2

Some of Statistics Canada's Contributions to Survey Methodology

Jean-François Beaumont, Susie Fortier, Jack Gambino, Mike Hidiroglou, and Pierre Lavallée

Statistics Canada, Ottawa, ON

The conduct of large scale surveys in national statistical offices follows a well-established process that involves a large number of steps. These steps, which are described in Statistics Canada's *Quality Guidelines* (Statistics Canada, 2008), begin with the setting of objectives for the survey and end with dissemination and documentation. In this chapter, we describe some of the important developments that have taken place at Statistics Canada to improve the intermediate steps. Some of the steps that we will describe include record linkage, sample design, editing, imputation, weighting, estimation, variance estimation, data analysis, benchmarking, and the use of time series. Many of these developments were necessary to improve the quality of the resulting published estimates, to reduce costs, and to aid in the building of either specific or generalized software to automate the methodology.

2.1 Introduction

When most Canadians think of Statistics Canada, they probably think of one of three things: the monthly consumer price index, the monthly unemployment rate or the Census. It is not an exaggeration to say that they are seeing just the tip of an enormous information iceberg. Statistics Canada produces a large volume of information on a daily basis — so much so that since 1932 it has had a publication called *The Daily* (accessible on the web at `http://www.statcan.gc.ca`). *The Daily* was literally a daily (weekday) paper publication initially and is now published on the web. The web version of *The Daily* first appeared in 1995. On any given day, *The Daily* typically covers several major releases, and each release may include hundreds, sometimes thousands, of estimates. For example, when the Labour Force Survey results are published in the early part of each month, the media focus is on the national employment and unemployment numbers, but in fact, thousands of estimates are released, broken down by geography, age and sex, and covering not only employment and unemployment, but also type of occupation and industry of employment, wages and other variables.

The information published by Statistics Canada is of interest not only to the general public but also to policy makers at all levels of government, as well as to non-governmental organizations, business planners, social and economic researchers, and so on. The range of topics for which the agency produces information is very broad. It includes data related to the economy (labour, commerce, trade, transportation, energy, etc.) and society (health, education, travel, tourism, the justice system, etc.). Many data users may not realize how much lies behind the numbers that they use to make decisions and conduct their studies: where the data come from, how they are collected and aggregated, and what their strengths and weaknesses may be. One of our goals in this chapter is to bring to light the statistical methods underlying the information that is disseminated by the agency.

These methods have evolved over time to satisfy the changing needs of data users and to take advantage of innovations due to technological change. Statistics Canada's mathematical statisticians (often referred to internally as "survey methodologists" or simply "methodologists") have made major contributions to the branches of statistics known as survey methodology and sample survey theory and methods. A second goal of this chapter is to discuss these contributions and, in the process, introduce numerous domains that may be unfamiliar not only to the general public but to other statisticians as well. Some of these domains are familiar to all statisticians: estimation, variance estimation, data analysis, and time series. But some may not be: imputation, multiple frames, and record linkage.

As a result of its research and development program, Statistics Canada is viewed as a world leader in survey methodology and its methods and approaches to research and development have been used as models by numerous national statistical organizations throughout the world.

The chapter is organized as follows. Section 2.2 presents innovations related to sample design. Sample design is a very important part of a survey. It involves all the steps needed to define and specify the population frame, sample size, sample selection method, and sample estimation method. Once this is established, the data are collected and processed. At this stage, survey statisticians need to validate and sometimes replace incorrect or missing data. This is called edit and imputation and is discussed in Section 2.3. Section 2.4 highlights various contributions from Statistics Canada to the estimation process. This is where the data, possibly in combination with auxiliary information from other sources, are used to estimate the values of interest in the population. The quality of this estimation is often measured in terms of variance, and variance estimation is presented in Section 2.5. Section 2.6 focuses on how survey statisticians handle the challenging task of data analysis, especially in the context of a complex survey design. Finally, Statistics Canada's contributions to the field of time series — defined as sequences of estimates ordered in time — are featured in Section 2.7.

2.2 Sample Design

Sample surveys are carried out by selecting samples of persons, businesses or other entities such as farms (called units) that we survey in order to get information. Sample selection is often done by randomly selecting certain units from a list that we call a sampling frame. This sampling frame represents the set of units for which we want to produce information; this is what makes up the target population. Obtaining the sampling frame is often difficult and costly because we want it to be complete (all units of the target population are contained in the frame), unique (every unit is present only once in the frame), and exclusive (no units of the frame are outside the target population).

The sample design describes the way that the random sample is selected. For instance, we may decide to divide the sampling frame into subgroups (called strata) to deal with more homogeneous subpopulations. We may also choose to select the sample using selection probabilities proportional to the size of the units. All of this is contained in the sample design. We will describe more aspects of the sample design as we discuss some of the contributions of Statistics Canada in this domain.

2.2.1 Record Linkage

Data from different sources are increasingly being combined to augment the amount of information that we have. This is the case especially with the creation of sampling frames. Often, the sampling frame is built by combining different sources using record linkage. Another application of record linkage is to aid in data collection by using address information to match publicly available telephone numbers with sampled households.

The concepts of record linkage were introduced by Newcome et al. (1959), and formalized mathematically by Fellegi and Sunter (1969). As described by Bartlett et al. (1993), record linkage is the process of bringing together two or more separately recorded pieces of information pertaining to the same unit (individual or business). Record linkage is sometimes called exact matching, in contrast to statistical matching. In the latter case, linkages are based on similar characteristics rather than unique identifying information.

The purpose of record linkage is to link the records from two files. If the records contain unique identifiers, then the matching process is trivial. Unfortunately, often a unique identifier is not available and then the linkage process needs to use some probabilistic approach to decide whether two records, coming respectively from each file, are linked together or not. With this linkage process, the probability of having a real match between two records is calculated. Based on the magnitude of this probability, it is then decided whether the two records can be considered as really being linked together or not.

With the approach of Fellegi and Sunter (1969), we choose an upper threshold and a lower threshold to which each linkage probability is compared. These thresholds are chosen to reduce the linkage errors coming from accepting bad links and rejecting good links. During the linkage process, if the linkage probability between two records is greater than the upper threshold, the link between these records is considered a true link. If the linkage probability is lower than the lower threshold, we consider that we do not have a link. Last, if the linkage probability is between the lower and upper thresholds, then we have a possible link, and we will decide on this link by manual resolution. This is generally done by looking at the data, and also by using auxiliary information.

In addition to frame creation, Statistics Canada uses record linkage in various other applications. It is used, for example, to add information to files for sociological or economic analysis and in the coverage studies of the census.

2.2.2 Multiple Frames

To cover the target population well, one sampling frame may not be sufficient. For example, for agriculture surveys, we may have available a list of farms, which can be used as a list frame. But the list frame may be incomplete (for example, newer farms may not be on it). We can also select farms by first selecting small land areas and then farms within those areas; the set of land areas is an area frame. The survey may decide to use both the list and area frames to select its sample. When the units involved in each frame are the same, it may be possible to combine the frames into a single one using, for instance, record linkage. However, in the agriculture example, the different frames do not contain the same type of unit and, in this case, we may decide to use the frames separately. We then talk about the use of multiple frames.

