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

Survey Methodology

Sample survey theory and methods: Past, present, and future directions

by J.N.K. Rao and Wayne A. Fuller

Release date: December 21, 2017

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 STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

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

Standards of service to the public

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

Note of appreciation

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

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
p preliminary
r revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
E use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2017
Vol. 43, No. 2, pp. 145-160
Statistics Canada, Catalogue No. 12-001-X

Sample survey theory and methods: Past, present, and future directions

J.N.K. Rao and Wayne A. Fuller[1]

Abstract

We discuss developments in sample survey theory and methods covering the past 100 years. Neyman's 1934 landmark paper laid the theoretical foundations for the probability sampling approach to inference from survey samples. Classical sampling books by Cochran, Deming, Hansen, Hurwitz and Madow, Sukhatme, and Yates, which appeared in the early 1950s, expanded and elaborated the theory of probability sampling, emphasizing unbiasedness, model free features, and designs that minimize variance for a fixed cost. During the period 1960-1970, theoretical foundations of inference from survey data received attention, with the model-dependent approach generating considerable discussion. Introduction of general purpose statistical software led to the use of such software with survey data, which led to the design of methods specifically for complex survey data. At the same time, weighting methods, such as regression estimation and calibration, became practical and design consistency replaced unbiasedness as the requirement for standard estimators. A bit later, computer-intensive resampling methods also became practical for large scale survey samples. Improved computer power led to more sophisticated imputation for missing data, use of more auxiliary data, some treatment of measurement errors in estimation, and more complex estimation procedures. A notable use of models was in the expanded use of small area estimation. Future directions in research and methods will be influenced by budgets, response rates, timeliness, improved data collection devices, and availability of auxiliary data, some of which will come from "Big Data". Survey taking will be impacted by changing cultural behavior and by a changing physical-technical environment.

Key Words: Data collection; History of survey sampling; Probability sampling; Survey inference.

1 Introduction

This paper was prepared at the invitation of Dr. Danny Pfeffermann, 2015 President of the International Association of Survey Statisticians, who provided the ambitious title. The paper was presented at the meetings of the International Statistical Institute in Rio de Janeiro, Brazil in 2015.

The title defines an area too large for us to address in a single paper. Furthermore, there are a number of review papers that address the topics of the title, including Kish (1995), Bellhouse (2000), Rao (2005), Bethlehem (2009), Brick (2011), Groves (2011), and Brewer (2013). Our discussion draws on those papers, but we do not attempt completeness. We provide a brief appraisal of the three topics and project a number of current situations into the future. Our aim is to stimulate further discussion, especially on the future directions. Beyond the discussion of controversies related to purposive sampling, we will concentrate on probability-based sampling. Because survey sampling is an applied field, some of the problems encountered and methods employed in practice will be addressed. Our discussion is most relevant for large general purpose samples, the surveys where we have the most experience. Likewise, our knowledge of applications is concentrated in Canada and the United States.

The paper is organized as follow. Section 2 presents the early landmark contributions from 1920-1960. Inferential issues are covered in Section 3. The paper concludes with a discussion on the future in Section 4.

1. J.N.K. Rao is Distinguished Research Professor, Carleton University, Ottawa, Canada, K15-5B6. E-mail: jrao@math.carleton.ca; Wayne A. Fuller is Distinguished Professor Emeritus, Iowa State University, Ames, IA, USA. 50011. E-mail: waf@iastate.edu.

2 Early landmark contributions: 1920-1960

Kiaer (1897) is perhaps the first to promote sampling (or what was then called the representative method) over complete enumeration (census), although the oldest reference can be traced back to 1000 BC. In the representative method, the objective is for the sample to mirror the parent finite population and this may be achieved either by balanced sampling on known auxiliary totals, through purposive selection or by random sampling leading to equal inclusion probabilities. By the 1920s the representative method was widely used. The International Statistical Institute (ISI) played a vital role by creating an expert committee to report on this method. Bowley's (1926) contribution to the ISI report includes his fundamental work on stratified random sampling with proportional allocation, leading to equal inclusion probabilities. Bowley (1936) states that the "first application of this principle" of inferring the population from the sample was the 1912 study in Reading. Bowley specified the sampling procedure for that study as a systematic sample from a list of houses. Bowley called the systematic procedure a "pure method of sampling" and stated, "This is literally the method of stratified sampling". Bowley gives a number of examples where systematic sampling was used after 1912. Bowley (1936) emphasized the importance of a complete frame and equal probabilities of selection. But it was Neyman (1934) who laid the foundations of probability sampling (or design-based approach). He demonstrated that stratified random sampling is preferable to balanced (representative) sampling as it was used then. He also introduced the concept of efficiency and optimal sample allocation, now called Neyman allocation, that minimizes the total size of the sample for a specified precision by relaxing Bowley's condition of equal inclusion probabilities. In fact, Tchuprow (1923) derived the Neyman allocation ten years earlier, in a paper discovered after the Neyman paper appeared. Neyman (1934) also showed that for large enough samples one could obtain confidence intervals on the population mean of a variable of interest such that the frequency of errors in the confidence statement in repeated sampling does not exceed the limit prescribed in advance, "whatever the unknown properties of the population". In recent years, balanced sampling, originally advocated by Gini and Galvani, has been refined to incorporate the nice features of both probability sampling and balanced sampling on known auxiliary totals (Deville and Tillé, 2004). The new balanced sampling method is now used in Europe, especially in France, to select samples for establishment surveys. A second method of probability controlled selection is rejective sampling, introduced by Hájek (1964) as a method for controlling the sample size in Poisson sampling. Fuller (2009a) extended the procedure to restrict acceptable samples to the set where estimates of the means of auxiliary variables are close to the population mean.

