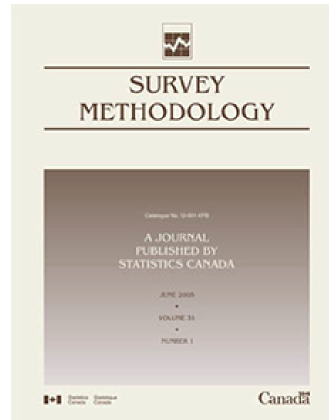
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Comments on “Progress in survey science and practice: yesterday-today-tomorrow”



by J. Michael Brick

Release date: June 30, 2025

Statistics Canada Statistique Canada

Canadä

How to obtain more information

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

You can also contact us by

Email at infostats@statcan.gc.ca

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

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

Standards of service to the public

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

Note of appreciation

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

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2025
Vol. 51, No. 1, pp. 23-27
Statistics Canada, Catalogue No. 12-001-X

Comments on "Progress in survey science and practice: yesterday-today-tomorrow"

J. Michael Brick[1]

Abstract

Survey science appears to be in a critical condition and its future direction is unclear. This paper diagnoses the situation and poses important questions to researchers and users of surveys. My discussion emphasizes the role of design in survey science and the implications of data collected without design considerations.

Key Words: Big data; Measurement; Nonprobability samples; Representation; Total survey quality.

I would like to congratulate and thank Professor Särndal on his timely contribution that confronts so many of the issues swirling about our discipline. In addition to being a leader in the technical aspects of sampling and surveys for many years, Prof. Särndal has provided profound insights into the foundations of survey science (e.g., Cassel, Särndal and Wretman, 1977). It is a privilege to add my thoughts to those he presents on this topic.

Clearly, the time has come to have a deep discussion on the issues and confusion that currently exist (Buelens, Boonstra, van den Brakel and Daas, 2012; Stedman, Connelly, Heberlein, Decker and Allred, 2019). Prof. Särndal does this by situating the issues within the context of the philosophy of science, posing key questions with which we must struggle so that we can find a path toward more progress. I particularly admire the way he places the controversy in the research tradition of Laudan (1977). Laudan's research tradition seems a more comprehensive framework than the paradigm of Kuhn (1970). I also agree with his premise that there has been only one research tradition in the roughly 100 years since the initial works by Bowley (1926) and Neyman (1934) coalesced into the current view of survey science. In my view, model-based estimation, as many understood the work of Royall and his colleagues (e.g., Royall, 1970) might have been a serious challenge, but as discussed below, I believe this work stayed within the research tradition.

Prof. Särndal does not propose the solution or the next research tradition to carry us forward. He is too wise to make such a presumption. However, he challenges us to evaluate the situation and consider our options. He focuses attention on "official statistics" which has always placed importance on quality, however quality might be defined.

The breadth of the issues raised is immense, and I can only discuss a few of those. I begin with favorite topics of mine, nonresponse and costs. I then argue to amend or clarify Särndal's definition of the sample assumption and the implications that has on considering Big Data and nonprobability samples. Finally, I provide comments on cumulative progress and the future, recognizing that the future is what happens rather than what we say will happen.

1. J. Michael Brick, Westat, 1600 Research Blvd, Rockville MD, 20850, USA. E-mail: mikebrick@westat.com.

Nonresponse and costs

The two major challenges to our current research tradition that are identified are the high rates of nonresponse and the high cost of data collection for probability samples. I and many others (e.g., Kalton, 2023) agree that these are the key threats to our research tradition, yet they are very different. Nonresponse has grown to such proportions that it may (or may have already in some cases) destroy the theoretical edifice of probability sampling as it is practiced. I recall asking Prof. Särndal at a conference about a decade ago whether he considered probability samples with low response rates to be "probability samples" and he immediately answered "No." His quick response suggested to me he had been considering various aspects of this for a long time before I asked.

