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

Survey Methodology

Author's response to comments on "Progress in survey science and practice: yesterday-today-tomorrow"

by Carl-Erik Särndal

Release date: June 30, 2025

Statistics Canada Statistique Canada

Canada

How to obtain more information

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

You can also contact us by

Email at infostats@statcan.gc.ca

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

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

Standards of service to the public

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

Note of appreciation

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

Survey Methodology, June 2025
Vol. 51, No. 1, pp. 61-80
Statistics Canada, Catalogue No. 12-001-X

Author's response to comments on "Progress in survey science and practice: yesterday-today-tomorrow"

Carl-Erik Särndal[1]

Abstract

This rejoinder is arranged as a series of themes or issues, inspired by the original article, and addressed, to varying degrees, in the six discussions. Among the themes: probability sampling and other paradigms in survey science; the role of the national statistical institutes in the growth of survey science; recent breakthroughs in the use of administrative data in statistics production, with multiple data inputs; the research tradition: a finite population and a well-behaved sample; deepened awareness, in recent decades, of the tradition and its ramifications; the theory track and the role of the academic sector; attempts, over time, at resolving problems; imperfections in the data collection, in the realized sample; nonresponse treatment, responsive design, panel surveys; realpolitik in national statistics production: a realistic approach to meet urgent demands for statistical information.

Key Words: Administrative data sources; Finite population; National statistical institutes; Paradigms in survey science; Probability sampling paradigm; Representative sample; The academic sector; The research tradition; The theory track; Valid conclusions.

1. Introduction

It is a privilege to attempt a reply to six insightful discussions of my article on progress in survey science. Together they are highly valuable also in that the six distinguished authors draw on different backgrounds and experience in our discipline. For some, this comes from a career in, or close to, the national statistical system of their country. For others, it comes from a more theoretical work environment, as in academia. For yet others, the background is a combination of both.

Progress is fundamental in science; it is necessary. My preparation for a talk – at the 2022 workshop of the BNU Network on Survey Statistics – directed my attention to "progress". (BNU is short for Baltic-Nordic-Ukrainian.) I recalled my reading some years earlier of the book titled "Progress and its Problems", by Larry Laudan (1977), philosopher of science. He proposes "research tradition" as a novel concept, in his critique of Thomas Kuhn's (1962) philosophy and the paradigm concept.

Hence, "research tradition" became a key instrument for my article. I used it to reflect on "progress" as it applies to survey science, but without an ambition to bring forth the most distinctive steps and dates in that long progress. My article is not an attempt to write the history of survey science. Others have done that very well.

The continued progress in survey science will take paths that are not fully evident to me, nor do they seem so to the discussants of my article. The discussions helped my further understanding on a number of points.

1. Carl-Erik Särndal, Professor Emeritus, Statistics Sweden, Lilla Norregatan 12, SE27135 Ystad, Sweden. E-mail: carl.sarndal@telia.com.

2. Science is evolving

Time passing, a well-established science may find itself in a period of turbulence or unrest, as when significant changes in conditions and outlook are happening. Scientific revolution and paradigm shift may occur. This is neither unusual nor unexpected. Survey science is no exception.

As Constance Citro notes, paradigms once universally accepted must evolve or die. I am glad and thankful for her insightful discussion. One reason is that words in her 2014 article in *Survey Methodology* were my inspiration and starting point for this inquiry into progress in survey science. She had noted there that "official statistical offices need to move from the probability sample survey paradigm of the past 75 years".

Citro's words reflected, I felt, a genuine feeling that staying rigidly within the confines of the probability sampling paradigm could stall or delay a much-needed development, in this particular case towards multiple data inputs for surveys. I can share this "ill-at-ease" with something that may, in some situations, seem a burdening constraint.

A ruling paradigm is a dominating presence. In a period of changing conditions, it may impede the progress and flow of a discipline. A paradigm shift may occur and bring a change. I return to some of Citro's comments in the following. At one point, I wondered whether much theoretical content in survey science has meant a distancing from the objectives and obligations of the national statistical system in a country or a group of countries. I think this is a very real issue in need of an examination, in more depth in some other forum than this one.

3. The probability sampling paradigm on trial

The intrinsic benefits and advantages of probability sampling are denied by essentially no one. But those advantages do not seem any more to outweigh perceived disadvantages. The marginal benefit, so to speak, of probability sampling is no longer sufficiently evident to the national statistical institutes (the NSIs) and other key participants, acting as they are under pressing conditions and cost saving objectives. Probability sampling will continue to serve as a part of survey science and practice, but not in the same dominating presence.

Dissatisfaction with the probability sampling paradigm comes from two sides. One reflects today's difficult survey conditions: a high cost of reaching precisely those in the designated probability sample, substitutes being strictly speaking disallowed by the theory. Follow-up of high nonresponse is costly.

Another dissatisfaction, in the eyes of some observers, rests in a perceived preoccupation in the survey science literature with theoretical content. I can share this concern. Probability sampling opened the way for a wide stream of mathematical/statistical contributions. This took many forms and was documented in hundreds of articles, especially over the past 50 years.

Such profusion is normal, particularly during a phase of "normal science", to use another of Kuhn's terms. Some of the effort brought valuable progress, some of it remains without any clear chance of implementation in survey practice. But it is a manifestation of the fact that intriguing theories were formulated; I return to this later.

