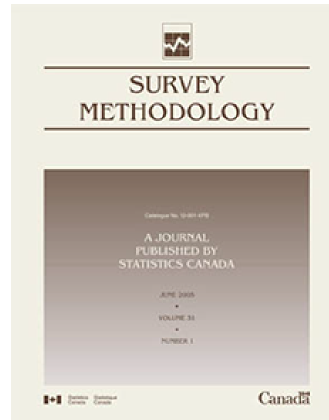
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Trends and directions in sample survey theory and methods



by J.N.K. Rao and Sharon L. Lohr

Release date: June 30, 2025

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

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, the Agency has developed standards of service which its employees observe in serving its clients. 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.

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2025
Vol. 51, No. 1, pp. 81-113
Statistics Canada, Catalogue No. 12-001-X

Trends and directions in sample survey theory and methods

J.N.K. Rao and Sharon L. Lohr[1]

Abstract

Rao (1999) summarized trends in sample survey theory and methods at the turn of the millenium. We provide an updated discussion of some current trends in survey design and estimation methods for the 50th anniversary of *Survey Methodology*. Recent innovations in survey design include research on anticipating nonsampling errors at the design stage and development of balanced and adaptive sampling designs to take advantage of detailed sampling frame information or data gathered during the survey process. Nonparametric and machine learning methods are increasingly used for data editing as well as for model-assisted estimation and nonresponse adjustments. Small area models have been expanded to incorporate spatial and time series information, increase the flexibility and robustness of the linking and variance models, benchmark to large-area direct estimators, and (for unit level models) account for informative sampling designs. The increasing availability of large administrative datasets, sensor and satellite data, and convenience samples has spurred research on how to use these sources – on their own and when integrated with probability samples. We conclude by discussing some frontiers for survey research.

Key Words: Data integration; Machine learning; Nonprobability samples; Small area estimation; Survey design; Survey quality.

1. Introduction

We are grateful to the editor for inviting us to participate in this issue celebrating the contributions of *Survey Methodology*. From its first issue in June 1975, the journal has provided a forum for research "dealing with all phases of methodological development in surveys" (Platek, 1999, page 109). Its widespread availability and attention to practical problems have made the journal an essential resource for survey researchers around the world.

In this article, we summarize some of the trends in survey research over the 50 years of *Survey Methodology*'s existence. We then look at some possible future directions. This article updates "Some Current Trends in Sample Survey Theory and Methods" (Rao, 1999), which reviewed the state of survey sampling at the turn of the millennium, and concentrates on aspects of the discipline that have emerged or changed since then. Rao (2005), Brick (2011), and Rao and Fuller (2017) provided other reviews of recent research in survey sampling.

A review such as this is necessarily limited to a few topics and cannot mention all important developments of the past quarter century. We refer the reader to other papers for reviews of important developments about confidentiality protection, imputation, and measurement error. While attention continues to be focused on improving estimates of the population total $Y = \sum_{i=1}^{N} y_i$ of a study variable y for a population of size N, increasing emphases are placed on new uses of auxiliary information $\mathbf{x}$ for design, estimation, and nonresponse adjustment. The interplay of technological developments and statistical theory opens up new vistas for survey research.

1. J.N.K. Rao, Distinguished Research Professor, Carleton University, Ottawa, Ontario, Canada, K1S 5B6. E-mail: jrao@math.carleton.ca; Sharon Lohr, Professor Emerita, Arizona State University. E-mail: sharon.lohr@asu.edu.

The principal steps for a sample survey identified by Rao and Fuller (2017, page 1) – "data collection and processing, estimation and analysis of data" – remain the same, although procedures and emphases within those steps have changed over the years. Section 2 discusses recent innovations in survey design, data collection, and survey processing, emphasizing developments that lead to improvements of the overall quality of a survey. Section 3 describes advances in estimation through improved weighting, modeling, and variance estimation. It also examines developments for dealing with nonresponse. Section 4 discusses some recent innovations in small area estimation, Section 5 describes work on integrating data from multiple sources, and Section 6 summarizes trends and identifies some opportunities for future research.

2. Survey design and data collection

Rao (1999) emphasized the importance of total survey design, which aims to minimize the mean squared error (MSE) of estimates, considering not only sampling error but also errors from data collection, data processing, measurement error, undercoverage, and nonresponse.

2.1 Data collection and processing

In the late twentieth century, most surveys were conducted in person (if full coverage was desired) or by telephone (for a lower-cost option). Rao (1999, page 4) commented that "Telephone surveys have become popular in recent years." While telephone surveys are still conducted in some regions, many survey organizations have turned to other modes as telephone response rates have plummeted (Kennedy and Hartig, 2019, reported telephone response rates of 6 percent in 2018).

Two survey modes in particular have become more prevalent in North America since 2000. Address lists updated using information from the United States Postal Service now provide nearly complete coverage of U.S. households (Harter, Battaglia, Buskirk, Dillman, English, Fahimi, Frankel, Kennel, McMichael, McPhee, Montaquila, Yancey and Zukerberg, 2016). The development of this address-based sampling frame has led to an increased number of household surveys in which initial contact is by mail. Nonresponse follow-up may then be done by modes with higher response rates such as in-person visits. The 2020 U.S. Census and the 2021 Canadian Census both made initial contact with households through a letter, offered multiple mode choices for filling out the questionnaire with encouragement to respond using the internet, and followed up with nonrespondents in person. The improved address frames for the censuses have almost full coverage and use of cheaper data collection modes for initial contacts reduces costs.

Many surveys collected by private organizations now use online panels, in which samples for individual surveys are drawn from a panel of people who have been recruited as potential survey participants. The panel members may be recruited by probabilistic methods (for example, through a mail survey of randomly chosen addresses) or may form a nonprobability sample (for example, if panel members are recruited through online advertisements). As of 2019, more than 80 percent of U.S. public opinion polls used online nonprobability panels (Kennedy, Hatley, Lau, Mercer, Keeter, Ferno and Asare-Marfo, 2021).

These innovations in data collection have spurred a great deal of empirical research on how mode, or use of multiple modes, affects survey quality (Couper, 2017; De Leeuw, 2018; Olson, Smyth, Horwitz, Keeter, Lesser, Marken, Mathiowetz, McCarthy, O'Brien, Opsomer, Steiger, Sterrett, Su, Suzer-Gurtekin, Turakhia and Wagner, 2021). For example, Kennedy, Mercer and Lau (2024) compared population characteristics estimated from respondents to six online samples with benchmarks from high-quality U.S. government surveys. Three of the online samples had "opt-in" recruitment; the other three recruited panel members through an address-based mail probability sample and then administered surveys online. Although response rates for the probability samples were less than 10 percent, the average absolute error after calibration was more than twice as high for the opt-in samples as for the probability samples, with especially high errors for persons aged 18-29, persons who self-identified as Hispanic, and questions related to receiving government benefits. Kennedy et al. (2024) suggested that some opt-in survey participants may tend to say "yes" regardless of the question asked, and called for more research on the accuracy of demographic information used for weighting that opt-in respondents provide.

Another emerging area is the potential of artificial intelligence methods for data collection and processing (Yung, Karkimaa, Scannapieco, Barcarolli, Zardetto, Sanchez, Braaksma, Buelens and Burger, 2018; United Nations Economic Commission for Europe, 2021). Beck, Dumpert and Feuerhake (2018) reported on 136 projects in national statistics offices that involved machine learning methods. The methods were most commonly used for classification, identifying duplicate records or outliers, and imputing missing items.

We give three examples to illustrate types of data processing problems that can be addressed with artificial intelligence methods. Hibben, Smith, Hoppe, Ryan, Rogers, Scanlon and Miller (2022) trained a natural language processing model to interpret responses to open-ended questions and distinguish valid responses from item nonresponse and gibberish answers.

Roberson (2021) addressed a problem of response burden in the 2017 U.S. Economic Census, which had a long and time-consuming questionnaire section on goods and services provided. Roberson explored whether businesses could alternatively supply Universal Product Codes, already in their databases, to the Census Bureau. A supervised machine learning method that classified products based on the Universal Product Codes achieved accuracy exceeding 90%.

Finally, there is a large and growing literature on machine learning methods for data editing and imputation. As one example, McClellan, Mitchell, Anderson and Zuvekas (2023) found that random forest and combinations of machine learning methods had greater accuracy than linear regression for imputing missing expenditure items in the U.S. Medical Expenditure Panel Survey. Lin and Tsai (2020) and Dagdoug, Goga and Haziza (2023a) reviewed and compared machine learning methods for imputing missing items from survey respondents.

Of course, machine learning methods cannot solve all problems and it is important to test their accuracy before implementation. For example, Ikudo, Lane, Staudt and Weinberg (2019) explored using random forest models to assign occupational classification from university human resources records, but found that job titles in human resource data are inconsistent across institutions and lack sufficient detail to allow accurate classification. Another important issue is data equity, ensuring that the algorithms do not systematically disadvantage some population subgroups, which will be discussed further in Section 6.

2.2 Survey design

Survey design articles in the first 25 years of *Survey Methodology* were mainly of three types: describing the design of a particular survey; implementing probability-proportional-to-size sampling; and allocating observations in stratified, two-phase, and multi-stage designs. Allocation continues to be a theme for survey design articles, but an increasing amount of research in the past 25 years has been devoted to topics such as anticipating nonsampling errors at the survey design stage, balanced sampling, and adaptive sampling.

Anticipating nonsampling errors at the survey design stage. Statistical organizations around the world have made great strides in improving sampling frame coverage, investigating measurement error, and reducing coding and data processing errors. The increased use of designed experiments allows survey researchers to study mode and question effects, within-household selection methods, nonresponse minimization techniques, variability among interviewers, and many other aspects of survey quality (Lavrakas, Traugott, Kennedy, Holbrook, de Leeuw and West, 2019). New technologies and data sources may be used to identify units that may be missing from sampling frames, or to identify subsets for oversampling (Datta, Ugarte and Resnick, 2020). For example, Hyman, Sartore and Young (2022) used webscraping to identify local food farms that might be missing from a list frame of farms, and Daas and van der Doef (2020) employed a text-based model to identify innovative companies by studying text on their websites.