While nonresponse is potentially a core threat, probability samples have always been costly. Non-response has increased that cost, but official statistics have been willing to pay the higher cost for the presumed quality associated with the product. The cost issue that is new is the vast amount of data, e.g., Big Data, that can be gathered for radically lower costs than was possible say twenty years ago. If someone offered me a new candy bar that costs 10 percent less than my favorite one, I may not be inclined to switch. But what if the price is 90 percent less? Nothing about the quality of the alternative has even been mentioned, but the much lower cost certainly has more appeal.

Designed samples

The definition of the sample assumption in the research tradition troubles me in some ways. I believe a key ingredient in the sample assumption is that the sample data are gathered using a known design. The design should be known even if the implementation is not completely rigorous, the sample design is not reproducible, and randomization is not included. Adding explicit language about the design requirement would be helpful to me. This modification would clearly admit probability, quota and cut-off sampling methods, as well as observational studies where a small number of areas are chosen purposively but the participants in each area are selected using a well-defined (perhaps randomly selected) procedure.

The model-based challenge seemed to some researchers to reject sampling and the research tradition, but Royall and his colleagues only objected to randomization as a basis for inference. That objection alone was very controversial. In his 1970 paper, Royall showed that choosing the largest units was an optimal design under his assumptions. Later, other designs such as balanced sampling were developed for model-based approaches and all of these are, to my understanding, within the current research tradition.

The addition of design to the definition seems essential to my view of the current research tradition, and it discriminates that tradition from a research tradition where the source and nature of the sample is irrelevant. Prof. Särndal mentions the desire for a "scientifically structured sample" for official statistics so my amendment may not be too far from his concept. A scientifically structured sample seems impossible (or at least incredibly lucky) without some design of the data collection.

The design of the data collection is the foundation for the construction of the two pillars of total survey quality – representation and measurement. We almost always use models in conjunction with the collected data to identify quality by describing deviations from the ideals of representation and measurement. The design-based reliance on randomization was a huge advance because, absent other departures from the ideal, it made it possible to objectively quantify sample representativeness. Measurement never had this advantage, but aspects of measurement quality were still feasible.

Another feature of the design assumption that distinguishes survey science from most other statistical sciences is that one sample can produce reliable statistics for a vast array of topics, often with a single estimation weight. In survey science reasonable accuracy is often the objective rather than optimality for a single statistic.

Big data and nonprobability samples

Many "Big Data" and nonprobability samples with which I am familiar would not qualify as designed. One antonym for "designed" is unintended, and in the data collection context this might entail lack of control. Big Data is captured, but not controlled by the data collector. In my view, this implies the quality of the product (a statistic) is almost always measured by comparison to external data sources. It also implies that the stability of the quality of the product is not within the control of the organization capturing the data. Official statistics should consider these implications carefully. Kalton (2023) discusses these issues very powerfully.

In the mid-1970s I worked on the Fatal Accident Reporting System (FARS) for the National Highway Traffic Administration. FARS gathered existing data on fatal motor vehicle accidents from police reports in the U.S. states – with many similarities to Big Data although it was not so inexpensive or big – and this is generally called administrative data. Because the police reported data were not intended for statistical purposes, our efforts were to exert some control over it so that the data were meaningful for informing national policies on topics such as seat belt usage, motorcycle helmets, and the role of alcohol in accidents. The control required coordination with the states and through them the police who recorded the data. Control involved training and the development and implementation of standards. Without this design component, the data could have led to counterproductive policies. Official statistics realized that having a census alone would not necessarily lead to good information for its intended purpose.

Of course, that does not mean Big Data are useless for inference, but it does place the challenge on proponents of Big Data to provide an acceptable quality measure for the statistics it generates. I am not aware of any mechanism for generating a quality measure from Big Data itself. The measurement pillar of total survey quality is still missing, and those interested in quality must decide if they are "really" interested in quality. Of course, low response rates in probability samples raise some of the same issues in this regard, but with probability samples there is at least a structure that may provide avenues to dealing with the measurement of quality.