4. The growth of survey science

My article may – indirectly – give an impression of addressing survey science from the limited perspective of statistical inference, applied in the context of the finite population and a sample from it. I do not completely deny it. As Robert Fay notes, there are two sides of our discipline, an empirical science side and a mathematically oriented side, and my text may have given, implicitly, more attention to the latter.

When the roots and the history of survey science are traced and told – as several have done in excellent ways – then the story begins, unavoidably it seems, with the name of a courageous personality, the head of a national statistical agency in the 1890's, and his revolutionary view that a "well-placed smallish sample can be enough" to portray a whole country, in his case Norway and its population. Survey science took off from that point, from a slow start to a prodigious presence, as Risto Lehtonen briefly recalls in his discussion. Our profession takes pride in this overture and the expansion that followed.

It may be possible to write the history of survey science as springing from a different seed, another originating concept, than "a smallish sample from a finite population of objects". It has perhaps been done.

It is as if that Scandinavian initiative and its many consequences set forever a mark of high prestige in survey science on this aspect of the field: to achieve "best" or "optimal" accuracy of estimates, in some mathematical/statistical sense.

It is not surprising that the idea of "just a smallish well-placed sample", with its myriad of intriguing possibilities, attracts the curiosity of theoretically oriented minds. This inspired much research, some of it of a highly theoretical nature. Prestigious scientific journals accept those kinds of contributions.

We may also recall how generations of students of the discipline were trained. They were exposed, in university courses, to the long-winded formulas of survey sampling. Those were abundant in the classical texts that students at the time lived by, and came to admire, texts from the fifties and sixties, by Cochran (1953), by Hansen, Hurwitz and Madow (1953), and by Kish (1965).

Those books give wonderful insights into statistical thinking in surveys. The rather inexperienced first- or second-year university student saw them clad in mathematical language. The ultimate proof of the pudding was the teacher's careful demonstration of a low or a zero bias, and a low variance, compared with competing methods. Highly accurate estimation is evidently the objective.

One can take a view that achieving the "absolutely best" or "theoretically optimal" method – under sometimes very special assumptions and conditions hard to guarantee in practice – may not be an ideal path

for survey science to pursue, particularly not in the eyes of an NSI with a mission to produce statistics that are "valid" or at least "acceptably good" for the country.

Warnings can be heard today about risks of "invalid estimates" when non-probability sample sources – big data and others – form part of the estimation material. Here, the word "invalid" seems as yet another reminder that it is "high accuracy" that can save the situation. But times are changing. Scientific revolution may bring a completely new perspective on "good enough, sufficiently valid, estimates".

5. Paradigm as a distinctive concept

The exact meaning of "paradigm" is not clear. It has been pointed out that Thomas Kuhn, in the second edition of his famous book, used the word in 21 different ways. A paradigm is commonly understood as a set of concepts or thought patterns in a given field. It is a system, a framework for scientific thought and effort. It may be of a theoretical nature. But often in modern language, "paradigm" is just a popular word, used for a theme, a concept or an idea, with no claim to "universality".

Early on, probability sampling was not a technique invented, proposed and named "paradigm" by a well initiated observer. It was a fertile concept for sampling and data collection. It turned out to have a great convincing power throughout the scientific community. It was a painstakingly slow process to get there. Much later, when Kuhn's influence had made "paradigm" an everyday word, it came to be called "the probability sampling paradigm".

To my mind, "a ruling paradigm" has the ring of a high ideal, confirmed, perhaps after a long struggle, in the collective mind of the scientific community, so as to become the respected norm. Anybody with good insight and overview can propose a framework for thought in his/her field. Whether or not it will gain the convinced adherence that makes it a "universally accepted paradigm" is a different matter.

More than one of the discussants seem to regret, reluctantly, the loss of probability sampling as a dominating presence and guiding light in survey science. They see no apparent replacement, none of the same unifying caliber as probability sampling.

To elevate a new line of thought into a status of a "universally accepted paradigm" takes a powerful concept, preferably present also in the name. A solid paradigm needs a transparent underlying idea that "holds together" its components.

The probability sampling paradigm satisfied this requirement, through the notion of probability. The paradigm sent a message that was appreciated and understood by everyone, professionals as well as the general public: a sample selection that is random but fair, giving every population unit a chance to be selected, with a known probability of selection, making possible also probability statements about the probable error of computed unbiased estimates.

Nevertheless, it took a long time to make "this simple idea" accepted and implemented. Long ago, when survey response was nearly one hundred percent, published statistics could justifiably be declared valid or

trustworthy if their computed standard error was sufficiently low. Not so anymore. Theory has not measured up to today's difficult survey conditions. High nonresponse tarnishes estimates; unknown bias afflicts them; accuracy is unknown. Theory has, as it were, abandoned reality. But point estimates continue to be produced, in the national interest; this is necessary.

6. Other paradigms

My expressed fear that "a new paradigm seems far away" was dissipated by Constance Citro and Eric Rancourt. They name new paradigms: "the mixed data (or blended data) sources paradigm" (Citro) and "the admin-first paradigm" (Rancourt). Both signal important steps forward in their respective settings, the U.S. federal statistical system and Statistics Canada. Moreover, the message they send is of much broader scope.