Response rates in many countries have declined dramatically since 2000, however, and although nonresponse models may reduce bias, there is no guarantee that the bias is removed (see Section 3.2). Mercer, Lau and Kennedy (2018) found that the quality of the available auxiliary information affected bias reduction more than the particular nonresponse model used. Consequently, the development of sampling frames with rich auxiliary information about the population units allows improved nonresponse models as well as more efficient sampling designs. For surveys of persons, these frames may come from population registers or from efforts such as the U.S. Census Bureau Frames project, which merges information from geospatial, business, job, and demographic frames (National Academies of Sciences, Engineering, and Medicine, 2023, Chapter 4). Remote sensing technology provides detailed information for sampling frames in agriculture, forestry, and environmental studies. The switch from telephone surveys to address-based sampling also increases the auxiliary information available for sample design and nonresponse adjustments, because neighborhood characteristics are known for each sampled housing unit.

There has been increasing use of machine learning models to anticipate nonsampling errors when designing surveys (Buskirk and Kirchner, 2020). For example, Oberski and DeCastellarnau (2021) used random forest to predict measurement error variance of survey questions from key characteristics coded for each question. Kern, Weiß and Kolb (2021) used machine learning models to predict nonresponse in future waves of a panel survey, allowing likely nonrespondents to be targeted in advance.

Balanced sampling. A simple random sample is guaranteed to be representative in expectation but a particular sample, especially if small, can produce statistics for some characteristics that are far from the population values. Balanced sampling takes advantage of auxiliary information $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iJ})^T$ present in the sampling frame for each population unit i, and restricts the set of samples that can be chosen to those for which the estimated population totals for those variables, $\hat{\mathbf{X}}$, equal the known population totals $\mathbf{X} = \sum_{i=1}^{N} \mathbf{x}_i$. Stratified sampling is a special case in which $\mathbf{x}$ is the vector of stratum indicators. It may not be feasible to achieve a design that stratifies on the cross-classification of sex, age, race, region, education, income, and other variables, but balanced sampling allows selection of a sample that matches the moments or category proportions of each variable separately.

Balanced sampling allows for much more flexibility than stratification when selecting a sample from a frame with rich auxiliary information, which is why it is increasingly being employed in countries with population registers. It can be thought of as a way to pre-calibrate the survey on the balancing variables; generalized regression estimation has the same relation to balanced sampling as poststratification bears to stratified sampling.

A balanced sample may be chosen purposively (as in Royall and Herson, 1973) or probabilistically (as in stratified random sampling). A balanced sample chosen using probability sampling methods combines the efficiency advantages of balancing with the model-robustness afforded by probability sampling. Fuller (2009) proposed a rejective probability sampling procedure to ensure approximate balancing in the sense of rejecting a selected probability sample unless the weighted mean of an auxiliary vector $\mathbf{x}$ is within a specified distance of the corresponding population mean. The use of probabilistic balanced sampling has increased greatly since the development of the cube method for computing approximately balanced designs with fixed inclusion probabilities (Deville and Tillé, 2004; Leuenberger, Eustache, Jauslin and Tillé, 2022) and of algorithms for calculating variances of estimates from balanced samples (Deville and Tillé, 2005; Breidt and Chauvet, 2011; Kim and Wu, 2013; Tillé, 2020).

Falorsi and Righi (2008, 2015) employed a balanced sampling approach when small area estimates are desired across a number of different domain types and it is not feasible to use the cross-classification of the different partitions. The allocation is designed to ensure that the direct (or indirect, if needed) small area estimates achieve a predetermined mean squared error for each domain of interest. Chauvet, Deville and Haziza (2011) applied principles of balanced sampling to imputation by randomly sampling donors for hot deck imputation while satisfying a set of constraints.

Spatially balanced sampling designs are often desirable for ecology or forestry surveys. These force the sample to be spread out over the study area, thus reducing the spatial correlations among units in the sample.

Tillé (2020, Chapter 8) and Robertson and Price (2024) summarized methods for selecting spatially balanced probability samples, along with measures for evaluating the spatial balance. Grafström and Ringvall (2013) derived designs for achieving spatial balance as well as balance on auxiliary variables obtained via remote sensing.

Adaptive survey designs. In classical probability sampling, the design is fixed before data collection commences – the strata are determined, primary sampling units (psus) are selected according to the prespecified probabilities, and the data are collected according to the predetermined protocol. In adaptive survey designs, information from units measured early in the process is used to modify (adapt) later data collection efforts.

Some adaptive survey designs result in probability samples – for example, adaptive cluster sampling starts with a small probability sample of psus, and then modifies selection probabilities for subsequent psus according to the data already collected (see Thompson, 2017, for a review). When the initial sample is not chosen probabilistically, however, the final sample from an adaptive procedure is not a probability sample. For example, respondent-driven sampling has become popular in recent years as a method for collecting data from a rare or hard-to-identify population such as recent immigrants or minority groups (Heckathorn and Cameron, 2017; Gile, Beaudry, Handcock and Ott, 2018). Members of a convenience sample from the hard-to-identify population are asked to name other members of the population for inclusion of the sample, who in turn are asked to suggest other population members, and so on until the desired sample size is attained. Because the initial sample consists of easy-to-recruit population members, a respondent-driven sample is a convenience sample and strong modeling assumptions are needed to make inferences to the population.

Another form of adaptive design attempts to improve quality metrics by modifying data collection protocols for units that have not yet responded. Cases may be prioritized to increase the overall response rate, target low-propensity subpopulations, reduce the variance of nonresponse weighting adjustments, or control costs. Coffey, Damineni, Eltinge, Mathur, Varela and Zotti (2024) gave examples in which auxiliary information from external data sources may inform data collection decisions.

Groves and Heeringa (2006) coined the term “responsive survey design” for a procedure that divides the data collection period into phases and uses information accrued from earlier phases (for example, number of calls or observations noted by interviewers) to modify the data collection protocols for later phases. The idea of responsive design is not new – the two-phase approach for nonresponse follow-up of Hansen and Hurwitz (1946) is used for many surveys including the American Community Survey – but recent years have seen an increase in research that tailors survey recruitment to individual nonrespondents (Schouten, Peytchev and Wagner, 2018; Zhang and Wagner, 2024). Tourangeau, Brick, Lohr and Li (2017) wrote that response rate increases, bias reductions, and reductions in variability of estimated response propensities were modest in the applications of responsive designs they reviewed. However, most of those applications had low-information sampling frames and it is possible that responsive design methods may work better with rich frame information. Kaputa, Thompson and Beck (2019) stressed the importance of evaluating responsive design strategies through designed experiments.

3. Estimation

3.1 Weighting and use of auxiliary information

Most surveys use weighting and calibration to produce estimates of population characteristics. Weighting accounts for unequal selection probabilities in the sampling design and is also commonly used with estimated self-selection probabilities from nonresponse or in nonprobability samples. An advantage of weighting is that estimates are internally consistent with each other – for example, estimated population totals for partitioning subsets sum to the estimated total for the full population – because all estimates are calculated using the same set of weights. Calibration adjusts sample weights so that estimated population totals for auxiliary variables agree with the known population totals for those variables. Brick (2013) and Haziza and Beaumont (2017) reviewed research on weighting methods and Särndal (2007) reviewed the calibration approach.

For many surveys, control totals used for calibration come from other surveys and themselves have sampling and nonsampling errors. Dever and Valliant (2016) derived statistical properties of estimators that are calibrated to quantities derived from another survey, and they and Opsomer and Erciulescu (2021) proposed variance estimators that capture the variability from both sources. Haziza and Lesage (2016) concluded that a two-step approach to weighting adjustments – first adjusting weights according to estimated response propensities and then calibrating weights to population control totals – provides more robustness to assumed nonresponse mechanisms than a one-step approach with calibration alone. The two-step procedure removes bias if either the response propensity model or the calibration model is correct (see Section 3.2 for nonresponse models).

In some situations a large number of auxiliary variables could potentially be used for calibration. Using all of the variables in the calibration model could result in instability through inflated variances or weight spiking. One potential solution is to restrict the set of variables used in calibration so as to avoid multicollinearity. Cardot, Goga and Shehzad (2017) proposed reducing the dimension of the calibration variables through principal component analysis. Multicollinearity can also be reduced through use of ridge-regression-type penalized optimization criteria (Rao and Singh, 1997, 2009; Beaumont and Bocci, 2008; Goga, 2024).

Other researchers have employed machine learning methods to select variables and determine the form of the calibration adjustments. Survey practitioners have long used decision tree methods such as the chi-square automatic interaction detection (CHAID) algorithm (Kass, 1980) to select variables and form nonresponse adjustment cells (see, for example, Rizzo, Kalton and Brick, 1996). Since then, a number of researchers have taken advantage of other developments in machine learning to improve survey nonresponse adjustments. For example, Phipps and Toth (2012) built regression trees to study nonresponse patterns in the Occupational Employment Statistics survey. Lohr, Hsu and Montaquila (2015) explored the performance of various regression trees and random forest, which averages predictions from a large number of trees, for estimating response propensities. McConville and Toth (2019) established the asymptotic properties of poststratified estimators where the poststrata are chosen by a regression tree model. Opsomer and Riddles (2025) extended the CHAID approach to account for the survey design by incorporating a Rao-Scott correction into the splitting criterion.

Breidt and Opsomer (2017) reviewed model-assisted approaches for estimating population means and totals. A study variable y is predicted by a function $m(\mathbf{x})$ of auxiliary variables $\mathbf{x}$. The population total is then estimated by the difference estimator

$$\hat{Y}_d = \sum_{i=1}^{N} \hat{m}(\mathbf{x}_i) + \sum_{i \in \mathcal{S}} d_i \left[y_i - \hat{m}(\mathbf{x}_i) \right], \tag{3.1}$$

where $\mathcal{S}$ denotes the set of sampled units, $d_i = 1/\pi_i = 1/P(i \in \mathcal{S})$ is the design weight for unit i, and $\hat{m}$ estimates m. The estimator is asymptotically unbiased under regularity conditions because the second term corrects for potential bias in the prediction model. Wu and Sitter (2001), proposing the use of a general function m for calibration, employed an estimator similar to (3.1) but with $\left[\sum_{i=1}^{N} \hat{m}(\mathbf{x}_i) - \sum_{i \in \mathcal{S}} d_i \hat{m}(\mathbf{x}_i) \right]$ multiplied by a regression coefficient.