When using multiple frames, we are typically selecting a sample from each frame independently. Different approaches are then possible to produce estimates from the resulting samples. In the approach by Hartley (1962), who was the first to formalize estimation for multiple frames, estimates from each sample portion (i.e., the sample in the intersection of both frames, and the samples from the non-overlapping portions of the frames) are combined to produce a global estimate.

Bankier (1986) developed at Statistics Canada an alternative method for producing estimates from a multiple frame survey. Even though he confined his method to sample designs for which a stratified simple random sample is selected independently from each frame, the estimation technique can be applied to more complex sample designs. Bankier (1986) viewed a multiple frame sample as a special case of selecting two or more samples independently from the same frame. The idea was to combine the different samples selected from the frames, and to compute the selection probability of each unit of the combined sample. As a result, standard techniques for estimating parameters from a single frame can be applied to multiple frame samples. The resulting estimators have been found to be simpler to compute and are easier to extend to three or more frames than those given in Hartley (1962).

2.2.3 Indirect Sampling

In practice, there may be no available sampling frame that directly corresponds to the desired target population. We then have to use a sampling frame that is indirectly related to this target population. An example of this is the following. Suppose that estimates are required for young children, but that the sampling frame is a list of their parents. In this case, the target population is young children; however, they are selected in two steps. First, a sample of the parents is selected using the list of the parents, and second, a sample of their children is selected. Note that the children of a particular family can be selected through the father or the mother. We then have two populations that are linked with one another: one is associated with the sampling frame and the other with the target population (parents and children respectively in our example). This type of sampling is referred to as indirect sampling (Lavallée, 2002, 2007).

Estimation of population parameters, such as totals and means using data selected by indirect sampling is not straightforward, particularly if the links between the units of the two populations (the sampling frame and the target population) are not one-to-one. If we consider the previous example of families, it can be very difficult to assign a selection probability to each child of a selected family. Indeed, we could have selected a family through one or more of the parents but the computation of the selection probability of the family, and consequently that associated with each child, requires that the selection probability of each parent is known, whether the latter is selected or not. Lavallée (1995) developed the generalized weight share method (GWSM) to solve this estimation problem. The GWSM is a generalization of the weight share method described by Ernst (1989). We can also consider indirect sampling and the GWSM as a generalization of network sampling as well as adaptive cluster sampling. These two sampling methods are described by Thompson (2002) and Thompson and Seber (1996).

2.2.4 Optimal Stratification

Populations are typically not homogenous, and sampling from them would result in estimates with high variability. They are therefore split into subpopulations (called strata) that are relatively homogeneous, and mutually exclusive. For example, a population of persons or businesses can be partitioned into low, medium and high income strata. The stratification of the population depends on the type of survey that is carried out. For example, in the case of household surveys, the stratification is mainly based on geography (e.g., provinces), whereas for business surveys it is based on a combination of geography, industrial classification and size. The stratification of a population may also be constrained by the level at which estimates are required. For example, if estimates for a household survey are required for the larger cities in a given province, the stratification of that province will account for that requirement. Once the strata are formed, sample selection takes place independently within each of them.

The use of a stratified sample plan involves five different design operations: (i) choosing a stratification variable(s); (ii) determining the number of strata; (iii) stratifying the population given the chosen stratification variable(s); (iv) allocating the total sample size to the strata; (v) selecting the units within each stratum according to the adopted sampling design. Note that although any sampling design can be used in the fifth step, simple random sampling (SRS) without replacement is often used as the method for selecting the units. This procedure is extensively used for sampling from list frames, especially for business surveys.

Some populations, such as populations of businesses, can be highly skewed, i.e., there will be many small and medium sized units and just a few very big units. Efficient sampling of highly skewed populations requires that they be stratified into a take-all stratum and a number of take-some strata. In our example, the take-all stratum would consist of the biggest units. All units in the take-all stratum are selected with certainty, whereas units in the take-some strata are selected via a probability mechanism, such as SRS. Given that a quantitative stratification variable is highly correlated with the variables of interest, we search for optimal stratification boundaries that split the population into subpopulations in such a way that large units will be selected with certainty (take-all stratum) and smaller units will be selected with some positive sampling fractions (take-some strata). For example, we may want to stratify the frame by small, medium and large businesses based on a revenue variable. The problem is to find the stratum boundaries that will categorize the small, medium and large businesses.

Approximate cut-off rules for stratifying a population into a take-all and a single take-some stratum were given by Hidiroglou (1986). Lavallée and Hidiroglou (1988) considered the more complex case where the population is to be stratified into a single take-all stratum and several take-some strata. The algorithm of Lavallée–Hidiroglou minimizes the overall sample size required for the survey, given the precision of the estimator and the allocation scheme of the sample to the take-some strata. The allocation scheme used by the algorithm is power allocation, as described in the next section.

The Lavallée–Hidiroglou algorithm has been used extensively in most business surveys at Statistics Canada, and elsewhere. The main program where it has been used is the Unified Enterprise Survey (Li et al., 2011), a business survey that unifies more than 60 surveys for different industries.

2.2.5 Power Allocation

In many surveys, reliable estimates are required both at the national level and for subnational areas. These subnational areas can be, for example, the provinces of Canada. The requirement for subnational estimates usually results in the sample being stratified geographically. When the subnational areas vary considerably in size or importance (e.g., in terms of population in a social survey or total revenue in a business survey), this variation must be taken into account when deciding how to allocate the sample to strata. For example, allocating sample to provinces proportional to their size may produce good precision at the national level. However, it may cause certain subnational estimators (e.g., for the smallest provinces) to have poor precision. Alternatively, a sample allocation that achieves nearly equal precision for the subnational estimators may result in the national estimator having much lower precision than under the above scheme. A compromise between these two types of allocations was developed by Bankier (1988), which has been called power allocation.

Power allocations were proposed in the past by Carroll (1970) and Fellegi (1981). The power allocation of Bankier (1988) creates a compromise between two different sample allocations by the use of a parameter appearing as a power (in the mathematical sense). By setting the power to different values, we get various allocations, including the two mentioned in the previous paragraph as special cases: they correspond to the power 0 and 1, respectively. By choosing a suitable power between 0 and 1, we create a compromise allocation that produces subnational estimators of approximately equal precision, without reducing too much the precision at the national level.

Power allocations have been used in the design of several surveys at Statistics Canada. For example, they have been applied to a sample survey of business tax returns where the stratification size measures were based on sales. In addition, they have been used for samples stratified by province that evaluate different aspects of the Canadian census.

2.3 Edit and Imputation

Once the sample is selected, there are numerous ways of collecting data. For instance, a paper questionnaire can be filled in by the respondent and returned by mail, responses can be recorded by an interviewer on a computer during a telephone or personal interview, or an Internet questionnaire can be filled in directly by the respondent. Regardless of the data collection mode, it is very likely that some returned questionnaires will contain errors and missing values. Computer-assisted data collection can reduce the occurrence of certain types of error by integrating edit rules in the computer application. Edit rules are used to verify that responses are valid and consistent. An example of an edit rule is that the age of a mother must be greater than the age of any of her children. However, there is a limit to the number of edit rules that can be integrated in the computer application without unduly lengthening the time needed to complete the questionnaire. As a result, errors in the collected data must be anticipated in most surveys.