The 1930s witnessed a rapid growth in demand for socio-economic information, and the advantages of probability sampling in terms of greater scope, reduced cost, and greater speed relative to censuses, were soon recognized worldwide. This led to an increase in number and type of surveys based on probability sampling and covering large populations. Neyman's probability sampling (or design-based approach) was almost universally accepted and it became a standard tool for empirical research in social sciences and official statistics. It was also recognized that the precision of an estimator is determined largely by the sample size and not by the sampling fraction. The 1940's saw a number of studies on the properties of systematic sampling for different populations. See Madow and Madow (1944), Cochran (1946), and Yates (1948). Cochran (1977, Chapter 8) is an excellent discussion of systematic sampling, a discussion that makes clear why only model-based estimators of variance are possible. Also see Bellhouse (1988). In the early development of sampling theory, focus was on estimating totals and means and associated sampling errors. Non-sampling errors such as nonresponse, coverage errors, and measurement errors, were largely ignored in theoretical research.

We now list a few important post-Neyman theoretical developments in the design-based approach. Mahalanobis used multi-stage sampling designs for crop surveys in India as early as 1937. His classic 1944 paper (Mahalanobis, 1944) rigorously formulated cost and variance functions for the efficient design of surveys. He was instrumental in creating the National Sample Survey of India, the largest multi-subject continuing survey with full-time staff using personal interviews for socio-economic surveys and physical measurements for crop surveys. Sukhatme, who studied under Neyman, also made pioneering contributions to the design and analysis of large scale agricultural surveys in India, using stratified multi-stage sampling. Classic text books on sampling by Cochran (1953), Deming (1950), Hansen, Hurwitz and Madow (1953), Sukhatme (1954) and Yates (1949) benefited students as well as practitioners.

Survey statisticians at the U.S. Census Bureau, under the leadership of Morris Hansen, made fundamental contributions to sample survey theory and methodology, during the period 1940-1960. This period is regarded as the golden era of the Census Bureau. Hansen and Hurwitz (1943) developed the basic theory of stratified two-stage cluster sampling with one cluster (or primary sampling unit) within each stratum drawn with probability proportional to size (PPS) and then subsampled at a rate to ensure a self-weighting sample (equal overall probabilities of selection). Unequal probability selection of clusters can lead to significant variance reduction by controlling the variability arising from unequal cluster sizes. Another major contribution from the U.S. Census Bureau is the use of rotation sampling with partial replacement of households to handle response burden in surveys repeated over time, such as the monthly U.S. Current Population Survey for measuring unemployment rates. Hansen, Hurwitz, Nisselson and Steinberg (1955) developed simple but efficient composite estimators under rotation sampling. Rotation sampling and composite estimation are widely used in large-scale continuing surveys.

Prior to the 1950s, the primary focus was on estimating population totals and means. Woodruff (1952) of the U.S. Census Bureau developed a unified approach for constructing confidence intervals for quantiles (in particular, the median), applicable to general sampling designs. The procedure remains a cornerstone for quantile estimation (Francisco and Fuller, 1991).

After the consolidation of the basic design-based sampling theory, Hansen, Hurwitz, Marks and Mauldin (1951) and others paid attention to measurement or response errors in survey data. Under additive measurement error models with minimal model assumptions on the observed responses treated as random variables, total variance of an estimator can be decomposed into sampling variance, simple response variance and correlated response variance (CRV) due to interviewers.

Mahalanobis (1946) had developed the method of interpenetrating subsamples for assessing both sampling and interviewer errors. By assigning the subsamples at random to interviewers, both the total variance and the interviewer component can be estimated. The interviewer component can dominate total variance when the number of interviewers is small. To remove the CRV component due to interviewers, self-enumeration by mail was introduced in the 1960 U.S. Census.

Nonresponse in surveys was also addressed in early survey sampling development. Hansen and Hurwitz (1946) proposed two-phase sampling in which the sample is contacted by mail in the first phase and a subsample of nonrespondents is then subjected to personal interview, assuming complete response or negligible nonresponse at the second phase. This method was used recently in Canada when the compulsory long form sample census was replaced by a voluntary National Household Survey. After the change of Government in 2015, the Prime Minister of Canada reinstated the long form census. Two phase sampling is retained but to a lesser extent. The Hansen-Hurwitz two phase sampling method has also been used in other surveys including the American Community Survey.