These two proposed paradigms extend, in a broad sense, the probability sampling paradigm: they are about the data intake for statistics production. Both turn around one central issue: the effective uses of administrative sources of data for statistics production. To assemble thoughts and ideas on how to capitalize on hidden, less expensive but admissible data resources is indeed a timely task, one that appeals to many, or seems even necessary at this point.

I return in a later section to Citro's interesting account of new survey initiatives in the U.S., in the spirit of the mixed data paradigm. Rancourt does not mention specific examples where Admin-first influenced or changed key products at Statistics Canada; they probably exist there as well.

Rancourt makes the point that if a paradigm shift is coming, in the spirit of Admin-first, "it is a methodological shift rather than a statistical shift". Admin-first is about "first considering potential non-survey data sources before launching surveys". It says, I believe, that a future perspective will emphasize the structural and developmental task, the integration of sources of data, administrative and others, survey data as well as non-survey data, in order to best serve the statistical purposes.

This perspective would, perhaps, focus less on the statistical methods content, the kind that accompanies the classical survey cycle "sampling frame/sample selection/estimation". Yet Rancourt adds that "as long as inference is required, a statistical structure is needed", and that "the question is to how to… take all the available data sources into account in order to enable valid inference". That is, once we know how to assess "valid inference" under inputs from diverse data sources, we shall perhaps be able to mix them appropriately. It remains a future task for survey science.

Michael Brick notes in his discussion that "the focus on more data and cheaper data only do not qualify as powerful notions to me". However, Constance Citro is well aware of the dilemma when she makes the point that "effort is needed in documenting and evaluating other data sources to deeply understand their uses and limitations, adapting processing systems to smoothly integrate other sources with survey responses".

The paradigms of Citro and Rancourt appear to have evolved primarily in a local environment, the U.S. federal statistical system on the one hand, Statistics Canada on the other. Contributing to their high importance is the fact that they try to overcome a reluctance – with roots in North American history and cultural values – in regard to registers of micro data, especially at the individual level. They may become "universally accepted" names, inscribed in the every-day scientific jargon.

7. Administrative data in statistics production: Who has got it?

Countries differ in how well prepared and ready they are to benefit from administrative data in statistics production. Reasons are part historical, part cultural. As Citro and Rancourt are well aware, the Scandinavian countries are at the forefront in the use of administrative registers for national statistical purposes.

These countries rely on a long tradition in the development and maintenance of high-quality administrative registers of different kinds. This puts them in a privileged position, compared with many others, in the use of administrative sources. Some other European countries are now also becoming well advanced in this respect.

The book by Wallgren and Wallgren – referred to by Risto Lehtonen – took a comprehensive look, probably the first of its kind, at administrative registers, their uses, problems and possibilities, within a national statistical system. In its third edition, 2022, it is titled "Register-based Statistics: Registers and the National Statistical System". As these authors have appropriately pointed out elsewhere, to understand the possibilities of administrative data "you must change your statistical paradigm. … The new paradigm cannot be based on probability and inference theory, as these historically important parts of statistical science cannot help us when we work with registers".

In Scandinavia, some important national statistics are produced directly and reliably from registers, without any data collection for a selected sample. For sample surveys, registers serve in creating advanced sample designs, and as a powerful source of auxiliary variables – control variables – at the estimation stage.

In these countries, the blend of different administrative data inputs is natural in statistics production. It was a reality in Scandinavia well before this became known as "mixed data sources". Understandably, the development has been slower in many other countries, for historical and cultural reasons.

As Lehtonen points out, Denmark and Finland pioneered in carrying out a census entirely on administrative sources. He also mentions how, in Finland, administrative registers from different sources are merged with sample survey data to create a multi-purpose micro-level data file of high quality.

Lehtonen indicates how Scandinavian countries envisage official statistics production in a frame of mind where administrative data will be the prime source, and where a "direct data collection", by survey sampling, enters only to the extent needed to complement the administrative content. Both Citro and Rancourt favor, I believe, such a future also for the U.S. and Canada.

8. The role of the national statistical institutes

Survey science has an affinity and a privileged relationship with the NSIs and their mission to produce trustworthy numbers, official statistics, for the nation state, or for a union of states. It is not an exaggeration to say that the NSIs and their needs have to a considerable extent driven the development of survey science. Much progress in survey science was linked to the NSIs, realized over the years by individuals, or teams, who worked in, or in close association with, those agencies.

Whatever the future may hold in store for survey science, the NSIs will continue to fulfil their mission to produce, to the best of their ability, much-needed statistics for the nation, or for an association of nations. Survey scientists will continue to be inspired by signals from those agencies, especially the leading ones, about issues waiting to be resolved. This will grant continued progress in survey science.

In many countries, the NSI has earned a reputation for high reliability and quality. It is a source of pride for the NSI when it gets recognized in the national media for professional competence. Trustworthy statistics has a political dimension. The governmental link distinguishes survey science from several other branches of statistical science.

Some NSIs were particularly successful and influential; research groups of high competence produced remarkable results. An outstanding example is the U.S. Bureau of the Census, as when, in a legendary period, Morris Hansen and his group were active there. When I was a young statistics student in Sweden, teachers made me keenly aware of the far-away U.S. Bureau of the Census, its excellent reputation, and the path breaking work done there. I perceived an emphasis on high accuracy of statistics produced in the national interest, as a basis for far-reaching decisions on important policy matters.

