}$ and bias in $\{0, 0.01, 0.03\}$. The population proportion is 0.2 in domain $\{1\}$ and 0.3 in domain $\{1, 2\}$, and the overall population proportion is estimated using $\hat{Y}(\theta) / N$ for $\hat{Y}(\theta)$ in (3.2). The lines show the RMSE for each n for $\theta = 1$ (S_2 is not used at all), $\theta = 0$ (the estimated proportion in domain $\{1, 2\}$ comes from S_2 and S_1 contributes only for estimating the proportion in domain $\{1\}$), and $\theta = 1/2$. In the bottom row of plots, the bias from S_2 begins dominating the RMSE even for relatively small sample sizes from S_1. A small amount of measurement bias can cancel the supposed advantage from data integration. This example assumes the error in S_2 is from measurement bias, but is similar in spirit to the example in Meng (2018), which shows that even when the selection bias from a convenience sample is small, a simple random sample of size 400 may have more useful information than a convenience sample of size 500 million.

As Thompson (2019) noted, many of the methods that have been developed for combining data from multiple sources have been situation-specific, with solutions tailored to the particular circumstances of that problem. One would not expect these methods to perform as well, on average, for other situations because of regression-to-the-mean effects. Before adopting a data combination method, it may be desirable to perform additional simulation studies that consider outcomes when the model assumptions are not met.

Lohr and Raghunathan (2017) discussed issues for designing data collection systems that leverage multiple data sources, focusing on the situation in which a probability survey is used in conjunction with administrative data sources that cover parts of the population. They considered using administrative data sources for (1) improving the frame for the probability sample, (2) providing contextual information for interpreting the survey data, (3) providing information for nonresponse follow-up and bias assessment, and (4) designing the entire data collection system to take advantage of inexpensive data collection afforded by some of the frames while obtaining complete coverage from the probability survey. Thinking of the design problem in the multiple-frame paradigm can be helpful for the last point. Lohr and Raghunathan (2017) suggested that when Frame 1 is complete but expensive to sample, while Frame 2 is incomplete but less expensive to sample – this includes the situation considered in Section 3.4 of this paper – it may be desirable to use a two-phase screening survey for the sample from Frame 1 and rely on the sample from Frame 2 to supply information for domain $\{1, 2\}$. That is the strategy that Waksberg and colleagues followed for designing the NSAF.

Figure 5.1 Root mean squared error of estimated population proportion under differing amounts of overlap and bias.



When there may be measurement error or domain misclassification, however, a more robust design may be preferred. Optimal designs for dual-frame surveys allocate resources so as to minimize the variance of estimated population totals of interest for fixed cost. Designs that are optimal under Assumptions (A1) to (A6) are not necessarily optimal when some of those assumptions are violated. The multiple-frame structure allows consideration of potential design performance under relaxation of the assumptions.

Hartley (1962) showed that a dual-frame survey resulted in substantial improvements in efficiency for the situations in Figure 2.2(a, b) when data can be inexpensively obtained from Frame 2 and $N_{\{1,2\}}/N$ is large. But when Frame 1 is complete, and the costs are comparable or $N_{\{1,2\}}/N$ is small, the extra complexity from using a dual-frame survey may outweigh its advantages. If, in addition, there is likely to be domain misclassification or if y is measured differently across the surveys, a dual-frame survey will be more complicated than a single sample from Frame 1 and may produce biased estimates.

On the other hand, using multiple data sources can also help assess nonsampling errors. Hartley (1974) wrote that when he presented his work on multiple-frame surveys at a conference, a discussant suggested that a "fairer" comparison would be to compare the variance from a dual-frame sample with that from a single sample of the same cost from the incomplete but cheap frame. Hartley responded (page 107): "The difficulty about this is, of course, that the bias through incompleteness may be of a magnitude which would make the single frame survey useless. If no a priori information on this bias is available, the two frame survey can in fact be regarded as an economical method of *measuring* this bias *and* eliminating it."

Thus, it may be desirable to design the data collection system with multiple goals of (1) obtaining estimates of key population quantities with small mean squared error, (2) assessing nonsampling errors from data sources, and (3) providing information to improve future survey designs. Some of the issues to consider include:

- Quality and stability of data sources. Classical multiple-frame survey design theory assumes that the frames are fixed. But it may be desired to use alternative data sources in which the frame is changing over time (for example, web-scraped prices) to help provide more timely information in coordination with a probability survey. Theory is needed on how to do this. If relying on data supplied by an external source, will those data continue to be available, and in the same form?
- Measurement of domain membership. If possible, information should be collected from each source to allow accurate determination of domain membership. If the information items collected in administrative sources cannot be altered, sometimes items can be added to probability samples that allow domain determination.
- Redundancy. For the situation in Section 3.4, where a census of part of the population is supplemented by a probability sample, a screening design might be optimal for S_1. But a screening design does not allow assessment of potential differences in measurement from the two samples. Some degree of overlap may be desired among the data sources in order to assess differences among the domain estimates from different sources.