Attention was also given to inferences for unplanned subpopulations (called domains) such as age-sex groups within a state. Hartley (1959) and Durbin (1958) developed a unified theory for domain estimation applicable to general designs and yet requiring only existing formulae for population totals and means.

Most of the survey sampling theory in the early period was developed by official statisticians while academic researchers, especially in USA, paid little attention to survey sampling. An exception was Iowa State University, where faculty played a leading role from the early stages under the leadership of Cochran, Jessen and Hartley. Another institution making early contribution to survey practice and research is the Survey Research Center at the University of Michigan established in 1947, with Leslie Kish as one of its first members.

In the 1950s formal theoretical frameworks for design-based inference on totals and means were proposed by regarding the sample data as a set of sample labels together with the associated variables of interest. Horvitz and Thompson (1952) derived the well-known estimator with weight inversely proportional to the inclusion probability. Narain (1951) also proposed this estimator. Godambe (1955) developed a general class of linear estimators by letting the sample weight of a unit depend on the label as well as on the labels of the other units in the sample. He then showed that the best linear unbiased estimator does not exist in this general class even under simple random sampling.

3 Inferential issues: 1950 -

3.1 Theoretical foundations

Attempts were made to integrate sample survey theory with mainstream statistical inference via the likelihood function. Godambe (1966) showed that the likelihood function from the full sample data including labels, regarding the vector of unknown population values as the parameter, provides no information on the non-sampled values and hence on the population total or mean. This uninformative feature of the likelihood function is due to the inclusion of labels in the data which makes the sample unique. An alternative design-based route ignores some aspects of the sample data to make the sample non-unique and thus arrive at informative likelihood functions (Hartley and Rao, 1968; Royall, 1968). This non-parametric likelihood approach is similar to the currently popular empirical likelihood (EL) approach in mainstream statistical inference (Owen, 1988). The EL approach has been applied to sampling problems in recent years to estimate not only totals and means but also more complex parameters. So the integration efforts with main stream statistics was partially successful.

The model-dependent approach provides an alternative route to inference from survey data. The approach requires that the population structure obeys a specified super-population model. The distribution induced by the assumed model provides the basis for inferences (Brewer, 1963 and Royall, 1970). Such conditional (conditional on the sample) inferences can be appealing. However, the resulting estimators may be design inconsistent and, as such, they can perform poorly in large samples under model misspecification (Hansen, Madow and Tepping, 1983).

A hybrid approach, called the model-assisted approach, attempts to combine the desirable features of the design-based and model-dependent methods, see Cassel, Särndal and Wretman (1976). The approach typically includes the use of data external to the collected data, called auxiliary data. Procedures using auxiliary data include regression estimation, ratio estimation, and raking, methods with estimators linear in the variable of interest. Estimators using auxiliary information, particularly regression, were recognized very early as powerful estimators (Cochran, 1953). Computing power made regression estimation practical in the 1970's, but to be acceptable in large scale surveys the regression weights need to be nonnegative. An early definition of nonnegative weights is Huang and Fuller (1978). Deville and Särndal (1992) gave a general method of constructing weights for design consistent estimators. Model assisted methods entertain only design consistent estimators of the total that are also model unbiased under a working model. This approach is useful for large samples and it leads to valid design-based inferences in large samples, regardless of the validity of the working model. However, efficiency of the estimators does depend on the degree to which the working model approximates the true population structure. The most popular form of model-assisted estimators are known as generalized regression estimators (GREGs) and are implemented in survey software packages.

Theoretical results for probability-based sampling emphasize the first two moments of the sample statistics. Central limit theorems have been used to provide justification for normality-based confidence intervals. An early central limit theorem for simple random samples is that of Madow (1948). Hájek (1960) gave a central limit theorem for simple random sampling and a theorem for rejective sampling in Hájek (1964). Bickel and Freedman (1984) gave a central limit theorem for stratified random sampling. Recent literature considers both sequences of fixed finite populations and sequences of finite populations that are samples from a superpopulation (Fuller, 2009b; Section 1.3.2).

Variance estimation was very costly, nearly prohibitive, in the 1930's and 1940's, and remains expensive today. Replication was adopted as an efficient variance estimation method from the beginning. As we noted, an early replication form was introduced by Mahalanobis (1939, 1946) called "interpenetrating" samples by him, and called "random groups" by later authors. The method of random groups based on half samples, was used by the U.S. Census Bureau in the 1950's and 1960's. McCarthy (1966, 1969) developed and described balanced half-sample variance estimation. Also see Kish and Frankel (1970). Wolter (2007) contains an extensive discussion on balanced half samples. Also see Dippo, Fay and Morgenstein (1984), Kish and Frankel (1974), Krewski and Rao (1981), and Rao and Shao (1999). The jackknife and bootstrap are the current versions of early replication procedures. Wolter (2007, Chapter 4) credits Durbin (1959) with the first use of the jackknife in finite population estimation. The use of the bootstrap in the classical setting dates from Efron (1979) but application to unequal probability samples and finite populations is not immediate. Among the large number of papers on jackknife and bootstrap for survey samples are McCarthy and Snowden (1985), an early with-replacement sampling version, and Rao and Wu (1988), a modified bootstrap based on “rescaling” for survey samples. Sitter (1992) discussed several topics including suggestions for obtaining integer sample sizes. Antal and Tillé (2011) gave bootstrap methods appropriate for a wide range of designs, including Poisson sampling. Beaumont and Patak (2012) gave general bootstrap procedures.