Montanari and Ranalli (2005) and Breidt and Opsomer (2017) presented statistical properties of the estimator in (3.1) with local polynomial, spline, neural network, and other nonparametric models used for $m(\mathbf{x})$. McConville, Breidt, Lee and Moisen (2017) and Chen, Valliant and Elliott (2018) selected regression variables using the lasso, and Dagdoug, Goga and Haziza (2023b) derived asymptotic properties when the model is chosen using random forest. Lundy and Rao (2022), in a simulation study of machine learning methods for calibration, found that regression trees and lasso had similar performance in the case of categorical covariates leading to main effects and interactions, and both methods automatically handled multicollinearity among the auxiliary variables. Much of this research on model-assisted estimation has been done in the full response setting, but the results have also been applied to calibration for treating nonresponse (see Section 3.2).

Although these more general functions can result in a smaller variance for the estimated population total $\hat{Y}_d$, there are some advantages to using a linear regression function for m. Because $\hat{m}(\mathbf{x}) = \mathbf{x}^T \hat{\mathbf{B}} = \mathbf{x}^T \left(\sum_{k \in \mathcal{S}} d_k \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \sum_{k \in \mathcal{S}} d_k \mathbf{x}_k y_k$ is a linear function of y, the generalized regression estimator can be calculated as $\sum_{i \in \mathcal{S}} d_i g_i y_i$, where $g_i = \left[1 + (\mathbf{X} - \hat{\mathbf{X}})^T \left(\sum_{k \in \mathcal{S}} d_k \mathbf{x}_k \mathbf{x}_k^T \right)^{-1} \mathbf{x}_i \right]$. Calibration with a linear model achieves internal consistency by using the same set of adjusted weights, $w_i = d_i g_i$, for all estimates. Additionally, it is not necessary to know $\mathbf{x}_i$ for every population unit – $\mathbf{x}_i$ must be known for $i \in \mathcal{S}$ but the estimator depends on auxiliary information about unsampled units only through the population total $\mathbf{X}$.

When a nonlinear prediction function is used in (3.1), the calibration is conducted with respect to the population total of the predicted values, $\sum_{i=1}^{N} \hat{m}(\mathbf{x}_i)$, rather than with respect to $\mathbf{X}$. For most functions m, $\mathbf{x}_i$ must be known for all sampled and unsampled units to be able to calculate $\sum_{i=1}^{N} \hat{m}(\mathbf{x}_i)$, in contrast to linear regression. Use of nonlinear functions may also result in multiple sets of weights when models are fit to different study variables. For example, with a regression tree function for m, the estimator in (3.1) can be expressed in the form $\sum_{i \in \mathcal{S}} w_i y_i$ for a set of weights, but the weight w_i depends on the number of population units that fall into the terminal node containing unit i and will be a different value for a different study variable. One can obtain consistency across estimates by fitting the model to one study variable and using those weights for all estimates, but that might reduce efficiency for other study variables. Montanari and Ranalli (2009) proposed obtaining a single set of weights w_i that satisfy the multiple calibration constraints $\sum_{i \in \mathcal{S}} w_i \mathbf{u}_i = \sum_{i=1}^{N} \mathbf{u}_i$, where the vector $\mathbf{u}_i = (\hat{m}_1(\mathbf{x}_i), \ldots, \hat{m}_P(\mathbf{x}_i), \mathbf{z}_i^T)^T$ consists of separate model predictions for each of P key study variables as well as a subset $\mathbf{z}$ of the auxiliary variables $\mathbf{x}$. They suggested using ridge-type regression estimators to reduce the variability of the weights and meet range restrictions.

3.2 Nonresponse models

As response rates for household surveys have decreased, research on weighting, calibration, and imputation models for missing survey data has intensified. Many of the methods can be applied to nonprobability samples as well as to probability samples having nonresponse.

Weighting adjustment methods for nonresponse attempt to estimate the probability ϕ_i that unit i responds to the survey, called the response propensity. If the initial sample is chosen probabilistically, then the population total is estimated by $\hat{Y}_\phi = \sum_{i \in \mathcal{R}} y_i \big/ \left(\pi_i \hat{\phi}_i\right)$, where $\mathcal{R}$ is the set of respondents to the survey and $\hat{\phi}_i$ is an estimator of ϕ_i. Kim and Kim (2007) derived the expected value and variance of the estimated population total when $\hat{\phi}_i$ is estimated using a logistic regression model. The estimator $\hat{Y}_\phi$ is approximately unbiased when all response propensities exceed a fixed constant $k > 0$ and are accurately modeled by the assumed function. These are strong assumptions, as many surveys have some unshakable nonrespondents whose response propensities may be considered to be 0, and in practice the response mechanism may have a different functional form or may depend on unmeasured variables.

Recent developments in nonresponse modeling and imputation are too numerous to describe in this paper and we refer the reader to Kim and Shao (2022) for a discussion of multiple and fractional imputation, nonignorable missing data, and clustered data. Two developments, however, relate to other themes in this paper. The first, described in Sections 2.1 and 3.1, is the increased use of machine learning models for imputation and nonresponse weight adjustments. Ferri-García and Rueda (2020) reviewed machine learning methods for estimating response propensities in online surveys.

The second development, which is related to the model-assisted estimation discussion in the previous section, is research on double robustness for nonresponse models. In such an approach, as introduced by

Kott (1994) for survey data, two models are developed. The first model predicts the propensity of responding to the survey; the second (imputation) model predicts y for missing units from auxiliary information as $\hat{m}(\mathbf{x})$. A doubly robust estimator of the population total has a form analogous to (3.1):

$$\hat{Y}_{\text{DR}} = \sum_{i \in \mathcal{S}} \frac{\hat{m}(\mathbf{x}_i)}{\pi_i} + \sum_{i \in \mathcal{R}} \frac{y_i - \hat{m}(\mathbf{x}_i)}{\pi_i \hat{\phi}_i} \tag{3.2}$$

If $\mathbf{x}$ is known for every population unit, $\sum_{i=1}^{N} \hat{m}(\mathbf{x}_i)$ can be substituted for the first term in (3.2). The estimator is approximately unbiased if either the propensity model ϕ or the imputation model m is correctly specified, hence the name double robustness. If both models are misspecified, however, $\hat{Y}_{\text{DR}}$ can still have substantial bias; Kang and Schafer (2007, page 523) concluded that “in at least some settings, two wrong models are not better than one”. Chen and Haziza (2017, 2023) and Wu (2023) described recent work in doubly and multiply robust estimators in the presence of nonresponse.

3.3 Variance estimation

When the first issue of *Survey Methodology* was published in 1975, most variance estimates were calculated using formulas for the appropriate sampling design along with Taylor series methods. Random group methods, jackknife, and balanced repeated replication had been introduced for at least some survey designs before 1975 (Mahalanobis, 1939; Durbin, 1959; McCarthy, 1966) but lack of computational power and software limited their use in practice. By the time Rao (1999) wrote his review, rigorous asymptotic theory had been developed for replication variance estimation methods (see, for example, Krewski and Rao, 1981; Rao and Wu, 1985) and multiple statistical agencies were using jackknife or balanced repeated replication for variance estimation.

Replication variance estimation is now standard for many surveys. This has been facilitated by the widespread availability of software for constructing replicate weights and calculating variances. Replication methods have many advantages, including the ability to compute variances for many types of statistics without having to derive separate variance formulas. They also incorporate the effects of weighting adjustments on with-replacement variance estimates (which also estimate the without-replacement variances when first-stage sampling fractions are small), because the same nonresponse adjustment steps are performed on the full weights and on each set of replicate weights. The past 25 years have seen increased use of the rescaling bootstrap (Rao and Wu, 1988; Rao, Wu and Yue, 1992) because of its flexibility in terms of sample sizes and number of bootstrap estimates; Mashreghi, Haziza and Léger (2016) reviewed recent developments on the bootstrap with probability samples and Beaumont and Émond (2022) proposed a bootstrap method that may be used with small first-stage sampling fractions.

Replication methods such as jackknife and bootstrap are typically used when all point estimates are calculated using a single set of weights. The replicate weights are derived from that single set according to the particular replication method. When estimates depend on multiple sets of weights, or have sources of randomness additional to the survey design and nonresponse adjustments, standard replication methods need modifications to capture all of the sources of variability. Some model-assisted or doubly robust estimates rely on predictions of y as well as a final weight vector. Wu (2022) summarized variance estimation methods that can be used for doubly robust estimates.

Following the initial work of Kott and Stukel (1997), there have been a number of developments in using replication methods to calculate variances for two-phase samples. The procedures described in Kim and Yu (2011) involve creating separate sets of jackknife replicates for the phase-1 and phase-2 weight vectors. Opsomer, Breidt, White and Li (2016) proposed variance estimation through successive difference replication in two-phase samples, and Beaumont, Béliveau and Haziza (2015) related a proposed simplified two-phase variance estimator to replication methods.

Another situation in which estimation relies on more than just a single weight vector occurs when fitting multilevel models to survey data. In the composite likelihood method proposed by Yi, Rao and Li (2016), estimates of variance components depend on the first-stage psu inclusion probabilities and on the joint inclusion probabilities for pairs of elements within psus. Lumley and Huang (2024) developed a bootstrap method that estimates the with-replacement variance using replicate weights for pairs of observations.