  When an imputation model is developed for y based on relationships between y and $\mathbf{x}$ from a data source with incomplete coverage, there is a danger that this model will not apply to the other parts of the population. It may be desired to take a small sample from the uncovered part of the population for purposes of evaluating the model.
- Relative amounts of information for different domains. When data sources include administrative records or large convenience samples, there may be much more information about some parts of the population than others. The issue becomes how to obtain reliable information on the missing parts of the population. When that information comes from a

sample, there may be high weight variation. Levine and Harter (2015) studied the issue of weight variation in dual-frame telephone surveys. Some of the weight variation may be reduced by obtaining additional administrative data sources on underrepresented subpopulations, but there is a danger that, as organizations move away from expensive probability samples, some subpopulations will be omitted from all sources.

- Robustness to design assumptions. Designs that are optimal in theory often turn out to be less so in practice. Exploring the anticipated design performance under violations of the assumptions can be helpful for modifying a theoretically optimal design. In some cases, combining information across sources may result in worse estimates than using a single source, or it may be decided that the gains from combining data are not worth the extra trouble.

  Waksberg (1998) advised: "Do not treat statistical procedures as mechanical operations; be prepared for the unexpected." Having a design with some robustness to the assumptions gives flexibility for unexpected problems.

- Auxiliary information. Many of the methods for integrating data rely on auxiliary information to perform imputations or predict domain membership. Mercer, Lau and Kennedy (2018) argued that for calibration, the richness of the auxiliary information is far more important than the particular method used to calibrate, and the same is true for other data combination methods. Having rich auxiliary information (beyond demographic variables) allows for better data integration models – and for better assessment of their performance.

Waksberg argued that a survey statistician needs to look at the entirety of the problem, not just the optimal design for measuring a single variable. He said that a sampling statistician should "think not only about the specific questions that are asked, but the broader aspects of these questions: whether the questions make sense and can be solved, or whether they should be modified or changed. This is how I've tried to have people with whom I work think about the issues: Here's a question, how do you respond to this specific question? Is it the right question? What statistics will you get by a narrow interpretation of the question, and is there a better way to proceed?" (Morganstein and Marker, 2000, page 304).

In this paper, I have suggested that multiple-frame surveys can serve as an organizing structure for designing and evaluating data-integration systems. This can help clarify the strengths and weaknesses of each source and, perhaps, result in a better way to proceed.

Acknowledgements

I am grateful to the Waksberg Award Committee for selecting me for this honor, to Mike Brick for helpful discussions, and to the associate editor and two referees whose constructive suggestions improved this paper.

References

Aidara, C.A.T. (2019). Quasi random resampling designs for multiple frame surveys. *Statistica*, 79, 321-338.

Alleva, G., Arbia, G., Falorsi, P.D., Nardelli, V. and Zuliani, A. (2020). A sampling approach for the estimation of the critical parameters of the SARS-CoV-2 epidemic: An operational design. https://arxiv.org/ftp/arxiv/papers/2004/2004.06068.pdf, last accessed March 28, 2021.

Arcos, A., Martínez, S., Rueda, M. and Martínez, H. (2017). Distribution function estimates from dual frame context. *Journal of Computational and Applied Mathematics,* 318, 242-252.

Arcos, A., Rueda, M., Trujillo, M. and Molina, D. (2015). Review of estimation methods for landline and cell phone surveys. *Sociological Methods & Research,* 44, 458-485.

Baffour, B., Haynes, M., Western, M., Pennay, D., Misson, S. and Martinez, A. (2016). Weighting strategies for combining data from dual-frame telephone surveys: Emerging evidence from Australia. *Journal of Official Statistics*, 32, 549-578.

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.

Beaumont, J.-F. (2020). Are probability surveys bound to disappear for the production of official statistics? *Survey Methodology,* 46, 1, 1-28. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2020001/article/00001-eng.pdf.

Beaumont, J.-F., and Rao, J.N.K. (2021). Pitfalls of making inferences from non-probability samples: Can data integration through probability samples provide remedies? *The Survey Statistician,* 83, 11-22.