Nonprobability samples, as their name implies, are extremely diverse. Some are designed and would fit within the current research tradition. Many others are not designed, even many that use matching or other techniques from observational studies. The nonprobability samples I classify as "not designed" are those based on sources that are unidentifiable. I contend that matching participants from these sources to demographic targets does not make them designed. Consider a sample generated from people who order cod liver oil from the manufacturer (the source is unknown to us). I would have no confidence in any estimates from this sample for a health survey even if the sample matched the inference population demographics of interest exactly. I can't think of an internal diagnostic that would alert us to the poor quality of the estimate.

A frequent empirical approach to quality assessment is to compare nonprobability sample estimates to census or probability sample estimates. Having been involved in this type of work, I found it to have little value. The next nonprobability sample from the same source might give totally different results. Deming taught us that the way we ensure quality is to look at the processes rather than the product, but we can't do this with these types of nonprobability samples. Another approach for assessing quality is to rely on a model of the universality of the underlying mechanism. This model implies the source of the data does not matter. Studies, perhaps most powerfully Henrich, Heine and Norenzayan (2010), have convinced me this is an unrealistic assumption. A challenge for nonprobability samples is to develop a quality measure of representation, and that challenge has defied researchers over the decades (see Kruskal and Mosteller, 1979).

Cumulative progress and the future

Another key achievement of survey science highlighted by Prof. Särndal is the cumulative progress it has made over its history. But he questions whether this progress may be ending. We may be approaching, or have already reached the point, where nonresponse imperils the research tradition. My hope is that this not the case, and that high enough response rates can be achieved if we are willing to pay for the desired quality. Response rates in mandatory census data collection in the U.S. and Canada suggest this is not completely wishful thinking. However, the cost may mean a reduction in the volume of high-quality surveys and statistics. Could we be beginning an era in which there are many low-quality surveys and only a small number of high-quality surveys? We may already be in this world, and the consequences for policymakers failing to understand the differences in quality could be problematic.

I, like Prof. Särndal, do not see a new research tradition has yet emerged. Nor do I see the direction that a new research tradition might blaze to replace our current one. Algorithms to extract data more cheaply may be useful and publishable, but without a statistical underpinning for inference how far can such advances take us? Prof. Särndal indicated that a replacement research tradition must be based on a new and powerful statistical notion with appeal to researchers and users. I fully agree but also want to suggest a new practical notion may also be required. The simplicity of producing hundreds of estimates from a single set of survey weights is very powerful. Model-based estimation never really came to grips with the importance of this simplicity for practitioners.

I do not doubt such notions could be developed, but the focus on more data and cheaper data only do not qualify as powerful notions to me. My impression is that any such idea would dramatically change the way we think of producing statistics and evaluating the quality of those statistics. Perhaps, a seminal idea is percolating in the mind of a researcher somewhere that will liberate those of us immersed in the current research tradition.

References

Buelens, B., Boonstra, H.J., van den Brakel, J. and Daas, P. (2012). Shifting paradigms in official statistics: From design-based to model based to algorithmic inference. Downloaded June 12, 2023 from http://pietdaas.nl/beta/pubs/pubs/DiscPaper_Shift_paradigm.pdf.

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

Cassel, C.M., Särndal, C.-E. and Wretman, J.H. (1977). Foundations of inference in survey sampling. New York: John Wiley & Sons, Inc.

Henrich, J., Heine, S.J. and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61-83.

Kalton, G. (2023). Probability vs. nonprobability sampling: From the birth of survey sampling to the present day. *Statistics in Transition*, 24, 1-22.

Kuhn, T.S. (1970). Logic of discovery or psychology of research? In *Criticism and the Growth of Knowledge*, (Eds., I. Lakatos and A. Musgrave), Cambridge University Press.

Kruskal, W., and Mosteller, F. (1979). Representative sampling, III: The current statistical literature. *International Statistical Review*, 245-265.

Laudan, L. (1977). *Progress and its Problems; Towards a Theory of Scientific Growth*. Univ. of California Press.

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.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57(2), 377-387.

Stedman, R.C., Connelly, N.A., Heberlein, T.A., Decker, D.J. and Allred, S.B. (2019). The end of the (research) world as we know it? Understanding and coping with declining response rates to mail surveys. *Society & Natural Resources*, 32, 1139-1154.