As for modern problems of rising costs and declining survey response, Rancourt notes that “the fact that NSIs are not spending more is not due to a lack of willingness to pay for quality. Rather, it is related to the fact that the level of effort needed (and therefore costs) may not even be achievable (even if one were willing to pay) to attain the traditional levels of response”.

In today’s society, it is hard to get the desired level of support and participation from the general public. But the cited words also hint at a predicament within an NSI: a lack of sufficient personal and other resources for bringing creative and effective solutions. Research, in particular basic research, is after all not the principal role of an NSI; as many see it, the nation’s taxpayers should not have to pay for such activity.

Two segments in society influenced survey science, in its formation, growth and consolidation over the past one hundred years: the NSIs on the one hand, academia on the other. Both had a profound impact. Survey science navigated between two strong poles: the NSIs and academia. At a number of occasions, when a pressing problem emanated in the national statistics sphere, academia responded and invested effort; valuable results, solutions and new methods, resulted from this. The academic sector became especially apparent from the sixties and on, as I mention later.

9. Research tradition

To avoid the overused word "paradigm" I preferred another assembling concept in philosophy of science, namely, "research tradition". I wanted to see how that concept might be concisely formulated with respect to survey science, as an umbrella for much of the important work accomplished in that field, theory as well as practice. At the same time, it should hold continued hope for future activity. Continuity and continuation are essential.

A research tradition can be seen as a movement in time where a community of participants "take charge of a science", shaping its progress, while sharing, as it were, a silent agreement. The tradition unfolds and sets the course of that science. It is a wider concept than paradigm. Over a period of time, the tradition may entertain several paradigms, successively or simultaneously.

I formulated the research tradition in survey science in two assumptions: Firstly, a finite, perhaps large, population, a set of identifiable objects or units. Secondly, a smallish subset, a sample, serving as a basis for conclusions, or inference, about the population.

The two assumptions are "the bare bones" of survey science, in a sense self-evident, to a point where it may seem superfluous to even mention them. But there is good reason to reflect on them, as I try to do. They allow an extraordinary richness for activities, directions and opportunities to explore; they set survey science apart from most other areas of statistical science; they served as a broad umbrella for research.

10. The sample

As Robert Fay observes, "the sample assumption… perhaps should be regarded as the single distinguishing feature of the discipline". It is indeed the critical one, and will remain so. A key question came to occupy survey science for decades: How to make sure that, in the end, the realized sample is as good as possible for drawing valid conclusions about descriptive parameters of the whole finite population.

As Fay further notes, my formulation of the tradition "does not set sharp boundaries on what to include and what to exclude from the definition of a sample". As is generally understood, it should be "a quality sample", apt to deliver valid and trustworthy conclusions. I had suggested the term "scientifically structured sample". That description is still insufficient and imprecise. I agree with Brick that such a sample is impossible without "a design of the data collection"; he adds that design is the foundation of the two pillars of total survey quality, representation and measurement.

Five words capture the essence of the research tradition: "a smallish sample can be enough". It was a break-through when established about one hundred years ago. From then on, survey science built on the necessity, for cost and other practical reasons, to limit data collection to "just a minor part". The NSIs, some in particular, participated actively in this development.

The words “can be enough” express an uncertainty, a doubt, but also an exciting promise. Attempts at handing this uncertainty shaped survey science, as it unfolded in the last one hundred years. The tradition entertained different theories – to some degree disagreeing – to structure the thinking about this uncertainty; some are mentioned in section 11 of my original article.

For good reason, probability sampling grew into a powerful theory in the tradition. But samples drawn by other well-motivated designs are admissible; the tradition does not exclude non-probability sampling, something that may be important for the future of the tradition.

Sample selection and data collection that is convincingly justified by other means than probability sampling can still qualify as belonging to the tradition. But then a heavy responsibility rests on the design. As Brick points out, non-probability samples are diverse; some but not all are designed.

In any case, it is understood that a data collection procedure has yielded measurements on the important survey variables for the members of a selected sample. But perhaps not all. Some selected units may be non-responding; measurements are missing; the sample is then the set of responding units, the effective sample, the basis for computing estimates for the population.

Mary Thompson evokes a high-tech future scenario for data collection: “that we could find out what we need to know about the population without necessarily asking the members of a well-constituted sample to respond”. Perhaps not utopian, it suggests a brilliant future for the tradition, free of problems that have plagued it for a long time: nonresponse and missing data.

Over the years, the general public came to accept that a sample can be enough, provided it is “representative”. The word is just about unavoidable in a producer’s report – say in the national news media – about the results of a survey. After all, the sample is not the whole population. It must “represent” the whole population, everyone understands that. The term serves just to reassure the general public.

Much thought and speculation has been devoted to the term “representative”. The property was granted, in a particular sense, by probability sampling, when executed with one hundred percent response. But under non-response, the representativity is at risk. I return to the nonresponse question in Section 12 below.