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., Flores-Cervantes, I.F., Lee, S. and Norman, G. (2011). Nonsampling errors in dual frame telephone surveys. *Survey Methodology,* 37, 1, 1-12. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011001/article/11443-eng.pdf.

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

Brick, J.M., Shapiro, G., Flores-Cervantes, I., Ferraro, D. and Strickler, T. (1999). *1997 NSAF Snapshot Survey Weights*. Washington, DC: Urban Institute.

Burke, J., Mohadjer, L., Green, J., Waksberg, J., Kirsch, I.S. and Kolstad, A. (1994). Composite estimation in national and state surveys. In *Proceedings of the Survey Research Methods Section*, 873-878. Alexandria, VA: American Statistical Association.

Chauvet, G. (2016). Variance estimation for the 2006 French housing survey. *Mathematical Population Studies*, 23, 147-163.

Chauvet, G., and de Marsac, G.T. (2014). Estimation methods on multiple sampling frames in two-stage sampling designs. *Survey Methodology*, 40, 2, 335-346. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2014002/article/14090-eng.pdf.

Chen, S., Stubblefield, A. and Stoner, J.A. (2021). Oversampling of minority populations through dual-frame surveys. *Journal of Survey Statistics and Methodology*, 9, 626-649.

Chen, Y., Li, P. and Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115, 2011-2021.

Chipperfield, J., Chessman, J. and Lim, R. (2012). Combining household surveys using mass imputation to estimate population totals. *Australian & New Zealand Journal of Statistics*, 54, 223-238.

Choudhry, G.H. (1989). Cost-variable optimization of dual frame design for estimating proportions. In *Proceedings of the Survey Research Methods Section*, 566-571. Alexandria, VA: American Statistical Association.

Chu, A., Brick, J.M. and Kalton, G. (1999). Weights for combining surveys across time or space. *Bulletin of the International Statistical Institute*, 2, 103-104.

Citro, C.F. (2014). From multiple modes for surveys to multiple data sources for estimates. *Survey Methodology*, 40, 2, 137-161. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2014002/article/14128-eng.pdf.

Cohen, S.B., DiGaetano, R. and Waksberg, J. (1988). Sample design of the NMES survey of American Indians and Alaska Natives. In *Proceedings of the Survey Research Methods Section*, 740-745. Alexandria, VA: American Statistical Association.

Cunningham, P., Shapiro, G. and Brick, J.M. (1999). *1997 NSAF In-Person Survey Methods*. Washington, DC: Urban Institute.

Dever, J.A. (2018). Combining probability and nonprobability samples to form efficient hybrid estimates: An evaluation of the common support assumption. In *Proceedings of the 2018 Federal Committee on Statistical Methodology (FCSM) Research Conference*. https://nces.ed.gov/FCSM/pdf/A4_Dever_2018FCSM.pdf, last accessed July 7, 2021.

Deville, J.-C., and Lavallée, P. (2006). Indirect sampling: The foundations of the generalized weight share method. *Survey Methodology,* 32, 2, 165-176. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006002/article/9551-eng.pdf.

DiGaetano, R., Judkins, D. and Waksberg, J. (1995). Oversampling minority school children. In *Proceedings of the Survey Research Methods Section*, 503-508. Alexandria, VA: American Statistical Association.

Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

Ferraz, C., and Vogel, F. (2015). Multiple frame sampling. In *Handbook on Master Sampling Frames for Agricultural Statistics: Frame Development, Sample Design and Estimation*, 89-106. Rome: Food and Agriculture Organization of the United Nations.

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

Groves, R.M., and Lepkowski, J.M. (1985). Dual frame, mixed mode survey designs. *Journal of Official Statistics*, 1, 263-286.

Groves, R.M., and Lepkowski, J.M. (1986). An experimental implementation of a dual frame telephone sample design. In *Proceedings of the Survey Research Methods Section*, 340-345. Alexandria, VA: American Statistical Association.

Groves, R.M., and Wissoker, D. (1999). *1997 NSAF Early Nonresponse Studies*. Washington, DC: Urban Institute.

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

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

Haziza, D., and Lesage, É. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics,* 32, 129-145.

Hendricks, S., Igra, A. and Waksberg, J. (1980). Ethnic stratification in the California Hypertension Survey. In *Proceedings of the Survey Research Methods Section*, 680-685. Alexandria, VA: American Statistical Association.

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

Kim, J.K., and Rao, J.N.K. (2012). Combining data from two independent surveys: A model-assisted approach. *Biometrika,* 99, 85-100.

Kim, J.K., and Tam, S.-M. (2021). Data integration by combining big data and survey sample data for finite population inference. *International Statistical Review,* 89, 382-401.