3.2 Analytic use of survey data

As we have remarked, the early work on probability sampling emphasized totals and means and many estimation procedures were developed for official statistics. However, from the beginning, survey samples were used by social scientists to answer subject matter questions with relevance beyond the finite population sampled. Deming and Stephan (1940) and Deming (1953) gave explicit consideration to the difference between “enumerative” and “analytic” use of survey and census data, also see Hartley (1959). The analytic estimates are sometimes called estimates for a superpopulation. Early analysts often treated survey sample data as a simple random sample and constructed estimates on that basis. The potential for bias that arises from ignoring the design led to estimation theory for analytic estimates. One component is comprised of tests for the effect of weights on estimates, see DuMouchel and Duncan (1983), Fuller (1984), and Korn and Graubard (1995). A second component has been the development of design based theory for complicated statistics. See Fuller (1975), Rao and Scott (1981, 1984), and Binder and Roberts (2003). The third approach builds the sampling design into the model (Skinner, 1994 and Pfeffermann and Sverchkov, 1999). A number of computer packages (SAS, SUDAAN, R, STATA) are now available for probability-based statistics and standard errors. Many of the algorithms date from the work at Iowa State University (Hidiroglou, Fuller and Hickman, 1976).

3.3 Missing data

Almost all samples (and experiments) have missing and incorrect data. Missing data in survey sampling are placed in two categories; unit-missing and item-missing, where, as the name implies, a missing unit means that all items in the response record are missing. An indicator of the importance of missing data in survey research is the monograph set edited by Madow, Nisselson and Olkin (1983). One method of handling missing data is to report the nature and number of missing items and tabulate the remaining items. This was common in the early years, but the implied assumption of exchangeability in such a procedure was often not reasonable. An early method of correcting for unit nonresponse was to use a substitute respondent, often interviewing someone “close” to the nonrespondent. A common modification at the analysis stage was, and remains, post stratification. (Deming, 1953; Thomsen, 1973; Kalton, 1983 and Jagers, 1986). In the missing data literature, post strata are often called cells. Regression estimators are direct extensions of cell estimators and are an important method of correcting for missing data (Fuller and An, 1998). Weighting methods for handling unit nonresponse are reviewed in Brick and Montaquila (2009).

Various forms of imputation for item nonresponse have been used over time, with imputation performed by clerks prior to use of computers. An early formal model-based and computer-based imputation was the hot deck imputation procedure used by the U.S. Census Bureau in the 1947 Current Population Survey, see the description in Andridge and Little (2009). Improved computing power and theoretical advances (Little, 1982; Kalton and Kish, 1984; Rubin, 1974, 1976, 1987; Little and Rubin, 1987; Kim and Fuller, 2004) have made imputation a standard part of estimation for survey samples and an active area of research. Recent books are Kim and Shao (2013) and Little and Rubin (2014).

3.4 Small area estimation

The increased use of models for small-domain estimates is the result of the pairing of two factors. The first is the demand for estimates for small domains (e.g. geographic areas) in policy formulation, fund allocation and regional planning. The second is the large standard errors for many of the design-based domain estimators. Schaible (1996) and Purcell and Kish (1979) gave early examples of small area estimation, also see Gonzalez (1973) and Steinberg (1979). The U.S. Census Bureau used model-based methods for small area estimation as early as 1947 (Hansen et al., 1953; Vol. I, pages 483-486). More recently, linear mixed models involving both fixed and random effects have become important. Early uses of mixed models for small area estimation are Fay and Herriot (1979) and Battese, Harter and Fuller (1988). Some sets of small area estimates can be viewed as a reallocation of the domain estimates, retaining the direct design-consistent estimate of the grand total. Bayesian methods, in particular hierarchical Bayes, are increasingly being used because of the ability to handle complex models; see Rao and Molina (2015, Chapter 10). On the basis of growing demand, there has been a large increase in literature and the field now boasts regular meetings and a book (Rao, 2003) with a recent second edition, (Rao and Molina, 2015).

3.5 Survey practice

Sample design and estimation topics that we have discussed are critical parts of a survey operation, but represent a small fraction of the total. The quality of the final product is determined by frame materials, collection instrument, data collection, editing, processing, and presentation of results. Many error sources are difficult to measure, but those designing surveys make implicit cost estimates when they allocate resources to different parts of the survey operation. Groves and Lyberg (2010) is a review of attempts to enumerate the components of survey quality and to bring them under a single umbrella. They credit Deming (1944) for an early description of error sources in sample surveys and describe the contributions of Dalenius (1974), Anderson, Kasper and Frankel (1979), Groves (1989), Biemer and Lyerg (2003), among others. Groves and Herringa (2006) proposed tools for actively controlling survey errors and costs that can lead to responsive designs for household surveys. In particular, para data (measurements related to the process of collecting survey data) can be used to monitor field work, to make intervention decisions during data collection and to deal with measurement error, nonresponse and coverage errors (Kreuter, 2013).