In a purely model-based survey theory, a sample may be designated strictly on the grounds of the model assumptions. Such a sample is perhaps both non-representative and non-probability, but can belong in the tradition: there is a population, there is a sample. A heavy onus is on the believer in the model-based theory to guarantee the sample’s fitness for use. As Brick appropriately points out, one challenge for non-probability samples is to develop a quality measure of representation.

Because the sample is a crucial element in the research tradition, it does not, as formulated, cover cases where statistics for the nation are produced without any sampling, but entirely from administrate sources; it is a field of its own.

11. The population

In the 19th century, a number of emerging nation states established their own national statistical institute. A few forerunners had them earlier. It was an era where quantitative aspects of the state were becoming increasingly important, to assist the government and to mark the nation's place in the world.

In early times, the concept of "population" referred to the persons residing in the country. As Fay points out, "the notion of population is not a recent conceptual insight for humanity". For centuries, rulers and administrators of kingdoms or other states have wanted to know the number of citizens within their borders. Censuses, efforts to establish an exact people count, have a long history, be it for geopolitical or linguistic or religious or yet other reasons. For example, Canada takes pride in the 1666 census of New France under the direction of intendant Jean Talon.

The importance of "population" is accentuated in today's world through supra-national political formations, such as the European Union. Each member state, with a population of permanent citizens as well as non-citizens present there, is subject to much attention and scrutiny. Surveys that target a finite population have a promising future for that reason as well.

An important development was when registers of population-related events, such as births, marriages and deaths, started to be regularly kept. Adolph Quetelet was a central figure in the nineteenth century rapidly growing interest in measures of vital aspects of the nation and its population.

More recent is the concept of population as used in modern survey science, that is, the idea of a repertoire, a register or a list of units with an identity, labelled units, from which a sample can be drawn, with or without probability sampling. Survey science extended the idea of "population", so that today it can refer to a collection of units such as persons or households or farms or business firms or schools or yet other kinds, belonging to a country or other territory of interest.

The finite population concept with its labelled, identifiable units is fundamental. It shaped survey science from early on. Hopefully it shall remain possible, also in a future context with large unstructured sets of data, flowing in from various mixed sources, to uphold the notion of identifiable units. That is, to be able to confirm an existing unit as belonging to a finite population of interest, and to identify data points in a cloud as belonging to one unique member of that population.

I am optimistic enough to believe that the research tradition – the idea of the finite population and a sample from it – can continue to lead the discipline, even though the probability sampling paradigm that guided it for a long time may lose some of its importance.

12. Examples in the life of the research tradition

How did the research tradition – the activity set in the idea of a finite population and a sample from it – react in recent decades, in the face of problems arising? Examples in this section try to illustrate this.

The challenge is to possess a "sufficiently good" sample when the data collection ends. The sample that is actually realized, with its measurements on critical survey variables, may be a less than favorable starting point. It is nonetheless the material that the statistician has to work with in producing statistics that are, hopefully, reliable.

To meet the challenge was not always easy. The three examples that follow have in common that the sample may be more or less deficient in its representativity vis-à-vis the finite population. Nonresponse in a probability sample is a common cause. The examples illustrate how the tradition faced the predicament, devoting in each case much attention to the problem, with a plethora of published work as a result.

Example 1: Reacting to nonresponse.

"Get as high a response portion as possible" was a common strategy in collecting data from a designated probability sample. Alternatively, a goal was to get the survey response rate up to at least "a respectable level", say 70 percent, depending on what the current popular outlook may have been. This can cause interviewers to seek response primarily from units deemed to be "the easy cases", thus a questionable practice from a representativity point of view.

The set of responding units is not a probability sample; response probabilities are unknown. Nonresponse was blamed, for decades, and justifiably, for its harmful impact on survey quality, notably a high risk of biased estimation. Consequently, the tradition took a lively interest in creating "nonresponse adjustment techniques", as an outgrowth of the estimation theory.

"Identity" was a central notion in this. A unit can, in the estimation, be given a weight that reflects its known specific attributes. Weighting was to a degree successful, and is still widely practiced, to the point that some maintain that even a considerable nonresponse may not be such a serious problem for quality. The idea of an advantageous "weight system" has been fruitful in survey science. As Brick points out: "The simplicity of producing hundreds of estimates from a single set of survey weights is very powerful."

Imputation for missing values was another remedy that inspired much work. Thus both weighting and imputation are extensively documented fields in the scientific literature.

Example 2: Attitude to panel surveys.

The idea of the panel survey, when first conceived, was a clever variant in the research tradition: There is a well-defined finite population, and there is a sample from it, tailored to circumvent or reduce the nonresponse problem.

None of the discussants chose to comment on my words about panel surveys. Although much criticized in some circles, panel surveys cannot be avoided in a discussion of the research tradition. Much survey activity resides in that sector, especially by private survey institutes, in opinion research, including political opinion surveys. Sampling takes place from a panel, which may, by construction, consist of population members favorably inclined to participation in surveys.

Needless to say, a central question is the quality of the sample present at the end of the procedure. The panel, claimed to be a "mini-population", and commonly described by the unavoidable word "representative", may be recruited in several steps, where selection effects, such as nonparticipation and non-representativity, intervene. Be that as it may, in the end there is a sample from the target population. How well does it stand up to a rigorous scrutiny?