Kott, P.S., and Vogel, F.A. (1995). Multiple-frame business surveys. In *Business Survey Methods,* (Eds., B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge and P.S. Kott), 185-203. New York: John Wiley & Sons, Inc.

Lavallée, P. (2007). *Indirect Sampling*. New York: Springer.

Lavallée, P., and Rivest, L.-P. (2012). Capture-recapture sampling and indirect sampling. *Journal of Official Statistics*, 28, 1-27.

Lee, S. (2006). Propensity score adjustment as a weighting scheme for volunteer panel web surveys. *Journal of Official Statistics*, 22, 329-349.

Lee, S., and Valliant, R. (2009). Estimation for volunteer panel web surveys using propensity score adjustment and calibration adjustment. *Sociological Methods & Research,* 37, 319-343.

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

Levine, B., and Harter, R. (2015). Optimal allocation of cell-phone and landline respondents in dual-frame surveys. *Public Opinion Quarterly,* 79, 91-104.

Lin, D., Liu, Z. and Stokes, L. (2019). A method to correct for frame membership error in dual frame estimators. *Survey Methodology,* 45, 3, 543-565. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2019003/article/00008-eng.pdf.

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

Lohr, S.L. (2011). Alternative survey sample designs: Sampling with multiple overlapping frames. *Survey Methodology,* 37, 2, 197-213. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011002/article/11608-eng.pdf.

Lohr, S.L. (2014). When should a multiple frame survey be used? *The Survey Statistician,* 69 (January), 17-21.

Lohr, S.L. (2022). *Sampling: Design and Analysis, Third Edition*. Boca Raton, FL: CRC Press.

Lohr, S.L., and Brick, J.M. (2014). Allocation for dual frame telephone surveys with nonresponse. *Journal of Survey Statistics and Methodology,* 2, 388-409.

Lohr, S.L., and Raghunathan, T.E. (2017). Combining survey data with other data sources. *Statistical Science*, 32, 293-312.

Lohr, S.L., and Rao, J.N.K. (2000). Inference from 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, B., Peng, J. and Sahr, T. (2013). Estimation bias of different design and analytical strategies in dual-frame telephone surveys: An empirical evaluation. *Journal of Statistical Computation and Simulation,* 83, 2352-2368.

Lu, B., Sahr, T., Iachan, R., Denker, M., Duffy, T. and Weston, D. (2013). Design and analysis of dual-frame telephone surveys for health policy research. *World Medical & Health Policy*, 5, 217-232.

Lu, Y. (2014a). Chi-squared tests in dual frame surveys. *Survey Methodology*, 40, 2, 323-334. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2014002/article/14096-eng.pdf.

Lu, Y. (2014b). Regression coefficient estimation in dual frame surveys. *Communications in Statistics – Simulation and Computation*, 43, 1675-1684.

Lu, Y., and Lohr, S. (2010). Gross flow estimation in dual frame surveys. *Survey Methodology*, 36, 1, 13-22. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2010001/article/11248-eng.pdf.

Lu, Y., Fu, Y. and Zhang, G. (2021). Nonparametric regression estimators in dual frame surveys. *Communications in Statistics – Simulation and Computation*, 50, 854-864.

Marks, E., and Waksberg, J. (1966). Evaluation of coverage in the 1960 Census of Population through case-by-case checking. In *Proceedings of the Social Statistics Section*, 62-70. Alexandria, VA: American Statistical Association.

Mecatti, F. (2007). A single frame multiplicity estimator for multiple frame surveys. *Survey Methodology*, 33, 2, 151-157. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2007002/article/10492-eng.pdf.

Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, 12, 685-726.

Mercer, A., Lau, A. and Kennedy, C. (2018). *For Weighting Online Opt-In Samples, What Matters Most?* Washington, DC: Pew Research.

Metcalf, P., and Scott, A. (2009). Using multiple frames in health surveys. *Statistics in Medicine*, 28, 1512-1523.

Montanari, G.E. (1987). Post-sampling efficient QR-prediction in large-sample surveys. *International Statistical Review*, 55, 191-202.

Montanari, G.E. (1998). On regression estimation of finite population means. *Survey Methodology*, 24, 1, 69-77. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1998001/article/3911-eng.pdf.

Morganstein, D., and Marker, D. (2000). A conversation with Joseph Waksberg. *Statistical Science*, 15, 299-312.

National Academies of Sciences, Engineering, and Medicine (2017). *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Washington, DC: National Academies Press.

National Academies of Sciences, Engineering, and Medicine (2018). *Federal Statistics, Multiple Data Sources, and Privacy Protection: Next Steps*. Washington, DC: National Academies Press.