4 The future

We can project a number of current situations into the future. Budgets will be tight and requests for products will expand. There will be demand for forecasts, and for improved access by users. There will be requests for statistics to be produced more rapidly and, naturally, with no compromise in quality. There will be pressure to bring estimates from different sources into agreement.

We expect faster computing to influence all aspects of the field. More complex edit and imputation algorithms will be developed. The time from collection to publication will be shortened. More complex analyses will be performed on survey data. Record linkage procedures will be improved. Data will be made available in different forms. Searchable databases where the user provides queries will become more common. The use of auxiliary data of all kinds, and in particular administrative data, will increase. Administrative data will be used both as auxiliary data and as the direct estimates for certain items. Citro (2014) gives examples of items where administrative data can be used to replace answers to questions in a questionnaire. Uses of auxiliary data where matching to collected data is imperfect will be a research area.

Modern communication methods and social media have resulted in vast quantities of data, much generated with short term and poorly identified purpose. The term “Big Data” is not well defined, but most would agree that social media data are a part of Big Data. The AAPOR report on Big Data (2015) is an excellent analysis of the potential and the challenges associated with Big Data. Tam and Clarke (2015) and Pfeffermann (2015) discuss the issues from the perspective of a governmental statistical organization. As part of modern society, social media are of interest to social scientists in their own right. Therefore, indexes and summaries of these data are, and will be, produced. An example is the University of Michigan Social Media Job Loss Index. Sampling has a large role to play in the creation of products from these data.

A challenge is transforming some types of Big Data into a form useful as auxiliary data. One example is the Porter, Holan, Wikle and Cressie (2014) use of Google trends of Spanish words as functional covariates to estimate state proportions of people speaking Spanish using American Community Survey estimates as dependent variables in small area models.

One of the often quoted advantages of samples relative to censuses is cost. The cost structure has changed with increased computing power and seems destined to continue to change. In the United States, the National Land Cover Database is a census of land cover (Han, Yang, Di and Mueller, 2012). Classification procedures are expected to improve so that use of such data as auxiliary data will increase. Data collection agencies will invest more in constructing improved auxiliary data files at the population level so that some data now collected on a sample basis will be collected at a population level. The same types of data development will continue for population and business statistics.

Of necessity, our discussion has little on collection. The way in which data collection procedures have been modified with changing technology is perhaps more obvious than the link between technology and theory. For the links to theory see Bellhouse (2000). Computer-assisted data collection is the evolving standard. The use of geo-location technology can be expected to increase. It is safe to forecast the increased use of remote sensing and remote data collection devices. For example, it would be easy to incorporate physical data collected by something like the Apple Watch or Fitbit into a health study. Larger and less attractive monitoring devices are currently in use in physical activity surveys (van Remoortel, Giavedoni, Raste, Burtin, Louvaris, Gimeno-Santos, Langer, Glendenning, Hopkinson, Vogiatzis, Peterson, Wilson, Mann, Rabinovich, Puhan, Troosters and PROactive consortium, 2012).

The recent experience is that phone and personal interview data collection is becoming more and more difficult. Respondents are facing expanded organized data collection activities. The ubiquitous questionnaire on satisfaction for everything from medical services to tooth paste surely must impact an individual’s willingness to respond. It seems reasonable to forecast increased difficulty in obtaining cooperation for traditional methods of data collection. Associated with that trend will be increased study of the nature of non-respondents and of non-response. Likewise efforts will be made to adapt data collection to the changing methods of communication.

Nonprobability samples have been a part of survey activity throughout the post-Neyman period. In particular, quota sampling is commonly used in marketing research and other areas for cost reasons (Sudman, 1966; 1976). Moser and Stuart (1953) and Stephan and McCarthy (1958) made early comparisons between quota sampling and probability sampling. Cochran (1977, page 136) says “The quota method seems likely to produce samples that are biased on characteristics such as income, education and occupation, although it often agrees with the probability samples on questions of opinion and attitude”. Use of procedures such as post stratification and regression estimation in nonprobability samples has continued at pace with use in probability samples. The changing nature of human communication offers opportunities for both model-based and probability-based procedures. Because of cost structures, new methods such as web-based procedures will often be used first in nonprobability settings and for nongovernmental purposes.

As matching procedures improve and as demand for detailed data increases, disclosure limitation procedures and associated research will receive increased attention.

Survey sampling is an application discipline, functioning in the current social, geographic, cultural, and technological world. To forecast how our field will be impacted by social and cultural changes, even in the short run, is a challenge. Will the fact that one must assume that almost all of one’s public activity and a great deal of one’s private activity has potential of being recorded lead to a more relaxed attitude in responding to questions? Will improved monitoring devices make respondents more willing to permit their physical activities be monitored? Or will all of the incidental monitoring lead to a reaction against organized data collection? Will increased availability of results based on collected data have a positive or negative effect on data collection efforts? What is the impact of various Social Media?