In addition to the numerous developments in replication variance estimation, there have also been advances in linearization methods. The traditional approach expresses the parameter of interest as a function of a vector of population totals $\mathbf{Y}$ and approximates the variance using derivatives of the function, which may be defined explicitly or implicity, with respect to the components of $\mathbf{Y}$ (Binder, 1983). If $\theta = g(\mathbf{Y})$ and $\hat{\theta} = g(\hat{\mathbf{Y}})$, then one can define a linearized variable

$$u_i = \left.\left(\frac{\partial g(\mathbf{a})}{\partial \mathbf{a}}\right)^T\right|_{\mathbf{a}=\mathbf{Y}} \mathbf{y}_i, \tag{3.3}$$

and, under regularity conditions, $V(\hat{\theta})$ is approximated by the variance of $\hat{U} = \sum_{i \in \mathcal{S}} d_i u_i$, the survey estimate of $U = \sum_{i=1}^N u_i$. Binder (1996) recommended a variant of (3.3) in which the derivative is evaluated at $\hat{\mathbf{Y}}$. Demnati and Rao (2004) noted that $\hat{\theta}$ can be alternatively expressed as a function $f(\mathbf{d})$ of the vector of design weights and wrote $\hat{\theta} \approx \sum_{i \in \mathcal{S}} d_i z_i$, where $z_i = \left.\left(\frac{\partial f(\mathbf{b})}{\partial b_k}\right)\right|_{\mathbf{b}=\mathbf{1}}$. Graf (2011) linearized the estimator of θ by taking derivatives with respect to the sample inclusion indicators and evaluating the derivatives at the inclusion probabilities.

4. Small area estimation

Sample surveys are typically designed to provide reliable estimates of totals and means for the target population as well as for subpopulations (domains) with sufficiently large samples. Those estimates use domain-specific data only and are called direct estimators. However, the demand for reliable statistics for local areas (called small areas) has grown greatly over the past 30 years or so. Direct estimators for small areas are either unreliable because of small sample sizes within areas or not feasible due to zero area-specific sample sizes. It is therefore necessary to make use of sample data in related areas through implicit or explicit linking models. When the first issue of *Survey Methodology* appeared in June 1975, only a few applications of small area methods based on implicit models had been done. Most of those methods relied on synthetic estimators, such as assuming that the area means within a poststratum are approximately equal to the stratum mean (National Center for Health Statistics, 1968). Synthetic estimation is based on strong implicit models while explicit models take account of errors in modeling and permit model checking and validation.

Explicit area level models, starting with the groundbreaking work of Fay and Herriot (1979), led to many extensions of the Fay-Herriot (FH) area level model and to a variety of applications for estimating area means. Unit level models also received a lot of attention, starting with the work of Battese, Harter and Fuller (1988). The book by Rao (2003) provided the underlying theory and methods for area level and unit level models. Due to many useful extensions of the basic area level and unit level models reported in the literature after 2003, a second edition of the book, Rao and Molina (2015), provided a comprehensive account of the developments to 2014. Many new extensions have appeared in the literature since 2015. A recent book by Morales, Esteban, Pérez and Hobza (2021) provided details of mathematical developments in small area estimation (SAE) with R code for applications. Molina and Rao (2023), Sugasawa and Kubokawa (2020), and Ghosh (2020) traced the history of SAE in the past 50 years and covered some current topics related to model-based SAE.

The basic area level FH model for estimating area means consists of a sampling model on the direct estimators and associated model on the area means linking the area means θ_i through area level covariates $\mathbf{z}_i$. Direct estimators are obtained from area-specific unit level data taking account of the survey design. The sampling model is given by $\hat{\theta}_i = \theta_i + e_i, i = 1, \ldots, m$ where m is the number of areas and the sampling errors e_i are assumed to be independent with mean 0 and known variance ψ_i. For simplicity, we assume that all the areas in the population are sampled. The linking model is given by $\theta_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i$, where $\boldsymbol{\beta}$ is the vector of model parameters, and v_i is a random area effect that has zero mean and common variance σ_v^2, and the area effects are assumed to be independent. Noting that the two models match, a combined model, called the FH model, is given by $\hat{\theta}_i = \mathbf{z}_i^T \boldsymbol{\beta} + v_i + e_i$, which is a special case of a linear mixed model. An empirical best linear unbiased predictor (EBLUP) of θ_i, without distributional assumptions, can be obtained. The EBLUP can be expressed as a weighted combination of the direct estimator and a regression synthetic estimator. More weight is given to the synthetic estimator when the sampling variance is large relative to the variance of the random area effect. Alternative estimators include empirical best (EB) and hierarchical Bayes (HB) predictors, based on parametric assumptions for EB and HB and in addition prior distributions on model parameters $\boldsymbol{\beta}$ and σ_v^2 for HB. EB and HB methods can handle more general cases and the HB method provides exact inferences unlike EBLUP and EB which require appealing to asymptotic theories for estimating the mean squared prediction error (MSPE) of the estimators. Rao (1999) described linearization methods of estimating MSPE and subsequently jackknife and parametric bootstrap methods of estimating MSPE have been proposed, see Rao and Molina (2015), Section 6.2

Area level models have the advantage of taking survey design into account through the sampling model and using area level covariates in the linking model. Area level covariates are more readily available than unit level covariates. Because of those advantages, the basic FH model has been extended in several different directions as well as addressing several practical issues. Practical issues addressed include: (1) Area level covariates subject to sampling or measurement errors. (2) Unknown sampling variances. (3) Linking model incorrectly specified. (4) Benchmarking area level estimators to a reliable direct estimator at an aggregate level. (5) Uncertain area effects involving the absence of area effect beyond the covariates for several areas; see Rao and Molina (2015), Section 6.4. Ghosh, Ghosh, Maples and Tang (2022) used global-local priors for HB estimation to address (5).

Extensions of the basic FH area level models include: (1) Use of cross sectional and time series data to borrow strength cross sectionally as well as over time. (2) Spatial models involving spatial correlation among the random area effects. (3) Bivariate FH models. (4) Estimating area proportions for a binary variable or polytomous variable using a logistic linear mixed model formulation. (5) Subarea or two-fold models such as modeling counties (subareas) within states (areas): see Rao and Molina (2015), Chapter 8. Spatial models (extension 2) are of interest when the area means of geographically close areas exhibit spatial variation and the available area level covariates are not adequate for explaining the spatial variation. Moreover, if some areas are not sampled, then the traditional FH model uses synthetic estimates based only on the covariates associated with those areas. On the other hand, spatial models can lead to improved estimates by modifying the synthetic estimates using the spatial model. Chung and Datta (2020) used Bayesian spatial models and the HB approach to estimate the means of sampled and non-sampled areas.

Turning to unit level models, again we assume all the areas are sampled, for simplicity. The unit level population data are given by $\{(y_{ij}, \mathbf{x}_{ij}), j = 1, \ldots, N_i, i = 1, \ldots, m\}$ and the population model is a nested error regression model $y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + v_i + e_{ij}, j = 1, \ldots, N_i, i = 1, \ldots, m.$ Here i and j denote area and unit within the area, y_{ij} and $\mathbf{x}_{ij}$ denote a variable of interest and associated covariate vector, v_i is an area random effect assumed to be independent of the unit error e_{ij}, and both are independent and identically distributed with zero mean and common variances σ_v^2 and σ_e^2, respectively. Area sample sizes are denoted by n_i and the sample is assumed to obey the population model, which implies non-informative sampling or absence of sample selection bias within areas. Under the above set up, EBLUP estimators of area means can be obtained without normality assumptions as weighted combinations of a sample regression estimator and a regression synthetic estimator. We assume there are no errors in linking y_{ij} to the associated covariate vector $\mathbf{x}_{ij}$. Extensive work has been reported on estimating the MSPE of the EBLUP; see Rao and Molina (2015), Chapter 7.

Fay (2018) argued that when unit level data are available, the weighted sample mean in the area level model should be replaced by the sample regression estimator and demonstrated that the resulting EBLUP for the area level model is comparable in efficiency to the EBLUP under the unit level model. In contrast, the use of weighted sums as the direct estimator in the area level model can lead to considerable loss of efficiency compared to the EBLUP under the unit level model (Hidiroglou and You, 2016). Practical issues addressed include robust estimation of area means in the presence of outliers, pseudo-EBLUP estimation based on the sample design weights and model misspecification; see Rao and Molina (2015), Section 7.6. Sun, Berg and Zhu (2024) reviewed work on multivariate unit-level models and proposed an extension in which study variables can be continuous or discrete. Pfeffermann and Sverchkov (2007) and Verret, Rao and Hidiroglou (2015) proposed models to handle informative sampling within areas. Verret et al. (2015) also showed that the pseudo-EBLUP handles informative sampling quite well.

Unit level models permit the estimation of complex parameters such as small area poverty indicators used to construct poverty maps. The World Bank uses poverty rate, poverty gap, and poverty severity as indicators of poverty. EB and HB estimation of poverty indicators under unit level models have been studied; see Guadarrama, Molina and Rao (2016) and Rao and Molina (2015), Sections 9.4 and 10.7.3.

Both area and unit level models have been extended to use functions other than linear regression for the linking model. Opsomer, Claeskens, Ranalli, Kauermann and Breidt (2008) and Rao, Sinha and Dumitrescu (2014) used penalized spline regression to allow the data to determine the shape of the curve in unit level models. Lohr and Mendez (2009), Krennmair and Schmid (2022), and Beaumont, Bocci, Bosa and Sombo (2024) have studied the use of random forest in small area models.

5. Data integration and nonprobability samples

Several of the topics mentioned in earlier sections involve integrating data from multiple sources. For example, calibration and small area estimation use data from external sources for control totals or as model inputs. Some imputation methods rely on external data sources for values to be imputed. In recent years, as survey costs have risen, there has been an increase in research on integrating administrative datasets, sensor data, social network data, and other non-survey data sources to produce statistics. Many statistical offices are moving toward multi-source statistics in order to take advantage of datasets collected for non-statistical purposes (De Waal, van Delden and Scholtus, 2020). Research on using multiple data sources to produce statistics is reviewed in Lohr and Raghunathan (2017), Thompson (2019), Zhang and Chambers (2019), Beaumont (2020), Yang and Kim (2020), Rao (2021), and Cabrera-Álvarez (2022).