After data collection, the goal of editing is to detect invalid responses and inconsistencies, and treat these errors either through a follow-up with respondents or through an automated algorithm. Statistical agencies long ago recognized that following up all the respondents with suspicious or erroneous values was a costly activity. In the context of business surveys, Latouche and Berthelot (1992) suggested coping with this issue by following up only a subset of these respondents, namely those that are expected to have a non-negligible impact on the survey estimates. They proposed three score functions to prioritize respondents to be followed up and concluded that their approach, often called selective editing, led to estimates that are similar to those obtained after following up all the respondents. Granquist and Kovar (1997) argued in favor of selective editing as a tool for reducing costs while maintaining the quality of estimates. They also suggested that editing should not only be viewed as a correction tool but also as a process that provides information about the quality of survey data. Such information could then serve as a basis for future improvements to the whole survey.

Although edit rules can be used to detect inconsistencies, they do not generally tell us how to resolve them. To do so, the first step is to determine which values need to be modified to resolve inconsistencies. This step is called error localization. Imputation, which is the process of replacing values that need modification or filling in missing values, can then be carried out. The goal of imputation is to produce a complete file that satisfies all the edit rules. Such a file simplifies the production of estimates for ultimate users who may not be familiar with more complex statistical techniques dealing with inconsistent and missing values.

Prior to the seminal paper of Fellegi and Holt (1976), there was no unified theory of edit and imputation. They developed such a theory and proposed the minimum-change principle for error localization. Their principle consists of changing the smallest number of values in a record so as to remove inconsistencies. They also proposed an algorithm to solve this minimization problem. The minimum-change principle is still used and is implemented in Statistics Canada's BANFF edit and imputation system. A description of BANFF can be found in Kozak (2005).

Inconsistencies represent one type of error. There are also other types of potential errors that need to be detected such as outliers, which are unusually large or small values. There are numerous univariate outlier detection techniques in the literature. In practice, survey analysts are often interested in detecting outliers in the ratio of collected values from one period to the next. Univariate outlier detection methods can be applied to ratios. However, ratios tend to be more volatile for smaller previous-period values than for larger previous-period values in business surveys. This implies that the straight application of univariate techniques would tend to detect too many outliers for smaller previous-period values and not enough for larger previous-period values. Hidiroglou and Berthelot (1986) proposed a simple method for tackling this problem that is used in many surveys in different statistical organizations and is implemented in BANFF. See Section 2.4.2 for a more general discussion on outliers.

We have already noted that, once errors have been identified and follow-ups have been completed, imputation can be carried out to complete data cleaning. Imputation replaces missing values or values that need modification by other values, such as predicted values from a linear regression model. A key feature of imputation methods is that they rely on auxiliary information available for all the sample records, including those with missing or erroneous values. In business surveys, different types of imputation methods are used. One popular imputation method is linear regression imputation. Another popular imputation method, especially in social surveys, is donor imputation. It replaces missing values or values that need modification by the value of a "donor" chosen among the set of records that are deemed to have correct values. Characteristics of different imputation methods in business surveys are discussed in Kovar and Whitridge (1995). Most of these methods are implemented in BANFF. Greater detail about imputation methods and the associated theory can be found in Haziza (2009).

In the context of donor imputation in the Canadian census, it was noted that some implementations of the minimum-change principle sometimes led to imputed records that seem implausible. A new imputation approach was developed to handle this issue; see, e.g., Bankier et al. (1999). Instead of localizing errors and then imputing, Bankier et al. proposed to reverse this order. In their approach, close donors are first determined for each record with missing or inconsistent values, which is called a recipient. Then, for each donor-recipient pair, imputation actions are found such that the chosen actions minimize a weighted average of two distances. An imputation action is chosen at random for a recipient among the set of best imputation actions for that recipient. This approach is an attempt to compromise between satisfying the minimum-change principle and plausibility. It has been implemented in the Canadian Census Edit and Imputation System (CANCEIS) developed at Statistics Canada.

2.4 Estimation

Once the data have been collected and validated, they are used to estimate parameters of interest, including simple statistics such as means, totals, and proportions as well as more complex parameters. The latter are usually linked to the analysis of the survey data. Examples include the estimation of regression parameters and chi-squared tests.

Formally, estimation is the survey process in which unknown population parameters are estimated using data from a sample, possibly in combination with auxiliary information from other sources. Estimation results are used to make inferences about these unknown parameters, that is, to draw conclusions about characteristics of the complete population.

Survey data are typically obtained via a sampling procedure, and estimators of the parameters of interest must reflect this. The first step is to assign a design weight to each of the responding sampled units. The design weight is the inverse of the probability of selection associated with this unit. For example, if one percent of a population is selected completely at random, then the probability of selection for each unit is $1/100$ and the sampling weight is 100. If the sampling design is multi-stage or multi-phase, the design weight of a given unit will be the product of the design weights at each stage or phase. This ensures that the estimates are unbiased, or approximately unbiased, in the sense that the average value of the estimator over all possible samples is equal to the population parameter being estimated.

Once the design weights have been computed, they may need to be adjusted for nonresponse when all or almost all data for a sampled unit are missing. Nonresponse occurs when information for some units cannot be collected. This can happen for a variety of reasons, including refusals, inability to contact the unit (e.g., because of incorrect contact information or because the potential respondent is on vacation), natural disasters (e.g., floods, major winter storms) and so on. Because of nonresponse, design weights are adjusted upward based on the assumption that the responding units represent both responding and non-responding units. These modified weights are called survey weights. For a given sampled unit i, we denote its corresponding survey weight by w_i. Estimates are then obtained for the parameters for arbitrarily defined subsets of the population, i.e., for domains. Common domains of interest are the individual strata and the entire population. Domains may also cut across the design strata; for example, the domain may consist of all companies that have been in existence for less than five years. If we are interested in estimating totals for a given domain d, then the estimate for that domain is given by

$$\hat{Y}_{(d)} = \sum_{i \in s_d} w_i y_i,$$

where y_i is the variable of interest and s_d is the part of the sample s that belongs to that domain. In the example, s is the sample of companies and s_d is the set of companies in the sample that are less than five years old.

Auxiliary data, such as age-sex population counts and other information from external sources, are usually incorporated in the weighting process at Statistics Canada. There are two main reasons for using auxiliary data in estimation. The first reason is that it is often important for the survey estimates to match known population totals or estimates from another, more reliable, survey. This can be viewed as a desire for consistency. The second reason is that the resulting estimators will be more efficient (in the sense of smaller variance) than those only weighted with the survey weights, if the auxiliary data are well correlated with the variables of interest. Examples of estimators that incorporate auxiliary data range from the ratio estimator that uses a single auxiliary variable to more complex estimators that use several variables via regression or calibration. This range of estimators can be obtained via the calibration theory given in Deville and Särndal (1992). Calibration adjusts the survey weights by incorporating auxiliary variables. The resulting weights $\tilde{w}_i$ known as calibrated weights are obtained by minimizing a function of the calibrated weights and the survey weights subject to a benchmark constraint. Given that x_i are the auxiliary data, the constraint requires that the sum of the product between the calibrated weights $\tilde{w}_i$ and the auxiliary data ($\sum_{i \in s} \tilde{w}_i x_i$) equals the known population totals ($\sum_{i \in U} x_i$), where U represents the whole population.