This discussion makes clear that factors external to our discipline will determine our future activities. We will be required to adapt in data collection, data processing, and data presentation-dissemination.

Acknowledgements

We thank Graham Kalton for comments and suggestions that led to improvements in the original draft. We thank the four discussants, Graham Kalton, Sharon Lohr, Danny Pfeffermann and Chris Skinner, for their supplements on the history, insightful observations on the present, and comments on the future of survey sampling. We elected not to prepare a rejoinder because we found much to appreciate and little basis for disagreement.

References

AAPOR Big Data Task Force (2015). *AAPOR Report on Big Data*. https://www.aapor.org/AAPOR_Main/media/MainSiteFiles/images/BigDataTaskForceReport_FINAL_2_12_15_b.pdf.

Anderson, R., Kasper, J. and Frankel, M. (1979). *Total Survey Error: Applications to Improve Health Surveys*. San Francisco, CA: Jossey-Bass.

Andridge, R.H., and Little, R.J. (2009). The use of sample weights in hot deck imputation. *Journal of Official Statistics,* 25, 21-36.

Antal, E., and Tillé, Y. (2011). A direct bootstrap method for complex sampling designs from a finite population. *Journal of the American Statistical Association,* 106, 534-543.

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

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

Bellhouse, D.R. (1988). Systematic sampling. In *Handbook of Statistics*, (Eds., P.R. Kreshnaiah and C.R. Rao), Elsevier, 6, 125-145.

Bellhouse, D.R. (2000). Survey sampling theory over the twentieth century and its relation to computing technology. *Survey Methodology,* 26, 1, 11-20. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2000001/article/5174-eng.pdf.

Bethlehem, J. (2009). The rise of survey sampling. Discussion Paper (09015), Statistics Netherlands, The Hague.

Bickel, P.J., and Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics,* 12, 470-482.

Biemer, P.P., and Lyberg, L. (2003). *Introduction to Survey Quality*. New York: John Wiley & Sons, Inc.

Binder, D.A., and Roberts, G.A. (2003). Design-based and model-based methods for estimating model parameters. In *Analysis of Survey Data*, (Eds., R.L. Chambers and C.J. Skinner), Wiley, Chichester, UK, 29-48.

Bowley, A.L. (1926). Measurement of the precision attained in sampling. *Bulletin of the International Statistical Institute,* 22(1), 1-62.

Bowley, A.L. (1936). The application of sampling to economics and sociological problems. *Journal of the American Statistical Association,* 31, 474-480.

Brewer, K.R.W. (1963). Ratio estimation and finite populations: Some results deducible from the assumption of an underlying stochastic process. *Australian Journal of Statistics,* 5, 93-105.

Brewer, K.R.W. (2013). Three controversies in the history of survey sampling. *Survey Methodology,* 39, 2, 249-262. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2013002/article/11883-eng.pdf.

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

Brick, M.J., and Montaquila, J.M. (2009). Non response and weights. In *Handbook of Statistics,* (Eds., D. Pfeffermann and C.R. Rao), Elsevier, Amsterdam, 29A, 163-185.

Cassel, C.M., Särndal, C.-E. and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. *Biometrika*, 63, 615-620.

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

Cochran, W.G. (1953). *Sampling Techniques.* New York: John Wiley & Sons, Inc.

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

Dalenius, T. (1974). Ends and Means of Total Survey Design. Report in "Errors in Surveys", Stockholm University.

Deming, E. (1944). On errors in surveys. *American Sociological Review*, 9, 359-369.

Deming, E. (1950). *Some Theory of Sampling*. New York: John Wiley & Sons, Inc.

Deming, W.E. (1953). On a probability mechanism to attain an economic balance between the resultant error of non-response and the bias of non-response. *Journal of the American Statistical Association*, 48, 743-772.

Deming, W.E., and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 4, 427-444.

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

Deville, J.-C., and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91, 893-912.

Dippo, C.S., Fay, R.E. and Morgenstein, D.H. (1984). Computing variances from complex samples with replicate weights. *Proceedings of the American Statistical Association*, *Section on Survey Research Methods*, 489-494.

DuMouchel, W.H., and Duncan, G.J. (1983). Using sample survey weights in multiple regression analysis of stratified samples. *Journal of the American Statistical Association*, 78, 535-543.

Durbin, J. (1958). Sampling theory for estimates based on fewer individuals than the number selected. *Bulletin of the International Statistical Institute*, 36, 113-119.

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

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.

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, 366, 269-277.

Francisco, C.A., and Fuller, W.A. (1991). Estimation of quantiles with survey data. *Annals of Statistics*, 19, 454-469.

Fuller, W.A. (1975). Regression analysis sample survey. *Sankhyā, Series C,* 37, 117-132.

Fuller, W.A. (1984). Least squares and related analyses for complex survey designs. *Survey Methodology*, 10, 1, 97-118. Paper available at http://www.statcan.gc.ca/pub/12-001-x/1984001/article/14352-eng.pdf.