For a comparison, suppose a national statistical institute carries out a survey with a probability sample from the complete population register, and realizes only a 50% unit response rate in that sample, resulting in computed estimates with an unknown bias. Can we then attach any great value or advantage to the probability sampling feature? The responding half of the sample is non-probabilistic, response probabilities being unknown.

Compare with a private survey institute that carries out the same survey with more than a 90% participation rate in the sample drawn from its panel, albeit not flawless. Which of the two surveys delivers the best quality, the better accuracy, the better value to society? It is a complex question without an easy answer, one that deserves to be taken seriously and analyzed in depth. But parties involved lack sufficient time and interest in doing this. Instead, features such as lower cost and quicker release of survey results will seduce producers as well as users.

It sounds like a double standard if we consider it acceptable to compute national statistics on a probability sample with 50% response or less, but unacceptable to compute such statistics on a high-participation, although non-probability, panel sample. Which lack of representativity is the more damaging one? It is a question seeking a convincing answer in a mathematical/statistical language.

There may be an intrinsic resistance, on the part of NSIs, intent on producing "statistics in the national interest", to use a methodology possibly less scientific, less academic, more subject to criticism.

Still, it may seem surprising that prestigious producers have not to any greater extent seized upon an opportunity to examine in more depth this variant of the research tradition, to upgrade it and develop their own refined version of it, gaining thereby high survey participation at low cost, and a quick release of statistics.

Example 3: Responsive design.

In this attempt, the research tradition seeks to establish a good sample by focusing on the course of the data collection. In a survey prone to a considerable nonresponse, it is about managing the data collection process so that, in the end, the effective sample, the response set, will bear strong resemblance to the probability sample that was drawn, the one that one would trust if all its members respond.

Thus it is about adapting the data collection to signals such as "we already have enough of those" – as one expert fittingly formulated it – and to manage instead the remainder of the data collection so that, in the end, one can dispose of a response set that is well balanced with respect to a vector of control variables.

Research on responsive design resulted in a large literature. On a negative note, such designs can be expensive and time consuming, requiring a detailed procedure for the day-to-day effort in the data collection period. Again, it is a question of marginal benefit: a perhaps less than convincing improvement, especially when time and cost must also be taken into account.

Discussion

The three examples in this section generated extensive activity within the research tradition. Did it result in any progress for survey science? Or is it rather about stand-still and trouble-shooting?

We like to think of "progress" as striking advances in theory, or as the implementation of remarkable novel techniques. But here, "progress" seems rather to be about the tradition's attempts to handle imperfections in the manifest sample from the finite population. It is progress in a sense similar to hard attempts at meeting goals in the battle against a threatening prospective in society, such as global warming.

Much of the blame lies on the nonresponse, high and rapidly rising in many surveys. Indirectly, one can thus blame the probability sampling theory itself. It requires, formally, that everyone in a designated probability sample must respond. It is a vulnerable theory that can be said to be unrealistic in today's conditions. The research tradition thus found itself engaging in "patch-up work", an understandable but apologetic attitude, wishing to alleviate the problem and still remain near the probability sampling ideal.

Brick feels that we may be approaching or have already reached the point where nonresponse imperils the research tradition. Nonresponse may have imperiled the probability sampling paradigm that served the tradition well. But has it imperiled the tradition? I do not think so.

In the future, the research tradition should certainly not get rid of vital concepts such as rigorous design and representativity. But if at all possible, it would be a liberation to rid the tradition of the nonresponse concept, this restrictive idea that it is only the set of units designated by a probability sampling that can deliver good survey results.

The research tradition does not stand and fall with the probably sampling paradigm. High nonresponse and high cost may imperil that paradigm. But the research tradition, set in its two assumptions, can survive and flourish. It is a challenge.

13. Historical interlude: deepened awareness of the tradition

I am going back in time to a period that one might characterize as "a deepened awareness of the tradition". It became apparent from the mid-fifties and onwards. Sampling techniques had been present and extensively practiced for several decades already, set in the tradition: the idea of a finite population of identifiable units, and a sample drawn from it. Attention had focused on sampling realized by one of the

classical designs: stratified, systematic, clustered, in two or more stages or phases, by probability-proportional-to-size, or yet others. These probability sampling designs could be successfully practiced, at that time, in environments of low or negligible unit nonresponse.

But then it was as if some protagonists were asking: Where does the tradition take us, at a closer look, with this idea of a finite set of units and a probability sample drawn from it? The ramifications and possibilities of the idea were examined. "Foundations of survey sampling" became a field of study. Conferences were devoted to the topic in the sixties. It gave the field an academic accent and flavor.

I hesitate to use the word "theoreticians", but prominent inputs in those debates came from statisticians in academia, knowledgeable in general statistical theory. A certain cleavage became apparent. More visible than in earlier times, a theory segment of survey science started to develop, alongside the pragmatic segment.

Mary Thompson's discussion recalls two striking contributions from academia: Godambe's (1955) view of inference for a finite population, including a famous "nonexistence theorem", and Basu's (1971) amusing "circus example". Both were important, not so much in terms of a direct influence on practice, as for the influence they had on the theoretical discourse in the discipline. It was an examination "from the inside", from the mid-fifties and on.