An important application of this method is used by most household surveys. The weights of individuals in the sample are adjusted so they add up to population-level age-sex totals, which are the auxiliary variables in this case. For example, if there are P boys aged 5–9 in the population, then the weights of boys in the sample who are in this age group are adjusted so they add up to P (this P corresponds to the final summation in the previous paragraph, with x equal to 1 for boys aged 5–9 and 0 otherwise). This is done simultaneously for all age groups and both sexes. In practice, household surveys that produce both person-level and household-level estimates (such as the unemployment rate and average household income, respectively), may impose an additional constraint on the weights: the weights of all persons in a given household are forced to be equal. This is sometimes referred to as an integrated weight and is described by Lemaitre and Dufour (1987).

2.4.1 Generalized Estimation System

The development of the Generalized Estimation System (GES) in the early nineties allowed Statistics Canada to consolidate numerous estimation procedures for cross-sectional surveys. GES is currently applied to a multitude of Statistics Canada surveys. Specifications for a general estimation system were initially written in 1990 and 1991. The methodology that underpins GES is described in Estevao et al. (1995). GES is built around the following elements: the sampling plan, the population parameters to be estimated, the use of auxiliary information, and domains of interest. The sampling designs covered by GES include not only simple designs such as stratified simple random sampling, where all units in a stratum have the same probability of selection, but also a variety of complex designs, where units may be selected in several steps and need not have the same probability of selection (e.g., the probability may be related to the size of each unit). GES computes estimates of totals, means, and ratios with their associated measures of reliability given that auxiliary information has been incorporated in the estimation process. This auxiliary information can cut across design strata, or be included within them. This allows the computation of most of the commonly used estimators in survey sampling, including separate and combined ratio or regression estimators (or intermediate combinations), post-stratified estimators (separate, combined or a hybrid version of those two procedures), and others such as the raking ratio estimator. Estimates and their associated measures of reliability are computed by GES for user-specified domains of interest.

2.4.2 Outliers

Outliers are observations that are numerically distant from the rest of the data. Grubbs (1969) defined an outlier as follows: "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs." Outliers can occur by chance in any survey, but they are often indicative of a population that has a heavy-tailed distribution. Such surveys run into the danger of including outliers in their sample if these populations have not been properly (or cannot be) stratified by size. In the case of business surveys, this will occur when there is tremendous growth in the size of some businesses that has not yet been recorded on the frame. For household surveys, this may occur when financial data are collected. The presence of outliers in the sample may result in unrealistically high or low estimates of population parameters, such as means and totals.

Even though an efficient sample design can minimize outlier problems, it cannot eliminate them. This means that outliers must be detected and treated. Outliers can be detected via univariate or multivariate methods depending on the dimensionality of the variables. Their impact on the estimates can be dampened by trimming their values (Winsorizing), reducing their weights, or using robust estimation techniques such as M-estimation. These three methods have been studied at Statistics Canada, and these procedures are discussed in Lee (1995) and in Beaumont and Rivest (2009).

In the context of simple random sampling without replacement, Rao (1971) suggested changing the weights of the outlier units to 1, and adjusting the weights of the remaining units in the sample. Hidiroglou and Srinath (1981) proposed alternative weight modifying procedures. Additional work on a Winsorization approach to outliers was done by Tambay (1988), who designed a procedure that performs well on data from skewed populations.

2.4.3 Composite Estimation

Monthly surveys such as the Canadian Labour Force Survey (LFS) deliberately retain a part of their sample in consecutive months. The LFS does this by changing (or rotating) only one sixth of its sample each month. The LFS sample is a set of dwellings (i.e., single family homes, apartment units, etc.). As a result, five sixths of the dwellings are the same in consecutive months, four sixths are the same two months apart, and so on. This is done primarily to reduce costs: interviews can be shorter in subsequent months, less travel is needed since subsequent interviews can usually be conducted by telephone, etc. A benefit of having some overlap in the sample is that it can be used to improve the quality of the estimates produced by the survey.

For example, to estimate the month-to-month change in employment, it is advantageous to focus on the five-sixths of the sample that is common because any observed differences are not due to sampling error. In fact, this estimate of change can be added to last month's overall estimate to update it, thereby giving us an alternative estimate of this month's employment. As a result, we then have two estimates for this month's total: the "usual" one based only on this month's sample, and a second estimate derived by updating last month's estimate using the common (or "matched" sample) to estimate change (Statistics Canada, 2008).

Whenever we have two estimates of the same quantity, it is natural to combine them in some way to obtain an estimate that we hope will be better than its parts. In the survey world this combined estimate is referred to as a composite estimate. In the case of the LFS, if we are interested in the number of employed people in the current month, then the composite estimate of employed is a linear combination of the above two estimates. It can be expressed as

$$\begin{aligned} &C \times (\text{“usual” estimate}) \\ &\qquad + (1 - C) \times (\text{last month's estimate} \\ &\qquad\qquad + \text{estimate of change based on the common sample}). \end{aligned}$$

The constant C is chosen to make the variance of this combination as small as possible. The estimator used by the Canadian LFS is a more elaborate version of this and is described by Singh et al. (2001), Gambino et al. (2001), and Fuller and Rao (2001).

2.4.4 Small Area Estimation

Sample surveys are generally designed to provide reliable estimates for large areas or major domains of a population. Direct survey estimates such as those computed by GES are likely to yield unacceptably large sampling errors if the domain size is small. Examples of such domains include small communities and small subpopulations. For example, the Labour Force Survey is designed to produce good estimates of unemployment at the province level, but unemployment estimates for small cities and for occupations such as nursing are not expected to be of high quality because of the limited sample size in those domains. However, there is a need for such estimates, and the methodology that has been developed to meet this demand is known as small area estimation. Small area estimation, for a given small area, essentially combines in an optimal manner the associated direct estimates with model-based estimates. The model-based estimates involve known population totals (auxiliary data) and estimates of the regression between the variable of interest and the auxiliary data across the small areas. In general, these models are classified into two groups: unit level models and area level models. Unit level models are generally based on observation units (e.g., persons or companies) from the survey and auxiliary variables associated with each observation, whereas area level models are based on direct survey estimates aggregated from the unit level data and related area-level auxiliary variables; see Rao (2003) for an overview of small area models.

The basic area level model is based on the methodology proposed by Fay and Herriot (1979). This model takes into account the survey design through the use of direct survey estimates and related design-based variance estimates. A variation of the Fay–Herriot model was proposed by You and Rao (2002a) to deal with the estimation of proportions such as the unemployment rate. Their approach was used at Statistics Canada in two instances: (i) You and Rao (2002a) estimated the Canadian census undercount via a specific unmatched model; and (ii) You (2008) proposed a time series model to estimate unemployment rates using Labour Force Survey data.

You and Rao (2002b) also contributed to the unit level approach for small area estimation. Their proposed estimator has several desirable properties; in particular, it satisfies a benchmarking property in that its estimators for the domains add up to a regression estimator of the overall total. This method has been used in many official statistics programs for different surveys.

The methods in the above two paragraphs have been incorporated in a Generalized Small Area Estimation prototype system that Statistics Canada is currently building (Estevao et al., 2012, 2013).