Fuller, W.A. (2009a). Some design properties of a rejective sampling procedure. *Biometrika*, 96, 1-12.

Fuller, W.A. (2009b). *Sampling Statistics*. New York: John Wiley & Sons, Inc.

Fuller, W.A., and An, A.B. (1998). Regression adjustments for nonresponse. *Journal of Indian Society of Agricultural Statistics*, 51, 331-342.

Godambe, V.P. (1955). A unified theory of sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 17, 269-278.

Godambe, V.P. (1966). A new approach to sampling from finite populations. *Journal of the Royal Statistical Society, Series B*, 28, 310-328.

Gonzalez, M.E. (1973). Use and evaluation of synthetic estimates. *Proceedings of the Social Statistics Section of the American Statistical Association*, 33-36.

Groves, R.M. (1989). *Survey Errors and Survey Costs*. New York: John Wiley & Sons, Inc.

Groves, R.M. (2011). Three eras of survey research. *Public Opinion Quarterly*, 73, 861-871.

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

Groves, R.M., and Lyberg, L. (2010). Total survey error: Past, present and future. *Public Opinion Quarterly*, 74, 849-879.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publications of the Mathematical Institute of the Hungarian Academy*, 5, 361-374.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistics*, 35, 1491-1523.

Han, W., Yang, Z., Di, L. and Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture*, 84, 111-123.

Hansen, M.H., and Hurwitz, W.N. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-362.

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

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

Hansen, M.H., Madow, W.G. and Tepping, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78, 776-793.

Hansen, M.H., Hurwitz, W.N., Marks, E.S. and Mauldin, W.P. (1951). Response errors in surveys. *Journal of the American Statistical Association*, 46, 147-190.

Hansen, M.H., Hurwitz, W.N., Nisselson, H. and Steinberg, J. (1955). The redesign of the census current population survey. *Journal of the American Statistical Association*, 50, 701-719.

Hartley, H.O. (1959). Analytical studies of survey data. In Volume in Honor of Corrado Gini, Instituto di Statistica, Rome, 1-32.

Hartley, H.O., and Rao, J.N.K. (1968). A new estimation theory for sample surveys. *Biometrika*, 55, 547-557.

Hidiroglou, M.A., Fuller, W.A. and Hickman, R.D. (1976). *SUPER CARP* Statistical Laboratory, Survey Section, Iowa State University, Ames, IA.

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

Huang, E.T., and Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section*, American Statistical Association, 300-305.

Jagers, P. (1986). Post-stratification against bias in sampling. *International Statistical Review/Revue Internationale de Statistique*, 54, 159-167.

Kalton, G. (1983). *Compensating for Missing Survey Data*. Survey Research Center, University of Michigan, Ann Arbor, Michigan.

Kalton, G., and Kish, L. (1984). Some efficient random imputation methods. *Communications in Statistics*, A13, 1919-1939.

Kiaer, A. (1897). The representative method of statistical surveys (1976 English translation of the original Norwegian), Oslo, Central Bureau of Statistics of Norway.

Kim, J.K., and Fuller, W.A. (2004). Fractional hot deck imputation. *Biometrika*, 91, 559-578.

Kim, J.K., and Shao, J. (2013). *Statistical Methods for Handling Incomplete Data*. CRC Press, Boca Raton, FL.

Kish, L. (1995). The hundred years' wars of survey sampling. *Statistics in Transition*, 2, 813-830.

Kish, L., and Frankel, M.R. (1970). Balanced repeated replications for standard errors. *Journal of the American Statistical Association*, 65, 1071-1094.

Kish, L., and Frankel, M.R. (1974). Inference from complex samples (with discussion). *Journal of the Royal Statistical Society, Series B*, 36, 1-37.

Korn, E.L., and Graubard, B.I. (1995). Analysis of large health surveys: Accounting for the sampling designs. *Journal of the Royal Statistical Society, Series A*, 158, 263-295.

Kreuter, F. (2013). *Improving Surveys with Paradata*. Hoboken: Wiley.

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

Little, R.J.A. (1982). Models for nonresponse in sample surveys. *Journal of the American Statistical Association*, 77, 237-250.

Little, R.J.A., and Rubin, D.B. (1987, 2014). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc., (Second Edition 2014).

Madow, W.G. (1948). On the limiting distribution of estimates based on samples from finite universes. *Annals of Mathematical Statistics*, 19, 535-545.

Madow, W.G., and Madow, L.H. (1944). On the theory of systematic sampling, I. *The Annals of Mathematical Statistics*, 15, 1-24.

Madow, W.G., Nisselson, H. and Olkin, I. (Eds.) (1983). *Incomplete Data in Sample Surveys*, 1, 2, 3, New York: Academic Press.

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

Mahalanobis, P.C. (1944). On large-scale sample surveys. *Philosophical Transactions of the Royal Society of London, Series B*, 231, 329-451.

Mahalanobis, P.C. (1946). Recent experiments in statistical sampling in the Indian Statistical Institute. *Journal of the Royal Statistical Society*, 109, 325-378.