Many of *Survey Methodology's* early papers on data integration concerned linking records from multiple datasets. Record linkage continues to be a primary method for integrating data – it can be used to augment the set of variables in a dataset, impute missing values by linking a survey nonrespondent to administrative records, or form a concatenated dataset of unique entities by identifying records in the multiple sources that belong to the same entity. Asher, Resnick, Brite, Brackbill and Cone (2020), Winkler (2021), and Binette and Steorts (2022) reviewed recent research in record linkage methods; Han and Lahiri (2019) studied statistical analysis with linked data; and Reiter (2021) summarized approaches that can account for uncertainty in linkages.

Accurate record linkage requires all data sources to contain detailed information that can be used to identify records belonging to the same entity. In this section, we concentrate on other methods that have been proposed for combining information from surveys with other sources and, in particular, on recent developments that do not require linking individual units.

There are numerous advantages for using non-survey data sources to augment (or sometimes replace) data from surveys. Using a dataset (for example, administrative records) that is already being collected for other purposes can reduce survey costs and burden. Alternative data sources can also provide information on population members who may be missing from a probability sample, such as determined nonrespondents or persons out of scope for the survey. A large administrative dataset may be able to provide detailed information for small subpopulations that is not available from a survey.

The major concern in using nonprobability samples for inference to a population is potential bias, from:

- Self-selection. Administrative records such as income tax records may have a well-defined population, but other types of surveys, such as opt-in online polls or social media data, are subject to self selection. The probability that a population unit participates in the sample is typically unknown.
- Undercoverage. Tax records lack information on persons not required to file income tax returns, electronic health records lack data on persons who do not use health care systems, and credit card databases lack information on cash transactions. For opt-in surveys, there are some persons in the population who will not have the opportunity to participate, or who will not participate under any circumstances – these people may be considered to have a participation probability of 0.
- Measurement error. The concepts measured in administrative or other non-survey data may not be exactly what the statistician wants to study. Measurement error may arise from question wording effects, recording errors, or even deliberate misreporting in some sources (as explored in Kennedy et al., 2024).

Keiding and Louis (2018) discussed these issues in the context of epidemiological studies. Of course bias from nonresponse, undercoverage, and measurement errors is also a concern for probability samples, and statistical methods for bias reduction with multiple data sources have built on those that have been developed for probability samples.

The problem of undercoverage can sometimes be addressed by employing two or more data sources in a multiple-frame approach. Undercoverage from one source can then be partially compensated by the other source. Combining the data, however, requires knowledge about the overlap in the two universes represented by the samples – that is, although analysts do not need the detailed identification information required for record linkage they must know whether a unit sampled from one data source is also in the universe for the other data source. Lohr (2021) looked at extensions of multiple frame approaches that might be considered for merging probability and nonprobability samples.

When the multiple sources are presumed to come from the same population or the overlap is unknown, hierarchical models can be used to combine their information. Lohr and Raghunathan (2017) reviewed hierarchical model integration methods through 2016. In many of these approaches, which are related to statistical methods used for meta-analysis, random effects terms model the heterogeneity across data sources. A population (or, in some cases, domain) mean from data source j, $\hat{\mu}_j$, is assumed to follow a $N(\mu + \Delta_j, \tau_j)$ distribution; if source j is a high-quality probability survey, the bias Δ_j is assumed to be 0. Cahoy and Sedransk (2023) allowed the survey means $\hat{\mu}_j$ to be clustered, with the form of the clustering estimated from the data, so that the posterior distribution of μ relies more heavily on surveys in the same cluster as the high-quality probability survey. Wiśniowski, Sakshaug, Perez Ruiz and Blom (2020) proposed a Bayesian method for reducing the variances of estimates from a small probability sample by deriving prior distributions for model parameters from a larger, parallel nonprobability sample. Nandram and Rao (2024) obtained small area estimates through a Bayesian model that integrated a probability sample with a nonprobability sample.

The multiple-frame and hierarchical modeling approaches discussed above require the study variable y to be measured in all of the data sources. In many situations of interest, however, no probability sample is available that measures y. Wu (2022) and Valliant (2024) reviewed two major approaches that have been taken to obtain estimates of the population total Y for a variable y measured only in a nonprobability (convenience) sample $\mathcal{S}_C$. Both approaches suppose the existence of a high-quality probability sample or census $\mathcal{S}_P$ that can be used to adjust for the selection bias. Although y is not measured in $\mathcal{S}_P$, both $\mathcal{S}_C$ and $\mathcal{S}_P$ measure a common set of auxiliary variables $\mathbf{x}$.

The first approach, similar to inverse propensity weighting for nonresponse in probability samples, expresses the convenience sample participation probability $\xi_i = P(i \in \mathcal{S}_C \mid \mathbf{x}_i, y_i)$ as a function of $\mathbf{x}_i$ and y_i and estimates Y by $\hat{Y}_1 = \sum_{i \in \mathcal{S}_C} y_i / \hat{\xi}_i$, where $\hat{\xi}_i$ is an estimator of ξ_i. The second approach predicts the study variable using a model $y = m(\mathbf{x})$ and estimates Y by $\hat{Y}_2 = \sum_{i \in \mathcal{S}_P} d_i \hat{m}(\mathbf{x}_i)$, where d_i is the weight for unit i associated with the probability sample. Wu (2022) also recommended a doubly robust approach that combines the inverse-propensity-weighted and imputation estimates similarly to (3.2): $\hat{Y}_3 = \sum_{i \in \mathcal{S}_C} \left[y_i - \hat{m}(\mathbf{x}_i) \right] / \hat{\xi}_i + \hat{Y}_2$.

Several methods have been proposed for estimating the convenience sample participation probabilities ξ_i (all of these methods have strong assumptions that are discussed below). Calibration or poststratification can be used if population totals for $\mathbf{x}$ are known. In many situations, however, population totals are known only for a limited set of variables (for example, age/race/sex groups) and a probability survey may be able to provide a richer set of auxiliary variables for reducing bias. Chen, Li and Wu (2020) proposed a pseudo-likelihood method for estimating ξ_i using $\mathcal{S}_P$. Savitsky, Williams, Gershunskaya and Beresovsky (2023) estimated ξ_i through stacking $\mathcal{S}_P$ and $\mathcal{S}_C$ and estimating the probability that a unit in the stacked sample comes from the convenience sample; the likelihood function in Wang, Valliant and Li (2021) is a special case when $\mathcal{S}_P$ is a census of the population.

As in Section 3.1, the model m used for imputation in the second approach may be parametric (for example, linear or logistic regression), nonparametric, or based on machine learning methods. Kaputa, Morris and Holan (2024) additionally incorporated spatial dependence into their HB model for estimating retail sales.

In contrast to inference from full-response probability samples, which depends primarily on the sampling design, statistical inferences from nonprobability samples depend upon strong – often unverifiable – assumptions. For approach 1, it is typically assumed that the sampling mechanism for $\mathcal{S}_C$ is ignorable (i.e., $P[i \in \mathcal{S}_C \mid \mathbf{x}_i, y_i] = P[i \in \mathcal{S}_C \mid \mathbf{x}_i]$) and that there is no undercoverage (i.e., all $\xi_i > 0$). For approach 2, it is typically assumed that the model m for predicting y is specified correctly and that the predicted values $\hat{m}(\mathbf{x}_i)$, obtained by fitting the model to $\mathcal{S}_C$, provide unbiased imputations for units in $\mathcal{S}_P$. For both approaches, it is assumed that the $\mathbf{x}$ variables are measured the same way on $\mathcal{S}_C$ as on $\mathcal{S}_P$ (for example, the same questions are used and there are no sources of differential measurement error). Standard errors or Bayesian credible intervals for estimates, discussed in Dever and Liao (2023), Wu (2022), and Savitsky et al. (2023), underestimate the uncertainty if the data integration approach does not remove all of the bias.

6. Discussion

Survey research has always been driven by information needs, data availability, and technological and methodological developments. Each of these aspects has changed since the review in Rao (1999). Accelerating demand for faster statistical information on increasingly fine subdivisions of the population has spurred research in small area estimation and data integration. Response rates for many household and establishment probability surveys have declined, while other types of data (for example, administrative, private-sector, satellite, and sensor data, as well as convenience samples) have become more plentiful but often have unknown quality and population coverage. Today's computers enable calculations and modeling approaches that were unthinkable in 1999. Technological advances have also affected how survey data are collected and processed; many surveys that would have been conducted face-to-face or by telephone in the past are now conducted over the internet, and machine learning methods are used to edit files and impute missing data.

We anticipate that trends in using computationally intensive models (as from machine learning) and relying on multiple data sources will continue in the near future. With decreasing response rates and increasing availability of other information, Beaumont (2020) asked whether probability surveys will disappear for the production of official statistics. We agree with his conclusion that, while alternative data sources can provide valuable information, “gold standard” data sources such as high-quality censuses and probability surveys are still essential for making inferences to the population. Almost all of the estimation methods reviewed in this paper on population or domain inference presuppose the existence of at least one data source that produces unbiased estimates for at least some variables. Continued research on improving quality of probability surveys – with innovations such as those described in Section 2 – will help ensure the existence of these representative data sources.

Much of the research on using multiple data sources has focused on using data sets that have already been collected (called “organic” data by Groves, 2011) together with existing probability surveys. We see a need for more research on design of data systems that include organic data, sources such as satellite and sensor data, and probability surveys. One option, discussed in Lohr and Raghunathan (2017), is to design probability surveys to supplement information available from other sources. For example, if an administrative dataset provides information on part of a population (for example, taxpayers), a supplementary survey might focus on obtaining information for the parts of the population missing from the administrative dataset. An optimal survey design in this framework would tailor the survey around the existing information to obtain estimates with low MSE for key estimates.

Low MSE, however, is not the only concern for an integrated data system. Other quality aspects such as granularity and timeliness need to be considered, and the system needs to be able to detect, and be robust to, changes in the organic data sources. For example, what happens if a data source changes or becomes unavailable? An administrative dataset might be discontinued if the program it supports is cancelled, a private data broker might stop collecting data or raise prices, or a large health insurance or credit card company might stop sharing data. It may be desirable to design the data system to have redundancy of information. Such redundancy would also allow better assessment of accuracy and measurement error in each source, and allow data sources to be monitored for changes in quality. More research is needed on how to take advantage of multiple data sources to evaluate assumptions and the quality of estimates (De Broe, Struijs, Daas, van Delden, Burger, van den Brakel, ten Bosch, Zeelenberg and Ypma, 2021). More research is also needed on confidentiality protection for blended data, as the additional information from the multiple sources may magnify disclosure risks (National Academies of Sciences, Engineering, and Medicine, 2024).