2.5 Variance Estimation

Estimates produced from a survey are subject to errors that need to be quantified. It is important to quantify these errors to give users of the associated estimates an idea of how reliable these estimates are. The most widely used statistic to measure this reliability is the variance. For convenience, a related measure, the coefficient of variation (CV) is often presented. The CV is a relative measure that is usually expressed as a percentage, making it easier to compare different estimates.

The variance (and the CV) measure one particular type of error, namely, sampling error. In survey sampling, in addition to sampling error there are also non-sampling errors.

Sampling error is the error that results from estimating a population characteristic by measuring a portion of the population rather than the entire population. The most commonly used measure to quantify sampling error is sampling variance. Sampling variance measures the extent to which different possible samples of the same size and the same design yield differing estimates of the same characteristic. Factors affecting the magnitude of the sampling variance include:

a) the variability of the characteristic of interest in the population (the more variable the characteristic in the population, the larger the sampling variance);

b) the size of the population (the size of the population only has an impact on the sampling variance for small to moderate sized populations); and

c) the sample design and method of estimation.

For sample designs that use probability sampling, the magnitude of an estimate's sampling variance can be estimated on the basis of observed variation in the characteristic among sampled units (i.e., based on the values observed for the units in the one sample actually selected). The estimated sampling variance depends on which sample was selected and varies from sample to sample. In other words, the estimated variance itself has a variance.

A non-sampling error, on the other hand, is a catch-all term for all other possible errors associated with the sample. These include under or over-coverage of the sampling frame, nonresponse, measurement and recall errors, etc. Although the non-sampling errors can be measured individually, there are no methods that combine them all into a single measure. One exception is that the variation introduced into estimation by either nonresponse adjustments or imputation can be incorporated in the variance.

The estimation of the sampling variance can become difficult if the estimated parameters are statistics that are not simple linear functions of the observed data or if the estimators incorporate auxiliary data. Examples of complex statistics include medians and functions of totals or means. An example of an estimator that incorporates auxiliary data is the calibration estimator discussed in Section 2.4. In this section, we focus on a number of developments that took place at Statistics Canada regarding the estimation of the sampling variance, as well as the estimation of variance reflecting nonresponse adjustments or imputation. These developments can be split into two broad classes: (i) Procedures based on Taylor linearization; and (ii) Procedures based on replication.

2.5.1 Taylor Linearization

For linear statistics, the estimated variance incorporates the first order and second order inclusion probabilities of the sample, and the observed values of the sample. Here, first order refers to the probability that a unit will be included in the sample, and second order refers to the probability that a given pair of units both will be selected. For nonlinear statistics, however, variance estimation can be much more difficult. Fortunately, the variance estimation procedures for linear statistics can be used to approximate the variance of nonlinear ones by using Taylor's theorem to approximate (or "linearize") the nonlinear statistics and then obtaining the variance of this linear approximation.

Binder (1983) used linearization to obtain the estimated variance of complex statistics represented as weighted estimating equations. Special cases of his approach include the generalized linear models described by Nelder and Wedderburn (1972), logistic regression, and log-linear models for categorical data. The Binder procedure is highly quoted in the literature, and is used to obtain variances for parameters other than the ones just mentioned.

Demnati and Rao (2004) presented a unified approach to deriving Taylor linearization variance estimators and applied it to a variety of problems. Their methodology produces the correct variance estimates for many complex statistics used in survey sampling. A non-comprehensive list of those complex statistics includes proportions, ratio estimates, generalized linear regression models, and the Wilcoxon two-sample rank-sum test.

2.5.2 Replication Methods

Replication methods involve selecting several subsets of the sample that are representative of the population, i.e., each so-called replicate is a subset of the sample. The number of units included in each replicate is roughly the same. Each replicate is selected using the same design. Since each replicate is essentially a sample itself, it can be used to produce an estimate. Therefore, estimates of the parameter of interest can be computed for each replicate in exactly the same way that they were computed for the full sample. It turns out that the variation in these replicate estimates is related to the variance we are trying to estimate, and this fact is used to obtain a valid variance estimate. There are several ways of creating the replicates. The two most widely used replication procedures at Statistics Canada are the jackknife and the bootstrap.

The jackknife. A very common way of estimating the variance using the jackknife method is to form each replicate by removing a single sampled unit (cluster or element) from the original sample, and recomputing the estimate for the parameter of interest. This procedure, known as the delete-one jackknife, can become time consuming for a survey with many clusters or elements, whereas the corresponding Taylor procedure requires fewer computations. Given an estimator and its corresponding variance expression in terms of the jackknife procedure, Yung and Rao (1996) showed how to linearize the jackknife variance estimator to obtain a linearization-type variance estimator for post-stratified and regression estimators.

Two-phase sampling poses interesting variance estimation challenges for complex estimators. A two-phase sampling design works as follows. A large first-phase sample is selected and some minimal information is collected at a small cost. The information collected at the first-phase is then used as auxiliary data for the selection of a subsample (the second-phase sample) drawn from the units of the first-phase sample. For example, the information collected at the first phase can be used as stratification variables. In the context of agriculture surveys, for example, the first phase of sampling can identify farms with specific crops, and the second phase can measure pesticide use on those crops.

To estimate the variance of complex estimators under two-phase sampling, Kott and Stukel (1997) tackled the problem using the jackknife technique. They showed that the jackknife may be used to estimate the variance for one common type of estimator (the reweighted expansion estimator) under certain conditions; it is not generally effective as a variance estimator for another type of estimator (the double expansion estimator). They found that the jackknife variance estimator can be nearly unbiased for the reweighted variance estimator, while the jackknife can fail as a variance estimator for the double expansion estimator. These results are found to be very useful at Statistics Canada because several social surveys use multi-phase sample designs, and also because a two-phase sample design is planned for the Integrated Business Statistics Program (Turmelle et al., 2012).

The bootstrap. The bootstrap differs from the jackknife in that random subsamples are selected with replacement from the original sample a number of times. Each random subsample is viewed as a replicate. The weights associated with each replicate depend on the original sampling design. Rao and Wu (1988) provide details on how these weights can be generated for fairly complex sample designs. For each subsample, the same estimation procedure used for the entire sample is repeated. The resulting estimates for each replicate are then used to estimate the variance. The bootstrap method most commonly used at Statistics Canada was originally described by Rao et al. (1992).

In practice, instead of resampling sampled units, the bootstrap can be carried out by attaching a series of adjustment weights to each observation in the original sample. Beaumont and Patak (2012) generated these weights in a novel way that maintains key properties of the sampling error (specifically the first two moments). It can be applied to most sampling designs. Most previous bootstraps, such as the one by Rao and Wu (1988), are special cases of their procedure.

2.5.3 Variance Estimation in the Presence of Imputation

The above variance estimation methods can deal with the variability due to selecting a sample of the finite population. There are many other sources of error. One important source of error that can be accounted for at the variance estimation stage is the error due to missing values. Many developments were made at Statistics Canada on this topic over the last 25 years, especially when imputation is used to compensate for the missing values. Two excellent reviews are Lee et al. (2001) and Haziza (2009). The earliest work on this subject at Statistics Canada is described in handwritten notes by Mike Hidiroglou from

1989. In these unpublished notes, Hidiroglou provided the required variance estimator for some special cases of regression imputation. The methodology is based on modeling the variable to be imputed.