O'Muircheartaigh, C., and Pedlow, S. (2002). Combining samples vs. cumulating cases: A comparison of two weighting strategies in NLSY97. In *Proceedings of the Survey Research Methods Section*, 2557-2562. Alexandria, VA: American Statistical Association.

Ranalli, M.G., Arcos, A., Rueda, M.d.M. and Teodoro, A. (2016). Calibration estimation in dual-frame surveys. *Statistical Methods & Applications*, 25, 321-349.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Rao, J.N.K. (2021). On making valid inferences by combining data from surveys and other sources. *Sankhyā, Series B*, 83-B, 242-272.

Rao, J.N.K., and Molina, I. (2015). *Small Area Estimation, 2nd Ed*. Hoboken, NJ: Wiley.

Rivers, D. (2007). Sampling for web surveys. Paper presented at the Joint Statistical Meetings.

Rueda, M.d.M., Arcos, A., Molina, D. and Ranalli, M.G. (2018). Estimation techniques for ordinal data in multiple frame surveys with complex sampling designs. *International Statistical Review*, 86, 51-67.

Saegusa, T. (2019). Large sample theory for merged data from multiple sources. *The Annals of Statistics*, 47, 1585-1615.

Särndal, C.-E., and Lundström, S. (2005). *Estimation in Surveys with Nonresponse*. Hoboken, NJ: Wiley.

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

Thompson, M.E. (2019). Combining data from new and traditional sources in population surveys. *International Statistical Review*, 87, S79-S89.

Traugott, M.W., Groves, R.M. and Lepkowski, J.M. (1987). Using dual frame designs to reduce nonresponse in telephone surveys. *Public Opinion Quarterly*, 51, 522-539.

Urban Institute and Child Trends (2007). *National Survey of America's Families (NSAF), 1997*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor].

Valliant, R., and Dever, J.A. (2011). Estimating propensity adjustments for volunteer web surveys. *Sociological Methods & Research*, 40, 105-137.

Waksberg, J. (1986). Discussion of papers on new approaches in telephone sample design. In *Proceedings of the Survey Research Methods Section*, 367-369. Alexandria, VA: American Statistical Association.

Waksberg, J. (1995). Distribution of poverty in census block groups (BGs) and implications for sample design. In *Proceedings of the Survey Research Methods Section*, 497-502. Alexandria, VA: American Statistical Association.

Waksberg, J. (1998). The Hansen era: Statistical research and its implementation at the US Census Bureau, 1940-1970. *Journal of Official Statistics*, 14, 119-135.

Waksberg, J., and Pritzker, L. (1969). Changes in census methods. *Journal of the American Statistical Association*, 64, 1141-1149.

Waksberg, J., Judkins, D. and Massey, J.T. (1997b). Geographic-based oversampling in demographic surveys of the United States. *Survey Methodology*, 23, 1, 61-71. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1997001/article/3107-eng.pdf.

Waksberg, J., Brick, J.M., Shapiro, G., Flores-Cervantes, I. and Bell, B. (1997a). Dual-frame RDD and area sample with particular focus on low-income population. In *Proceedings of the Survey Research Methods Section*, 713-718. Alexandria, VA: American Statistical Association.

Waksberg, J., Brick, J.M., Shapiro, G., Flores-Cervantes, I., Bell, B. and Ferraro, D. (1998). Nonresponse and coverage adjustment for a dual-frame survey. *Proceedings: Symposium 97, New Directions in Surveys and Censuses*, 193-198. Ottawa: Statistics Canada.

Wolter, K.M., Ganesh, N., Copeland, K.R., Singleton, J.A. and Khare, M. (2019). Estimation tools for reducing the impact of sampling and nonresponse errors in dual-frame RDD telephone surveys. *Statistics in Medicine*, 38, 4718-4732.

Yang, S., and Kim, J.K. (2020). Statistical data integration in survey sampling: A review. *Japanese Journal of Statistics and Data Science*, 3, 625-650.

Yang, S., Kim, J.K. and Hwang, Y. (2021). Integration of data from probability surveys and big found data for finite population inference using mass imputation. *Survey Methodology*, 47, 1, 29-58. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2021001/article/00004-eng.pdf.

Zhang, L.-C. (2019). Log-linear models of erroneous list data. In *Analysis of Integrated Data* (Eds., L.-C. Zhang and R.L. Chambers), 197-218. Boca Raton, FL: CRC Press.

Zhang, L.-C., and Chambers, R.L. (Eds.) (2019). *Analysis of Integrated Data*. Boca Raton, FL: CRC Press.