Data equity is an important consideration as machine learning algorithms and organic data become more widely used. The National Academies of Sciences, Engineering, and Medicine (2023) described data equity as “promoting the collection and use of data in which all populations, and especially those that have been historically underrepresented or misrepresented in the data record, are visible and accurately portrayed” (page 3). Nonresponse adjustments, imputation, model-assisted estimation, small area estimation, and data integration methods all rely on predictions from models, and the accuracy of these predictions depends on the quality of the data used to fit or train the model. In particular, predictions can be poor for subpopulations that are poorly represented in the training data because of selection bias or small sample size, and this can lead to inaccurate estimates for those subpopulations.

Careful attention to data equity issues is needed for the data integration methods described in Section 5. Some subpopulations may be missing from all data sources, or may have lower quality information available for data linkage or for identifying frame membership in a multiple-frame approach. If $\mathcal{S}_C$ has a large number of data records from a subpopulation that has only a handful of members in $\mathcal{S}_P$, estimates of the nonprobability sample participation probability ξ_i may be inaccurate. For approach 2, if the dataset $\mathcal{S}_C$ used to fit the model m contains few members of a population subgroup, then the model predictions applied to that subgroup may be suspect. Evaluating the quality of predictions is especially important for newer prediction methods such as those using machine learning. Erman, Rancourt, Beaucage and Loranger (2022) discussed equitable and responsible use of machine learning methods, and Statistics Canada's Framework for Responsible Machine Learning Processes called for assessments of training data quality, inference validity, and algorithm soundness (Bosa, 2022).

Finally, all of the methods discussed in this paper will benefit from continued research on the quality of data sources and on providing accurate measures of uncertainty. Kalton (2019, page S27) wrote: "Unless easily communicated quality metrics are developed and actively promoted for all types of sample designs, users may simply resort to the least expensive and fastest method without considering the quality of the estimates produced." Standard errors from probability sampling theory, which are accurate when the survey has full response and no measurement error, may substantially understate uncertainty about estimates when applied to low-response-rate probability samples or to nonprobability samples. Multiple data sources – and multiple methodologies for combining data – may allow survey statisticians to better understand non-sampling errors and incorporate these into measures of uncertainty.

The discipline of survey sampling has come a long way since the first issue of *Survey Methodology* in 1975, and many of the important developments have appeared in this journal. We look forward to another 50 years of innovation in its pages.

Acknowledgements

J.N.K. Rao's research was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to the editor and associate editor for their constructive comments and suggestions.

References

Asher, J., Resnick, D., Brite, J., Brackbill, R. and Cone, J. (2020). An introduction to probabilistic record linkage with a focus on linkage processing for WTC registries. *International Journal of Environmental Research and Public Health*, 17(6937), 1-16.

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83(401), 28-36.

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., Béliveau, A. and Haziza, D. (2015). Clarifying some aspects of variance estimation in two-phase sampling. *Journal of Survey Statistics and Methodology*, 3(4), 524-542.

Beaumont, J.-F., and Bocci, C. (2008). Another look at ridge calibration. *Metron*, 66(1), 5-20.

Beaumont, J.-F., Bocci, C., Bosa, K. and Sombo, S. (2024). The use of random forests in small area estimation. Paper presented on June 20, 2024 at the International Conference on Establishment Statistics, Glasgow.

Beaumont, J.-F., and Émond, N. (2022). A bootstrap variance estimation method for multistage sampling and two-phase sampling when Poisson sampling is used at the second phase. *Stats*, 5(2), 339-357.

Beck, M., Dumpert, F. and Feuerhake, J. (2018). Machine learning in official statistics. *arXiv preprint*. https://arxiv.org/abs/1812.10422.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review*, 51(3), 279-292.

Binder, D.A. (1996). Linearization methods for single phase and two-phase samples: A cookbook approach. *Survey Methodology*, 22(1), 17-22. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1996001/article/14389-eng.pdf.

Binette, O., and Steorts, R.C. (2022). (Almost) all of entity resolution. *Science Advances*, 8(12), eabi8021.

Bosa, K. (2022). Responsible use of machine learning at Statistics Canada. https://www.statcan.gc.ca/en/data-science/network/machine-learning.

Breidt, F.J., and Chauvet, G. (2011). Improved variance estimation for balanced samples drawn via the cube method. *Journal of Statistical Planning and Inference*, 141(1), 479-487.

Breidt, F.J., and Opsomer, J.D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science*, 32(2), 190-205.

Brick, J.M. (2011). The future of survey sampling. *Public Opinion Quarterly*, 75(5), 872-888.

Brick, J.M. (2013). Unit nonresponse and weighting adjustments: A critical review. *Journal of Official Statistics*, 29(3), 329-353.

Buskirk, T.D., and Kirchner, A. (2020). Why machines matter for survey and social science researchers: Exploring applications of machine learning methods for design, data collection, and analysis. In *Big Data Meets Survey Science: A Collection of Innovative Methods*, (Eds., C.A. Hill, P.P. Biemer, T.D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov and L.E. Lyberg), 9-62. Hoboken, NJ: Wiley.

Cabrera-Álvarez, P. (2022). Survey research in times of big data. *EMPIRIA: Revista de Metodología de las Ciencias Sociales*, 53, 31-51.

Cahoy, D., and Sedransk, J. (2023). Combining data from surveys and related sources. *Survey Methodology*, 49(1), 69-87. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2023001/article/00003-eng.pdf.

Cardot, H., Goga, C. and Shehzad, M.-A. (2017). Calibration and partial calibration on principal components when the number of auxiliary variables is large. *Statistica Sinica*, 27, 243-260.

Chauvet, G., Deville, J.-C. and Haziza, D. (2011). On balanced random imputation in surveys. *Biometrika*, 98(2), 459-471.

Chen, J.K.T., Valliant, R.L. and Elliott, M.R. (2018). Model-assisted calibration of non-probability sample survey data using adaptive LASSO. *Survey Methodology*, 44(1), 117-144. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2018001/article/54963-eng.pdf.

Chen, S., and Haziza, D. (2017). Multiply robust imputation procedures for the treatment of item nonresponse in surveys. *Biometrika*, 104(2), 439-453.

Chen, S., and Haziza, D. (2023). Doubly and multiply robust procedures for missing survey data. *The Survey Statistician*, 88, 75-85.

Chen, Y., Li, P. and Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, 115(532), 2011-2021.

Chung, H.C., and Datta, G.S. (2020). Bayesian hierarchical spatial models for small area estimation. Technical Report Statistics #2020-07, U.S. Census Bureau Center for Statistical Research & Methodology. https://www.census.gov/content/dam/Census/library/working-papers/2020/adrm/RRS2020-07.pdf.

Coffey, S.M., Damineni, J., Eltinge, J., Mathur, A., Varela, K. and Zotti, A. (2024). Some open questions on multiple-source extensions of adaptive-survey design concepts and methods. *Journal of Official Statistics*, 40(1), 16-37.

Couper, M.P. (2017). New developments in survey data collection. *Annual Review of Sociology*, 43, 121-145.

Daas, P.J., and van der Doef, S. (2020). Detecting innovative companies via their website. *Statistical Journal of the IAOS*, 36(4), 1239-1251.

Dagdoug, M., Goga, C. and Haziza, D. (2023a). Imputation procedures in surveys using nonparametric and machine learning methods: An empirical comparison. *Journal of Survey Statistics and Methodology*, 11(1), 141-188.

Dagdoug, M., Goga, C. and Haziza, D. (2023b). Model-assisted estimation through random forests in finite population sampling. *Journal of the American Statistical Association*, 118(542), 1234-1251.

Datta, A.R., Ugarte, G. and Resnick, D. (2020). Linking survey data with commercial or administrative data for data quality assessment. In *Big Data Meets Survey Science: A Collection of Innovative Methods*, (Eds., C.A. Hill, P.P. Biemer, T.D. Buskirk, L. Japec, A. Kirchner, S. Kolenikov and L.E. Lyberg), 99-129. Hoboken, NJ: Wiley.

De Broe, S., Struijs, P., Daas, P., van Delden, A., Burger, J., van den Brakel, J., ten Bosch, O., Zeelenberg, K. and Ypma, W. (2021). Updating the paradigm of official statistics: New quality criteria for integrating new data and methods in official statistics. *Statistical Journal of the IAOS*, 37(1), 343-360.

De Waal, T., van Delden, A. and Scholtus, S. (2020). Multi-source statistics: Basic situations and methods. *International Statistical Review*, 88(1), 203-228.

De Leeuw, E.D. (2018). Mixed-mode: Past, present, and future. *Survey Research Methods*, 12(2), 75-89.

Demnati, A., and Rao, J.N.K. (2004). Linearization variance estimators for survey data. *Survey Methodology*, 30(1), 17-26. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2004001/article/6991-eng.pdf.

Dever, J.A., and Liao, D. (2023). Variance estimation for probability and nonprobability establishment surveys: An overview. In *Advances in Business Statistics, Methods and Data Collection*, (Eds., G. Snijkers, M. Bavdaž, S. Bender, J. Jones, S. MacFeely, J.W. Sakshaug, K.J. Thompson and A. van Delden), 659-683. Hoboken, NJ: Wiley.

Dever, J.A., and Valliant, R. (2016). General regression estimation adjusted for undercoverage and estimated control totals. *Journal of Survey Statistics and Methodology*, 4(3), 289-318.

Deville, J.-C., and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91(4), 891-912.

Deville, J.-C., and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference*, 128, 569-591.

Durbin, J. (1959). A note on the application of Quenouille's method of bias reduction to the estimation of ratios. *Biometrika*, 46 (3/4), 477-480.

Erman, S., Rancourt, E., Beaucage, Y. and Loranger, A. (2022). The use of data science in a national statistical office. *Harvard Data Science Review*, 4(4), 1-18.