As Mary Thompson notes, Godambe (1955) "helped to stir a strong interest in rival approaches to inference". The theory development took note of that paper. On the other hand, prestigious national institutes, such as the U.S. Bureau of the Census, did not let themselves be overly disturbed by Godambe's message; it was seen rather as an outcome of academic thinking, with little or no consequence for practice. Convinced that their methodology provided good statistics for the nation, these organizations carried on as before.

Basu's (1971) "circus example" is a brilliant piece of statistical satire. Its delightful mockery of sacrosanct notions in survey sampling brought wide attention, not only in survey science circles.

We can ask: Who or what is it that Basu is making fun of, with his irreverent example? What are the principal targets, or victims, in this satire? Is it the theory itself, the pedantic assignment of known selection probabilities to the members of a finite population? Is it the principle to rely on just a small probability sample? Is it the survey statistician, impressed by theoretical notions of doubtful practical consequence?

In his example, Basu plays with the two basic concepts in the research tradition: The finite population with its identifiable units, and a sample, drawn with a design that gives non-zero inclusion probability to every unit. In addition, the example questions the value of a respected quality standard: Unbiased estimation, under the chosen design.

Basu skillfully manipulates the first two to make the third give the gullible circus statistician a ludicrous – nevertheless unbiased – point estimate. A bizarre population – fifty elephants – but a legitimate one. A silly sampling design, but one that will deliver a legitimate probability sample.

All the formalities are satisfied but the result is a disaster. It is as if Basu is telling the survey science community, with a twinkle in his eyes: Look, if you prefer to hold on to this research tradition, beware of its absurdity in some cases.

Again, the NSIs were not noticeably disturbed by this academic exercise; they carried on in the conviction that their intelligent survey designs were producing high quality numbers for the nation, leaving the academics alone to entertain themselves with bizarre survey situations.

In hindsight, did Basu's popular example cause more harm than good? A severe critic of survey science may maintain that its pillars are inapt, or shaky, unworthy of a serious science. How did this example affect the reputation of survey science, within the community of sciences?

Be that as it may, Godambe's nonexistence theorem and Basu's circus example were remarkable events on the survey science scene, as it was unfolding in the fifties and sixties. They offered thought-provoking views into the special nature of survey science, material for students to reflect upon, notably at the Ph.D. level.

It is true that the startling results of Godambe and Basu were set in unrealistic circumstances. There was indeed a finite population and a sample from it, but an absence of difficulties such as inadequate sampling frames or nonresponse. The two results bring out what one might call "blind corners" of the tradition. Nevertheless, they became parts of the theoretical heritage.

14. The theory track of the tradition

Survey science did rely on theory before 1970. But theory took on a distinctively new character from about that time. A segment of the research tradition engaged on "a theory track", enriching it with important theories, notably about the estimation of finite population parameters. This theory content broadened survey science. Much research of a theoretical nature was published.

This theory progression followed from "the deepened awareness". The debate on "foundations" foreboded "the formative seventies". Estimation theories were proposed. They were competing, to some degree contradicting and were much debated. Some are mentioned in Section 11 of my original article, bearing names such as model based, model assisted, calibration, small area; names seen in titles of many published research articles in recent times.

Other incisive theories addressed topics and practices such as questionnaire design and measurement, the character of nonresponse, the methodology of imputation, the approaches to mixed mode data collection. New terminology was created and became familiar language, as, for example, when terms such as "missing (completely) at random", "ignorable" and "non-ignorable" became focal points in the discourse on non-response.

The research flowing from these theories left some lasting imprints on statistics production, in the NSIs as in other survey environments. Much basic research was done in the academic sector, whose presence in survey science had become increasingly important from the seventies and on. One result of this was many Ph.D.'s in survey science, persons who, in their turn, contributed to methodology in the NSIs, in the positions that they were offered and held in leading agencies.

These developments also pointed to a certain polarization, "theory as opposed to practice". Some of the theoretical work struck the practitioners – some of them active in the NSIs – as rather remote and of little consequence.

Two cultures can thus be said to have developed and prospered in survey science. "The theoreticians" represented a culture rather different from that of "the practitioners". It was not always easy for the two camps to meet on a common ground and to communicate in a harmonious fashion.

A certain disappointment may have been felt at times. More than one practitioner certainly wondered, at some point: "What do we do with all this theory?" To achieve a more fruitful balance between theory and practice is a future objective for the tradition.

The theory track stands on its own feet. It is a significant branch of survey science. The assessment of progress and importance takes place to a considerable degree within that branch itself. Research in this track finds publication in leading statistical journals, some of a distinctly theoretical nature. Some of the content may be remotely connected only with the daily activities in an NSI; an implementation in practice of some of the specialized material seems remote. Nevertheless, this work is an important component of survey science.

In parallel to the progress of a mathematical character runs the important development of Total Survey Design as a basic structural framework in survey science. As Fay indicates, it draws on the empirical sciences. Thus, in recent decades, the research tradition experienced a distinction also from another point of view, that of the basic underpinnings. There are two outlooks within the tradition, different in aspiration and content, one of a mathematical orientation, the other of a social science orientation.

15. Uniformly excellent survey quality is a hard proposition

Brick raises an intriguing and challenging question: "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."

To assess and rank surveys in regard to quality is not easy, especially if we accept that "survey quality", as carefully formulated by some NSIs, is a multi-faceted concept, composed of a number of interacting

dimensions. Brick's supposition brings a follow-up question: does malpractice happen in the world of surveys?