McCarthy, P.J. (1966). Replication: An approach to the analysis of data from complex surveys. *Vital and Health Statisitcs*, Series 2, No. 14, National Center for Health Statistics, Public Health Service, Washington DC.

McCarthy, P.J. (1969). Pseudoreplication: Further evaluation and application of the balanced half-sample technique. *Vital and Health Statistics*, Series 2, No. 31, National Center for Health Statistics, Public Health Service, Washington, DC.

McCarthy, P.J. (1969). Pseudoreplication: Half-samples. *International Statistical Review/Revue Internationale de Statistique*, 37, 239-264.

McCarthy, P.J., and Snowden, L.B. (1985). The bootstrap and finite population sampling. *Vital Health Statistics*, 2-95, *Public Health Service Publication*, 85-1369, U.S. Government Printing Office, Washington, D.C.

Narain, R.D. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 3, 169-174.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-625.

Owen, A.B. (1988). Empirical likelihood ration confidence intervals for a single functional. *Biometrika*, 75, 237-249.

Pfeffermann, D. (2015). Methodological issues and challenges in the production of official statistics. *Journal of Survey Statistics and Methodology*, 3, 425-483.

Pfeffermann, D., and Sverchkov, M. (1999). Parametric and semiparametric estimation of regression models fitted to survey data. *Sankhyā, Series B,* 61, 166-186.

Porter, A.T., Holan, S.H., Wikle, C.K. and Cressie, N. (2014). Spatial Fay-Herriot model for small area estimation with functional covariates. *Spatial Statistics,* 10, 27-42.

Purcell, N., and Kish, L. (1979). Estimation for small domains. *Biometrics,* 35, 365-384.

Rao, J.N.K. (2003). *Small Area Estimation.* Hoboken: Wiley.

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

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

Rao, J.N.K., and Scott, A.J. (1981). The analysis of categorical data from complex sample surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association,* 76, 221-230.

Rao, J.N.K., and Scott, A.J. (1984). On chi-squared tests for multiway contingency tables with cell proportions estimated from survey data. *Annals of Statistics,* 12, 46-60.

Rao, J.N.K., and Shao, J. (1999). Modified balanced repeated replication for complex survey data. *Biometrika,* 86, 403-415.

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.

Royall, R.M. (1968). An old approach to finite population sampling. *Journal of the American Statistical Association,* 63, 1269-1279.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika,* 57, 377-387.

Rubin, D.B. (1974). Characterizing the estimation of parameters in incomplete data problems. *Journal of the American Statistical Association,* 69, 467-474.

Rubin, D.B. (1976). Inference and missing data. *Biometrika,* 63, 581-590.

Rubin, D.B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley & Sons, Inc.

Schaible, W.L. (Ed.) (1996). *Indirect Estimators in U.S. Federal Programs.* New York: Springer.

Sitter, R.R. (1992). A resampling procedure for complex survey data. *Journal of the American Statistical Association,* 87, 755-765.

Skinner, C.J. (1994). Sample models and weights. In *Proceedings of the Survey Research Methods Section,* American Statistical Association, 133-142.

Steinberg, J. (Ed.) (1979). *Synthetic Estimates for Small Areas: Statistical Workshop Papers and Discussion.* NIDA Research Monograph No. 24. U.S. Government Printing Office, Washington, D.C., U.S.A.

Stephan, F., and McCarthy, P.J. (1958). *Sampling Opinions*. New York: John Wiley & Sons, Inc.

Sudman, S. (1966). Probability sampling with quotas. *Journal of the American Statistical Association*, 61, 749-791.

Sudman, S. (1976). *Applied Sampling.* New York: Academic Press.

Sukhatme, P.V. (1954). *Sampling Theory of Surveys with Applications*. Iowa State College Press, Ames.

Tam, S.-M., and Clarke, F. (2015). Big data, official statistics and some initiatives by the Australian Bureau of Statistics. *International Statistical Review/Revue Internationale de Statistique*, 83, 436-448.

Tchuprow, A.A. (1923). On the mathematical expectation of the moments of frequency distributions in the case of correlated observations. *Metron*, 2, 461-493, 646-683.

Thomsen, I. (1973). A note on the efficiency of weighting subclass means to reduce the effects of non-response when analyzing survey data. *Statistisk Tidskrift*, 11, 278-283.

van Remoortel, H., Giavedoni, S., Raste, Y., Burtin, C., Louvaris, Z., Gimeno-Santos, E., Langer, D., Glendenning, A., Hopkinson, N.S., Vogiatzis, I., Peterson, B.T., Wilson, F., Mann, B., Rabinovich, R., Puhan, M.A., Troosters, T. and PROactive consortium (2012). Validity of activity monitors in health and chronic disease: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, 9.

Wolter, K.M. (2007). *Introduction to Variance Estimation.* New York: Springer-Verlag.

Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. *Journal of the American Statistical Association*, 47, 635-636.

Yates, F. (1948). Systematic sampling. *Philosophocal Transaction of the Royal Society of London, Series A*, A241, 345-377.

Yates, F. (1949). *Sampling Methods for Censuses and Surveys*. Griffin, London.