Falorsi, P.D., and Righi, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology*, 34(2), 223-234. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2008002/article/10763-eng.pdf.

Falorsi, P.D., and Righi, P. (2015). Generalized framework for defining the optimal inclusion probabilities of one-stage sampling designs for multivariate and multi-domain surveys. *Survey Methodology*, 41(1), 215-236. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2015001/article/14149-eng.pdf.

Fay, R.E. (2018). Further comparisons of unit- and area-level small area estimators. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 2057-2070, Alexandria, VA.

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(366a), 269-277.

Ferri-García, R., and Rueda, M.d.M. (2020). Propensity score adjustment using machine learning classification algorithms to control selection bias in online surveys. *PloS One*, 15(4), e0231500.

Fuller, W.A. (2009). Some design properties of a rejective sampling procedure. *Biometrika*, 96(4), 933-944.

Ghosh, M. (2020). Small area estimation: Its evolution in five decades. *Statistics in Transition*, 21(4), 1-22.

Ghosh, T., Ghosh, M., Maples, J.J. and Tang, X. (2022). Multivariate global-local priors for small area estimation. *Stats*, 5(3), 673-688.

Gile, K.J., Beaudry, I.S., Handcock, M.S. and Ott, M.Q. (2018). Methods for inference from respondent-driven sampling data. *Annual Review of Statistics and its Application*, 5, 65-93.

Goga, C. (2024). From traditional to modern machine learning estimation methods for survey sampling. *The Survey Statistician*, 90, 18-29.

Graf, M. (2011). Use of survey weights for the analysis of compositional data. In *Compositional Data Analysis: Theory and Applications*, (Eds., M. Graf, A. Buccianti and V. Pawlowsky-Glahn), 114-127. Hoboken, NJ: Wiley.

Grafström, A., and Ringvall, A.H. (2013). Improving forest field inventories by using remote sensing data in novel sampling designs. *Canadian Journal of Forest Research*, 43(11), 1015-1022.

Groves, R.M. (2011). Three eras of survey research. *Public Opinion Quarterly*, 75 (5), 861-871.

Groves, R.M., and Heeringa, S.G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A*, 169(3), 439-457.

Guadarrama, M., Molina, I. and Rao, J.N.K. (2016). A comparison of small area estimation methods for poverty mapping. *Statistics in Transition*, 17(1), 41-66.

Han, Y., and Lahiri, P. (2019). Statistical analysis with linked data. *International Statistical Review*, 87(S1), S139-S157.

Hansen, M.H., and Hurwitz, W.N. (1946). The problem of non-response in sample surveys. *Journal of the American Statistical Association*, 41(236), 517-529.

Harter, R., Battaglia, M.P. Buskirk, T.D., Dillman, D.A., English, N., Fahimi, M., Frankel, M.R., Kennel, T., McMichael, J.P., McPhee, C.B., Montaquila, J., Yancey, T. and Zukerberg, A.L. (2016). *Address-Based Sampling*. American Association of Public Opinion Research, Alexandria, VA. https://aapor.org/publications-resources/reports/.

Haziza, D., and Beaumont, J.-F. (2017). Construction of weights in surveys: A review. *Statistical Science*, 32(2), 206-226.

Haziza, D., and Lesage, É. (2016). A discussion of weighting procedures for unit nonresponse. *Journal of Official Statistics*, 32(1), 129-145.

Heckathorn, D.D., and Cameron, C.J. (2017). Network sampling: From snowball and multiplicity to respondent-driven sampling. *Annual Review of Sociology*, 43, 101-119.

Hibben, K.C., Smith, Z., Hoppe, T., Ryan, V., Rogers, B., Scanlon, P. and Miller, K. (2022). Toward a semi-automated item nonresponse detector model for open-response data. Paper presented at Federal Committee on Statistical Methodology Research and Policy Conference, https://www.fcsm.gov/assets/files/docs/2022-conference-docs/D1.2_Cibelli.pdf.

Hidiroglou, M.A., and You, Y. (2016). Comparison of unit level and area level small area estimators. *Survey Methodology*, 42(1), 41-61. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2016001/article/14540-eng.pdf.

Hyman, M., Sartore, L. and Young, L.J. (2022). Capture–recapture estimation of characteristics of US local food farms using a web-scraped list frame. *Journal of Survey Statistics and Methodology*, 10(4), 979-1004.

Ikudo, A., Lane, J.I., Staudt, J. and Weinberg, B.A. (2019). Occupational classifications: A machine learning approach. *Journal of Economic and Social Measurement*, 44(2-3), 57-87.

Kalton, G. (2019). Developments in survey research over the past 60 years: A personal perspective. *International Statistical Review*, 87 (S1), S10-S30.

Kang, J.D.Y., and Schafer, J.L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, 22 (4), 523-539.

Kaputa, S.J., Morris, D.S. and Holan, S.H. (2024). Bayesian multisource hierarchical models with applications to the Monthly Retail Trade Survey. *Journal of Survey Statistics and Methodology*, 12(5), 1567-1589.

Kaputa, S.J., Thompson, K.J. and Beck, J.L. (2019). An embedded experiment for targeted non-response follow-up in establishment surveys. *Journal of the Royal Statistical Society Series A*, 182(4), 1371-1391.

Kass, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Journal of the Royal Statistical Society Series C*, 29(2), 119-127.

Keiding, N., and Louis, T.A. (2018). Web-based enrollment and other types of self-selection in surveys and studies: Consequences for generalizability. *Annual Review of Statistics and its Application*, 5, 25-47.

Kennedy, C., and Hartig, H. (2019). Response rates in telephone surveys have resumed their decline. http://www.pewresearch.org/fact-tank/2019/02/27/responserates-in-telephone-surveys-have-resumed-their-decline/.

Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J. and Asare-Marfo, D. (2021). Strategies for detecting insincere respondents in online polling. *Public Opinion Quarterly*, 85(4), 1050-1075.

Kennedy, C., Mercer, A. and Lau, A. (2024). Exploring the assumption that commercial online nonprobability survey respondents are answering in good faith. *Survey Methodology*, 50(1), 3-21. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2024001/article/00013-eng.pdf.

Kern, C., Weiß, B. and Kolb, J.-P. (2021). Predicting nonresponse in future waves of a probability-based mixed-mode panel with machine learning. *Journal of Survey Statistics and Methodology*, 11(1), 100-123.

Kim, J.K., and Kim, J.J. (2007). Nonresponse weighting adjustment using estimated response probability. *Canadian Journal of Statistics*, 35(4), 501-514.

Kim, J.K., and Shao, J. (2022). Statistical Methods for Handling Incomplete Data, Second Edition. Boca Raton, FL: CRC Press.

Kim, J.K., and Wu, C. (2013). Sparse and efficient replication variance estimation for complex surveys. *Survey Methodology*, 39(1), 91-120. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2013001/article/11826-eng.pdf.

Kim, J.K., and Yu, C.L. (2011). Replication variance estimation under two-phase sampling. *Survey Methodology*, 37(1), 67-74. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011001/article/11448-eng.pdf.

Kott, P.S. (1994). A note on handling nonresponse in sample surveys. *Journal of the American Statistical Association*, 89(426), 693-696.

Kott, P.S., and Stukel, D.M. (1997). Can the jackknife be used with a two-phase sample? *Survey Methodology*, 23(2), 81-89. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1997002/article/3621-eng.pdf.

Krennmair, P., and Schmid, T. (2022). Flexible domain prediction using mixed effects random forests. *Journal of the Royal Statistical Society Series C*, 71(5), 1865-1894.

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. *The Annals of Statistics*, 9(5), 1010-1019.

Lavrakas, P.J., Traugott, M.W., Kennedy, C., Holbrook, A.L., de Leeuw, E.D. and West, B.T. (2019). *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*. Hoboken, NJ: Wiley.

Leuenberger, M., Eustache, E., Jauslin, R. and Tillé, Y. (2022). Balancing a sample almost perfectly. *Statistics and Probability Letters*, 180, 109229.

Lin, W.-C., and Tsai, C.-F. (2020). Missing value imputation: A review and analysis of the literature (2006-2017). *Artificial Intelligence Review*, 53, 1487-1509.

Lohr, S.L. (2021). Multiple-frame surveys for a multiple-data-source world. *Survey Methodology*, 47(2), 229-263. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2021002/article/00008-eng.pdf.

Lohr, S.L., Hsu, V. and Montaquila, J. (2015). Using classification and regression trees to model survey nonresponse. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 2071-2085. Alexandria, VA.

Lohr, S.L., and Mendez, G. (2009). Regression trees with survey data. Paper presented on August 6, 2009 at Joint Statistical Meetings, Washington, DC.

Lohr, S.L., and Raghunathan, T.E. (2017). Combining survey data with other data sources. *Statistical Science*, 32(2), 293-312.

Lumley, T., and Huang, X. (2024). Linear mixed models for complex survey data: Implementing and evaluating pairwise likelihood. *Stat*, 13(1), e657.

Lundy, E.R., and Rao, J.N.K. (2022). Relative performance of methods based on model-assisted survey regression estimation: A simulation study. *Survey Methodology*, 48(1), 49-71. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2022001/article/00003-eng.pdf.

Mahalanobis, P.C. (1939). A sample survey of the acreage under jute in Bengal. *Sankhyā: The Indian Journal of Statistics*, 4(4), 511-531.

Mashreghi, Z., Haziza, D. and Léger, C. (2016). A survey of bootstrap methods in finite population sampling. *Statistics Surveys*, 10, 1-52.

McCarthy, P.J. (1966). *Replication: An Approach to the Analysis of Data from Complex Surveys.* Washington, DC: National Center for Health Statistics.

McClellan, C., Mitchell, E., Anderson, J. and Zuvekas, S. (2023). Using machine-learning algorithms to improve imputation in the Medical Expenditure Panel Survey. *Health Services Research*, 58(2), 423-432.

McConville, K.S., Breidt, F.J., Lee, T.C. and Moisen, G.G. (2017). Model-assisted survey regression estimation with the lasso. *Journal of Survey Statistics and Methodology*, 5(2), 131-158.