A dataset will most likely have been imputed using a mixture of imputation procedures. Beaumont and Bissonnette (2011) provided a simple methodology to estimate variances for datasets that were imputed using such a mixture of procedures. This methodology has been implemented in the System for the Estimation of Variance due to Nonresponse and Imputation (SEVANI) developed at Statistics Canada; see Beaumont et al. (2010).

Yung and Rao (2000) looked at jackknife variance estimation in the presence of missing values for both the post-stratified and the GREG estimator. They were able to develop consistent jackknife variance estimators under weighting adjustments for unit nonresponse and under weighted mean and hot deck stochastic imputation procedures for item nonresponse. Following Yung and Rao (1996), they also derived corresponding jackknife linearization variance estimators.

2.6 Data Analysis

Up to the 1970s, survey research was mainly focused on the estimation of simple descriptive parameters of a finite population such as population totals, population means and ratios of two population totals. In this context, the quality of estimators of descriptive parameters is often assessed by estimating their sampling variance, as discussed in Section 2.5.

After the seventies, survey analysts became more interested in studying relationships between variables by postulating models, such as regression models which relate a variable of interest y to one or more explanatory variables $x_1, \ldots, x_p$. In this case, the parameters to be estimated are no longer descriptive parameters, but model parameters, sometimes called analytical parameters. In the regression example, these parameters would be the regression coefficients and correlations between variables. Analysts are usually interested in drawing conclusions about the unknown model parameters using sample survey data; i.e., they are interested in making inferences.

Throughout this chapter we have mentioned the variability due to the selection of a sample from the finite population. When making inferences about model parameters, an additional source of variability, namely the variability due to the model, must be considered. Rubin-Bleuer and Schiopu-Kratina (2005) developed a formal theoretical framework for inferences about model parameters in the joint model-design space. If the overall sampling fraction is small, i.e., if the size of the sample is small compared to the size of the population, the model variability can often be neglected; see, e.g., Binder and Roberts (2003). This simplifies variance estimation. For instance, we can use the linearization technique of Binder (1983) or replication techniques that account for the sampling design. However, it is not always appropriate to ignore the model variability. Methods were developed at Statistics Canada to handle this issue. Demnati and Rao (2010) extended their linearization method to account for the model variability while Beaumont and Charest (2012) extended the generalized bootstrap.

Often, survey analysts are interested in testing different types of hypotheses about model parameters. For instance, it may be of interest to test whether education is related to the propensity to leave low-income status. There are numerous procedures that can be used to test hypotheses in classical statistics. For survey data, these procedures need some adaptation to account for the sampling design. An early example of this is Fellegi (1980), who adapted such a procedure. After this, Statistics Canada statisticians, often in collaboration with university researchers, made numerous contributions to the analysis of survey data; see Hidiroglou and Rao (1987), Roberts et al. (1987), and Roberts et al. (2009). Beaumont and Bocci (2009) developed bootstrap tests of hypotheses that can be obtained by replicating a standard test statistic several times.

2.7 Time Series

The information published by Statistics Canada is often in the form of time series, i.e., as a sequence of estimates available for a specific period and ordered in time. Over the years, Statistics Canada has contributed to the development of many methods related to the dissemination and analysis of time series data. Two areas are particularly worth emphasizing: seasonal adjustment and trend-cycle estimation as well as benchmarking and reconciliation methods.

2.7.1 Seasonal Adjustment and Trend-Cycle Estimation Method

Seasonal adjustment is the process used to identify, estimate and, when appropriate, remove seasonal and/or calendar effects from a time series. These are the repetitive patterns that normally occur at the same time and in about the same magnitude every year. For example, construction activity slows down every winter and students seek summer jobs at about the same time every year. These patterns yield little information on the underlying socio-economic situation and make the data more difficult to interpret. As such, most key economic indicators published by national statistical organizations are seasonally adjusted, i.e., the seasonal effects are removed. Some of the most popular statistical methods used to perform seasonal adjustment are based on the X–11 variant of the Census Method II Seasonal Adjustment program. This method was thoroughly reviewed and documented in Ladiray and Quenneville (2001), a book mostly written at Statistics Canada as part of an exchange with the "Institut national de la statistique et des études économiques" (INSEE). The book quickly became an important reference for government agencies, macroeconomists and other users of economic data.

While the core X–11 method itself was developed at the US Census Bureau in the 1950s and 1960s, it underwent many enhancements and extensions over the years. A major one was proposed by Dagum in the mid-1970s and concerns the use of so-called ARIMA time series models in combination with the X–11 method, as described in Dagum (1978). To estimate the seasonal component of a data series at any given point in time, statisticians use moving averages on previous, current and future observations. Because information on future observations is not available, seasonal adjustment can be conducted using only previous and current values, and asymmetric moving averages. Dagum showed that the asymmetric approach yields higher revisions than using a forecasted value with symmetric filters (Dagum, 1982). Under her leadership, the X–11–ARIMA method (Dagum, 1980) was developed at Statistics Canada. Various other enhancements such as extra diagnostics and treatment of calendar effects were included in subsequent years (Dagum, 1988, 2000). The idea of using ARIMA forecasts in combination with the X–11 method is still in use and available today with newer methods (X–12–ARIMA and X–13–ARIMA–Seats). Dagum (1996) also established a method to further extract the economic signal from a seasonally adjusted series. This method helps identify turning points in the economic signal by minimizing unwanted ripples (false turning points) and revisions.

2.7.2 Benchmarking and Reconciliation Methods

When they deal with time series, statistical agencies also face various coherence issues. For example, statistical programs often have two sources of data measuring the same target variable: (i) a more frequent measurement (e.g., monthly) with an emphasis on accurate estimation of the period-to-period movement; and (ii) a less frequent measurement (e.g., annually) with an emphasis on accurate estimation of the level. These two sources are not always perfectly aligned. Benchmarking refers to techniques used to ensure coherence between time series data of the same target variable measured at different frequencies, for example, monthly and annually. It entails imposing the level of the benchmark series while minimizing the revisions to the observed movement in the sub-annual series as much as possible. Statistics Canada has been developing and adapting methods to solve this problem for many years. An important contribution came from Cholette (1984) who modified a well-known procedure from Denton (1971) to better take into consideration assumptions about the first terms of the series. Techniques for related issues such as non-binding benchmarking, interpolation, temporal distribution, calendarization, linkage and reconciliation were developed at Statistics

Canada and documented through the years (Dagum and Cholette, 2006). The methods have also been implemented as SAS procedures in a generalized system initially called Forillon (now G-series) (Latendresse et al., 2006; Bérubé and Fortier, 2009). More recently, research and development has continued on topics such as calendarization — an advanced application of benchmarking techniques used to transform values from a time series observed over varying time intervals into values that cover calendar intervals such as day, week, month, quarter and year (Quenneville et al., 2013).

2.8 Conclusion

Survey statisticians have played an important role in fulfilling Statistics Canada's mandate to provide information to Canadians on the economy and society. In the process, they have made significant contributions to statistical methodology in a wide variety of areas, as outlined in this chapter. As society evolves, and in particular, as technological change continually affects the ways in which a statistical agency can interact with citizens, new challenges present themselves. We can expect statisticians to continue to play an important role in dealing with these challenges. For example, Statistics Canada is in the process of adopting the Internet response option for most of its surveys, following its successful use in the 2011 Census of Population. Giving Canadians this alternative to traditional response modes (telephone, personal visit, paper questionnaire) is both logical and inevitable, but it leads to methodological issues that need to be addressed.