McConville, K.S., and Toth, D. (2019). Automated selection of post-strata using a model-assisted regression tree estimator. *Scandinavian Journal of Statistics*, 46(2), 389-413.

Mercer, A., Lau, A. and Kennedy, C. (2018). *For Weighting Online Opt-in Samples, What Matters Most?* Washington, DC: Pew Research. https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samples-what-matters-most/.

Molina, I., and Rao, J.N.K. (2023). Historical overview of small area estimation in the 50th birthday of the IASS. *The Survey Statistician*, 88, 23-35.

Montanari, G.E., and Ranalli, M.G. (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association*, 100(472), 1429-1442.

Montanari, G.E., and Ranalli, M.G. (2009). Multiple and ridge model calibration. *Proceedings of Workshop on Calibration and Estimation in Surveys*. Ottawa, ON: Statistics Canada.

Morales, D., Esteban, M.D., Pérez, A. and Hobza, T. (2021). *A Course on Small Area Estimation and Mixed Models*. Cham, Switzerland: Springer.

Nandram, B., and Rao, J.N.K. (2024). Bayesian integration for small areas by supplementing a probability sample with a non-probability sample. *Statistics Applications*, 22(1), 343-374.

National Academies of Sciences, Engineering, and Medicine (2023). *Toward a 21st Century National Data Infrastructure: Enhancing Survey Programs by Using Multiple Data Sources*. Washington, DC: The National Academies Press.

National Academies of Sciences, Engineering, and Medicine (2024). *Toward a 21st Century National Data Infrastructure: Managing Privacy and Confidentiality Risks with Blended Data*. Washington, DC: The National Academies Press.

National Center for Health Statistics (1968). *Synthetic State Estimates of Disability: Public Health Service Publication No. 1759*. Washington, DC: U.S. Government Printing Office.

Oberski, D.L., and DeCastellarnau, A. (2021). Predicting measurement error variance in social surveys. https://repositori.upf.edu/bitstream/handle/10230/48430/WP63_RECSM_sept2021.pdf.

Olson, K., Smyth, J.D., Horwitz, R., Keeter, S., Lesser, V., Marken, S., Mathiowetz, N.A., McCarthy, J.S., O'Brien, E., Opsomer, J.D., Steiger, D., Sterrett, D., Su, J., Suzer-Gurtekin, Z.T., Turakhia, C. and Wagner, J. (2021). Transitions from telephone surveys to self-administered and mixed-mode surveys: AAPOR task force report. *Journal of Survey Statistics and Methodology*, 9(3), 381-411.

Opsomer, J.D., Breidt, F.J., White, M. and Li, Y. (2016). Successive difference replication variance estimation in two-phase sampling. *Journal of Survey Statistics and Methodology*, 4(1), 43-70.

Opsomer, J.D., Claeskens, G., Ranalli, M.G., Kauermann, G. and Breidt, F.J. (2008). Non-parametric small area estimation using penalized spline regression. *Journal of the Royal Statistical Society Series B*, 70(1), 265-286.

Opsomer, J.D., and Erciulescu, A.L. (2021). Replication variance estimation after sample-based calibration. *Survey Methodology*, 47(2), 265-277. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2021002/article/00006-eng.pdf.

Opsomer, J.D., and Riddles, M.K. (2025). sCHAID: A tool for constructing nonresponse adjustment cells under a design-based framework. *Survey Methodology*, 51(1), 217-231. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2025001/article/00006-eng.pdf.

Pfeffermann, D., and Sverchkov, M. (2007). Small-area estimation under informative probability sampling of areas and within the selected areas. *Journal of the American Statistical Association*, 102(480), 1427-1439.

Phipps, P., and Toth, D. (2012). Analyzing establishment nonresponse using an interpretable regression tree model with linked administrative data. *The Annals of Applied Statistics*, 6(2), 772-794.

Platek, R. (1999). Survey Methodology – The first 25 years. *Survey Methodology*, 25(2), 109-111. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1999002/article/4874-eng.pdf.

Rao, J.N.K. (1999). Some current trends in sample survey theory and methods. *Sankhyā: The Indian Journal of Statistics, Series B*, 61(1), 1-25.

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

Rao, J.N.K. (2005). Interplay between sample survey theory and practice: An appraisal. *Survey Methodology*, 31(2), 117-138. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2005002/article/9040-eng.pdf.

Rao, J.N.K. (2021). On making valid inferences by integrating data from surveys and other sources. *Sankhyā: The Indian Journal of Statistics, Series B*, 83(1), 242-272.

Rao, J.N.K., and Fuller, W.A. (2017). Sample survey theory and methods: Past, present, and future directions. *Survey Methodology*, 43(2), 145-160. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2017002/article/54888-eng.pdf.

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

Rao, J.N.K., and Singh, A.C. (1997). A ridge-shrinkage method for range-restricted weight calibration in survey sampling. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 57-65. Alexandria, VA.

Rao, J.N.K., and Singh, A.C. (2009). Range restricted weight calibration for survey data using ridge regression. *Pakistan Journal of Statistics*, 25(4), 371-384.

Rao, J.N.K., Sinha, S.K. and Dumitrescu, L. (2014). Robust small area estimation under semi-parametric mixed models. *Canadian Journal of Statistics*, 42(1), 126-141.

Rao, J.N.K., and Wu, C.F.J. (1985). Inference from stratified samples: Second-order analysis of three methods for nonlinear statistics. *Journal of the American Statistical Association*, 80(391), 620-630.

Rao, J.N.K., and Wu, C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*, 18(2), 209-217. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1992002/article/14486-eng.pdf.

Reiter, J.P. (2021). Assessing uncertainty when using linked administrative records. In *Administrative Records for Survey Methodology*, (Eds., A.Y. Chun, M.D. Larsen, G. Durrant and J.P. Reiter), 139-153. Hoboken, NJ: Wiley.

Rizzo, L., Kalton, G. and Brick, J.M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, 22(1), 43-53. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/1996001/article/14386-eng.pdf.

Roberson, A. (2021). Applying machine learning for automatic product categorization. *Journal of Official Statistics*, 37(2), 395-410.

Robertson, B., and Price, C. (2024). One point per cluster spatially balanced sampling. *Computational Statistics and Data Analysis*, 191, 107888.

Royall, R.M., and Herson, J. (1973). Robust estimation in finite populations I. *Journal of the American Statistical Association*, 68(344), 880-889.

Särndal, C.-E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, 33(2), 99-119. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2007002/article/10488-eng.pdf.

Savitsky, T.D., Williams, M.R., Gershunskaya, J. and Beresovsky, V. (2023). Methods for combining probability and nonprobability samples under unknown overlaps. *Statistics in Transition*, 24(5), 1-34.

Schouten, B., Peytchev, A. and Wagner, J. (2018). *Adaptive Survey Design*. Boca Raton, FL: CRC Press.

Sugasawa, S., and Kubokawa, T. (2020). Small area estimation with mixed models: A review. *Japanese Journal of Statistics and Data Science*, 3, 693-720.

Sun, H., Berg, E. and Zhu, Z. (2024). Multivariate small-area estimation for mixed-type response variables with item nonresponse. *Journal of Survey Statistics and Methodology*, 12(2), 320-342.

Thompson, M.E. (2019). Combining data from new and traditional sources in population surveys. *International Statistical Review*, 87(S1), S79-S89.

Thompson, S.K. (2017). Adaptive and network sampling for inference and interventions in changing populations. *Journal of Survey Statistics and Methodology*, 5(1), 1-21.

Tillé, Y. (2020). Sampling and Estimation from Finite Populations. Hoboken, NJ: Wiley.

Tourangeau, R., Brick, J.M., Lohr, S.L. and Li, J. (2017). Adaptive and responsive survey designs: A review and assessment. *Journal of the Royal Statistical Society Series A*, 180(1), 203-223.

United Nations Economic Commission for Europe (2021). *Machine Learning for Official Statistics*. Geneva: United Nations. https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf.

Valliant, R. (2024). Hansen Lecture 2022: The evolution of the use of models in survey sampling. *Journal of Survey Statistics and Methodology*, 12(2), 275-304.

Verret, F., Rao, J.N.K. and Hidiroglou, M.A. (2015). Model-based small area estimation under informative sampling. *Survey Methodology*, 41(2), 333-347. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2015002/article/14248-eng.pdf.

Wang, L., Valliant, R. and Li, Y. (2021). Adjusted logistic propensity weighting methods for population inference using nonprobability volunteer-based epidemiologic cohorts. *Statistics in Medicine*, 40(24), 5237-5250.

Winkler, W.E. (2021). Cleaning and using administrative lists: Enhanced practices and computational algorithms for record linkage and modeling/editing/imputation. In *Administrative Records for Survey Methodology*, (Eds., A.Y. Chun, M.D. Larsen, G. Durrant and J.P. Reiter), 105-138. Hoboken, NJ: Wiley.

Wiśniowski, A., Sakshaug, J.W., Perez Ruiz, D.A. and Blom, A.G. (2020). Integrating probability and nonprobability samples for survey inference. *Journal of Survey Statistics and Methodology*, 8(1), 120-147.

Wu, C. (2022). Statistical inference with non-probability survey samples (with discussion). *Survey Methodology*, 48(2), 283-311. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2022002/article/00002-eng.pdf.

Wu, C. (2023). Calibration techniques for model-based prediction and doubly robust estimation. *The Survey Statistician*, 88, 86-93.

Wu, C., and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96(453), 185-193.

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.

Yi, G.Y., Rao, J.N.K. and Li, H. (2016). A weighted composite likelihood approach for analysis of survey data under two-level models. *Statistica Sinica*, 26(2), 569-587.

Yung, W., Karkimaa, J., Scannapieco, M., Barcarolli, G., Zardetto, D., Sanchez, J.A.R., Braaksma, B., Buelens, B. and Burger, J. (2018). *The Use of Machine Learning in Official Statistics*. https://unece.org/sites/default/files/2024-07/HLGMOS The use of machine learning in official statistics.pdf.

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

Zhang, S., and Wagner, J. (2024). The additional effects of adaptive survey design beyond post-survey adjustment: An experimental evaluation. *Sociological Methods & Research*, 53(3), 1350-1383.