Other examples where statisticians will continue to play an important role include dealing with decreasing response rates, meeting the increasing demand for estimates for small domains by developing methods for producing such estimates, and making more effective use of other sources of data such as tax data and administrative data, all while respecting privacy and confidentiality. It is also likely that massive administrative databases (often referred to as "big data") will start to play a bigger role in the statistical process. Ensuring that such data are used appropriately for official statistics will be another challenge for Statistics Canada's statisticians. With its well-established survey methodology program, Statistics Canada has the strengths needed to meet these challenges. As in the past, we can expect new developments that will benefit not only Statistics Canada, but the statistics community around the world as well.

About the Authors

Jean-François Beaumont is a chief in the Statistical Research and Innovation Division at Statistics Canada. He earned his MSc degree in statistics from Université Laval. His primary research and development activities are imputation of missing data, bootstrap variance estimation for complex sampling designs and estimation issues with survey data. He has been instrumental in the development of SEVANI, the System for the Estimation of Variance due to Nonresponse and Imputation.

Susie Fortier is chief of the Time Series Research and Analysis Centre at Statistics Canada. She received a BSc in statistics and an MSc in mathematics from the Université de Sherbrooke. She teaches various time series courses at Statistics Canada's Training Institute, and her research interests focus on time series reconciliation and benchmarking, seasonal adjustment and price index theory. She is production manager of the journal *Survey Methodology / Techniques d'enquête.*

Jack Gambino is director of the Household Survey Methods Division at Statistics Canada. He obtained an MSc from McGill University and a PhD in Statistics from the University of Toronto in 1982 and has worked at Statistics Canada since then. His primary research interests are in sample survey theory and methods and in Bayesian inference. He is an elected member of the International Statistical Institute.

Mike Hidiroglou is director of the recently created Statistical Research and Innovation Division (SRID) at Statistics Canada where he has worked since 1974, except for a period at the Office for National Statistics in the United Kingdom as Director of Methodology (2003–06). He obtained his PhD in 1974 from Iowa State University. His work at Statistics Canada has been dedicated to the development and implementation of methodology for business surveys, and his research covers a wide spectrum of topics. As director of SRID, he is responsible for the overall coordination of methodology research carried out at Statistics Canada, and the Editorship of Survey Methodology.

Pierre Lavallée is assistant director of the Business Survey Methods Division at Statistics Canada. He completed an MSc in 1988 at Carleton University, and a PhD at the Université libre de Bruxelles. He has worked with business surveys and agricultural surveys, as well as longitudinal social surveys such as the Survey of Labour and Income Dynamics. His published research includes the book *Le sondage indirect, ou la méthode généralisée du partage des poids* (Ellipses, 2002), which later appeared as *Indirect Sampling* (Springer, 2007). He gives courses and presentations regularly on subjects such as recent developments in indirect sampling and hard-to-reach populations.

Bibliography

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

Bankier, M. D. (1988). Power allocations: Determining sample sizes for subnational areas. *The American Statistician*, 42:174–177.

Bankier, M. D., Lachance, M., and Poirier, P. (1999). A generic implementation of the new imputation methodology. In *Proceedings of the Survey Research Methods Section*, pp. 548–553. American Statistical Association, Alexandria, VA.

Bartlett, S., Krewski, D., Wang, Y., and Zielinski, J. M. (1993). Evaluation of error rates in large scale computerized record linkage studies. *Survey Methodology*, 19:3–12.

Beaumont, J.-F. and Bissonnette, J. (2011). Variance estimation under composite imputation: The methodology behind SEVANI. *Survey Methodology*, 37:171–179.

Beaumont, J.-F., Bissonnette, J., and Bocci, C. (2010). *SEVANI, Version 2.3, Methodology Guide.* Technical Report, Statistics Canada, Ottawa, ON.

Beaumont, J.-F. and Bocci, C. (2009). A practical bootstrap method for testing hypotheses from survey data. *Survey Methodology*, 35:25–35.

Beaumont, J.-F. and Charest, A.-S. (2012). Bootstrap variance estimation with survey data when estimating model parameters. *Computational Statistics & Data Analysis*, 56:4450–4461.

Beaumont, J.-F. and Patak, Z. (2012). On the generalized bootstrap for sample surveys with special attention to Poisson sampling. *International Statistical Review*, 80:127–148.

Beaumont, J.-F. and Rivest, L.-P. (2009). Dealing with outliers in survey data. In *Handbook of Statistics, Vol. 29A, Sample Surveys, Design Methods and Applications*, pp. 247–280. Elsevier, Amsterdam.

Bérubé, J. and Fortier, S. (2009). PROC TSRAKING: A user-defined SAS procedure for balancing time series. In *Proceedings of the Business and Economic Statistics Section of the American Statistical Association, Washington, DC.*

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

Binder, D. A. and Roberts, G. R. (2003). Design-based methods for estimating model parameters. In *Analysis of Survey Data*, pp. 29–48. Wiley, Chichester.

Carroll, J. (1970). *Allocation of a Sample between States.* Technical Report, Australian Bureau of Census and Statistics, Canberra.

Cholette, P. A. (1984). Adjusting sub-annual series to yearly benchmarks. *Survey Methodology*, 10:35–49.

Dagum, E. B. (1978). Modelling, forecasting and seasonally adjusting economic time series with the X–11–ARIMA. *Journal of the Royal Statistical Society*, *Series D*, 27:203–216.

Dagum, E. B. (1980). *The X–11–ARIMA Seasonal Adjustment Method.* Technical Report, Statistics Canada, Ottawa, ON.

Dagum, E. B. (1982). The effects of asymmetric filters on seasonal factor revision. *Journal of the American Statistical Association*, 77:732–738.

Dagum, E. B. (1988). *The X–11–ARIMA/88 Seasonal Adjustment Method: Foundations and User's Manual.* Technical Report, Statistics Canada, Ottawa, ON.

Dagum, E. B. (1996). A new method to reduce unwanted ripples and revisions in trend-cycle estimates from X–11–ARIMA. *Survey Methodology*, 22:77–83.

Dagum, E. B. (2000). *The X–11–ARIMA/2000 Seasonal Adjustment Method: Foundations and User's Manual.* Technical Report, Statistics Canada, Ottawa, ON.

Dagum, E. B. and Cholette, P. A. (2006). *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series.* Springer, New York.

Demnati, A. and Rao, J. N. K. (2004). Linearization variance estimators for survey data. *Survey Methodology*, 30:4–13.

Demnati, A. and Rao, J. N. K. (2010). Linearization variance estimators for model parameters from complex survey data. *Survey Methodology*, 36:193–201.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, 66:99–102.

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

Ernst, L. (1989). Weighting issues for longitudinal household and family estimates. In *Panel Surveys*, pp. 135–159. Wiley, New York.

Estevao, V., Hidiroglou, M. A., and Särndal, C.-E. (1995). Methodological principles for a generalized estimation system at Statistics Canada. *Journal of Official Statistics*, 11:181–204.

Estevao, V., Hidiroglou, M. A., and You, Y. (2012). *Small-Area Estimation Unit-Level Model with EBLUP and Pseudo-EBLUP Estimation Methodology Specifications.* Technical Report, Statistics Canada, Ottawa, ON.

Estevao, V., Hidiroglou, M. A., and You, Y. (2013). *Fay–Herriot Area-Level Model with EBLUP Estimation Methodology Specifications.* Technical Report, Statistics Canada, Ottawa, ON.

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

Fellegi, I. P. (1980). Approximate tests of independence and goodness of fit based on stratified multistage samples. *Journal of the American Statistical Association*, 75:261–268.

Fellegi, I. P. (1981). Should the census counts be adjusted for allocation purposes? Equity considerations. In *Current Topics in Survey Sampling*, pp. 47–76. Academic Press, New York.

Fellegi, I. P. and Holt, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, 71:17–35.

Fellegi, I. P. and Sunter, A. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64:1183–1210.

Fuller, W. A. and Rao, J. N. K. (2001). A regression composite estimator with application to the Canadian Labour Force Survey. *Survey Methodology*, 27:45–51.

Gambino, J., Kennedy, B., and Singh, M. P. (2001). Regression composite estimation for the Canadian Labour Force Survey: Evaluation and implementation. *Survey Methodology*, 27:65–74.

Granquist, L. and Kovar, J. G. (1997). Editing of survey data: How much is enough? In *Survey Measurement and Process Quality*, pp. 415–435. Wiley, New York.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11:1–21.

Hartley, H. O. (1962). Multiple frame surveys. In *Proceedings of the Survey Research Methods Section*, pp. 203–206. American Statistical Association, Alexandria, VA.

Haziza, D. (2009). Imputation and inference in the presence of missing data. In *Handbook of Statistics, Sample Surveys: Theory, Methods and Inference*, pp. 215–246. Elsevier, Amsterdam.

Hidiroglou, M. A. (1986). The construction of a self-representing stratum of large units in survey design. *The American Statistician*, 40:27–31.

Hidiroglou, M. A. and Berthelot, J.-M. (1986). Statistical editing and imputation for periodic business surveys. *Survey Methodology*, 12:73–83.

Hidiroglou, M. A. and Rao, J. N. K. (1987). Chi-squared tests with categorical data from complex surveys. *Journal of Official Statistics*, 3:117–140.

Hidiroglou, M. A. and Srinath, K. P. (1981). Some estimators of the population total from simple random samples containing large units. *Journal of American Statistical Association*, 76:690–695.

Kott, P. S. and Stukel, D. M. (1997). Can the jackknife be used with a two-phase sample? *Survey Methodology*, 23:81–89.

Kovar, J. G. and Whitridge, P. (1995). Imputation of business survey data. In *Business Survey Methods*, pp. 403–423. Wiley, New York.

Kozak, R. (2005). The BANFF system for automated editing and imputation. In *Proceedings of the Survey Methods Section*. Statistical Society of Canada, Ottawa, ON.

Ladiray, D. and Quenneville, B. (2001). *Seasonal Adjustment with the X–11 Method*. Springer, New York.

Latendresse, E., Djona, M., and Fortier, S. (2006). *Benchmarking Sub-Annual Series to Annual Totals: From Concepts to SAS Procedure and SAS Enterprise Guider Custom Task*. Technical Report, Paper presented at the 2007 SAS Global Forum, Orlando, FL, April 2007.

Latouche, M. and Berthelot, J.-M. (1992). Use of a score function to prioritize and limit recontacts in editing business surveys. *Journal of Official Statistics*, 8:389–400.

Lavallée, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21:25–32.

Lavallée, P. (2002). *Le sondage indirect, ou la méthode généralisée du partage des poids*. Éditions de l'Université de Bruxelles (Belgique) and Éditions Ellipses (France).

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

Lavallée, P. and Hidiroglou, M. A. (1988). On the stratification of skewed populations. *Survey Methodology*, 14:33–43.

Lee, H. (1995). Outliers in business surveys. In *Business Survey Methods*, pp. 503–521. Wiley, New York.

Lee, H., Rancourt, É., and Särndal, C.-E. (2001). Variance estimation from survey data under single imputation. In *Survey Nonresponse*, pp. 315–328. Wiley, New York.

Lemaitre, G. and Dufour, J. (1987). An integrated method for weighting persons and families. *Survey Methodology*, 13:199–207.

Li, Y., Liu, W., Mathieu, P., Stardom, J., and Yeung, C. (2011). *Sampling for the Unified Enterprise Survey*. Technical Report, Statistics Canada, Ottawa, ON.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, *Series A*, 135:370–384.

Newcome, H. B., Kennedy, J. M., Axford, S. J., and James, A. P. (1959). Automatic linkage of vital records. *Science*, 130:954–959.

Quenneville, B., Picard, F., and Fortier, S. (2013). Calendarization with interpolating splines and state space models. *Journal of the Royal Statistical Society*, *Series C*, 62:371–399.

Rao, C. R. (1971). Some aspects of statistical inference in problems of sampling from finite populations. In *Foundations of Statistical Inference*, pp. 177–202. Holt, Rinehart and Winston, New York.

Rao, J. N. K. (2003). *Small Area Estimation.* Wiley, New York.

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*, 18:209–217.

Roberts, G., Rao, J. N. K., and Kumar, S. (1987). Logistic regression analysis of sample survey data. *Biometrika*, 74:1–12.

Roberts, G., Ren, Q., and Rao, J. N. K. (2009). Using marginal mean models for data from longitudinal surveys with a complex design: Some advances in methods. In *Methodology of Longitudinal Surveys*, pp. 351–366. Wiley, Chichester.

Rubin-Bleuer, S. and Schiopu-Kratina, I. (2005). On the two-phase framework for joint model and design-based inference. *The Annals of Statistics*, 33:2789–2810.

Singh, A. C., Kennedy, B., and Wu, S. (2001). Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design. *Survey Methodology*, 27:33–44.

Statistics Canada (2008). *Methodology of the Canadian Labour Force Survey.* Technical Report, Catalogue Number 71–526, Statistics Canada, Ottawa, ON.

Tambay, J.-L. (1988). Integrated approach for the treatment of outliers in subannual economic surveys. In *Proceedings of the Survey Research Methods Section*, pp. 229–234. American Statistical Association, Alexandria, VA.

Thompson, S. K. (2002). *Sampling,* Second Edition. Wiley, New York.

Thompson, S. K. and Seber, G. A. (1996). *Adaptive Sampling.* Wiley, New York.

Turmelle, C., Godbout, S., and Bosa, K. (2012). Methodological challenges in the development of Statistics Canada's new integrated business statistics program. In *Paper presented at the International Conference on Establishment Surveys IV.* Montréal, QC, June 2012.

You, Y. (2008). An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology*, 34:19–27.

You, Y. and Rao, J. N. K. (2002a). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics*, 30:3–15.

You, Y. and Rao, J. N. K. (2002b). A pseudo empirical best linear unbiased prediction approach to small area estimation using survey weights. *The Canadian Journal of Statistics*, 30:431–439.

Yung, W. and Rao, J. N. K. (1996). Jackknife linearization variance estimators under stratified multi-stage sampling. *Survey Methodology*, 22:23–31.

Yung, W. and Rao, J. N. K. (2000). Jackknife variance estimation under imputation for estimators using poststratification information. *Journal of the American Statistical Association*, 95:903–915.