It is clear that surveys can vary greatly in quality, depending on the context and the environment in which they are placed. Variable survey quality is a fact. High-quality as well as low-quality surveys are carried out.

If such a segmentation is undesirable, it is not evident how the profession can bring about an effective change. The codes and criteria at our disposal are not crisp enough to clearly identify "low-quality" cases, let alone to prevent those surveys from being carried out. Just about anyone can launch a survey. The respectful tradition will not, for example, publicly rule out some questionable panel survey practices as clearly low-quality work.

Surveys compose a vast domain of activity, with no firm rules or standards for their design and execution. A certain volume of "low-quality" output is both natural and unavoidable in a country or in a group of countries; the tradition has to live with that.

We generally consider statistics produced and released by a country's NSI as important; these products merit high attention in regard to survey quality. This holds also for surveys that are important components of research in social science or other fields.

The NSI is active in a wide range of statistical programs and productions, including the official statistics for the nation. The NSI has certain flagship products. The agency takes special pride in those, striving at highest possible quality for those. They may be surveys where highly specialized methodology was put into place, or newly created surveys in a virgin territory, or surveys on topics highly visible in the country's news media, such as labor market conditions, including the unemployment rate and its fluctuations. Other products, without being neglected, may not get that same privileged attention. As Rancourt puts it, "the level of effort needed… may not even be achievable".

An important private survey industry is present in some countries, in addition to the NSI. These private producers are not uniform in regard to survey quality. Some deliver work of the highest quality, others may not. Realistically, they must differ in resources, in statistical competence and knowhow.

Some are small private enterprises, without the capacity or the incentive to deliver top quality. Nevertheless they are well positioned and prosper financially in a lucrative survey market, where they carry out surveys in a standardized fashion on a variety of subjects.

It is a matter of some concern that the field may in fact find itself segmented into a high-quality sector and a low-quality sector. As Brick notes, "the consequences for policymakers failing to understand the differences in quality could be problematic". Outside a rather small community of highly competent and disciplined professionals one cannot take for granted an understanding of the statistical ideas and concepts that survey technology builds on. There is a communication gap.

Constance Citro suggests for the survey methods profession "to develop standards to help users understand the weight to give to one another set of estimates", and, as a further assistance, to contribute "sensitivity analyses … that data producers should perform and document for users to have reasonable confidence in the data".

16. Realpolitik in national statistics production

In her discussion, Constance Citro describes two recent initiatives in U.S. national statistics production, The National Experimental Well-being Statistics (NEWS), for improved income estimates, and the U.S. Household Pulse Survey. They are significant achievements under difficult conditions, in the spirit of "blended data inputs". They can be seen as a result of realpolitik, to conduct surveys of national importance while accepting today's difficult reality, and without necessarily respecting all the traditional norms. In this light, I see them as highly creative examples.

From my reading of her two accounts, I retain in particular the following items:

1. Extensive research lies behind these statistical products.
2. High rates of nonresponse and missing data affect these surveys.
3. Some of the released statistics from these programs are declared "experimental".
4. High and urgent need for these products obliges the agency to produce them quickly.

Those are important points. Statistics released by a prestigious producer, in this case the Bureau of the Census, tend to be taken virtually as "the truth". They are acted upon by decision makers and other users. Even if released statistics are signaled as experimental, they can be of substantial value to a user.

The considerable effort behind those statistics programs confirms the sense of responsibility that one associates with the producer. Competent statisticians at Census developed the methodology. This in itself is a quality assurance. It is an investment that will likely benefit other statistics programs at Census as well.

Their detailed reasoning is not known to me, but Citro describes a part of the development as "using a wide variety of administrative sources to improve weighting and imputations and to adjust for under-reporting". Classical survey methodology tools, such as weighting and imputation, are thus still retained in coping with missing data, and an access to "a wide variety of administrative sources" permits those tools to function sufficiently well.

When a national agency must, as in this case, meet an "overriding goal of fast release of estimates", it faces a dilemma. Whether released statistics are valid or invalid, in a fine-comb theoretical perspective, is in a sense irrelevant. Census releases them as "experimental"; users need them promptly and find them useful. This sounds like a paradox. But it does not make our science less respectable; on the contrary.

A certain disappointment lies in Constance Citro's words saying that in "the incumbent way of thinking… disproportionate effort… had been devoted, in my view, to issues of survey weighing and estimation to

adjust for unit nonresponse bias and variability as compared to the effort devoted to incorporating data sources that would be more accurate than survey responses for key items".

As I suggested, in the section above on "the theory track", some of the published theoretical content "may be remotely connected only with the daily activities in an NSI". Theory for the sake of theory was, and remains, a part of survey science. It does raise questions. How may the research tradition in survey science progress in ways that are better in tune with the NSIs and their problems? How may progress be achieved in a more harmonious fashion?

17. Concluding thoughts

My purpose in these notes was to reflect on "progress". Neither I nor the discussants have drafted a definite program, or course of action, to achieve a continued progress in survey science; it was not a primary purpose here.

As I said in the original article, a prestigious scientific field takes pride in cumulative progress. That is, to acquire more and more scientific insight, seeking persistently to bring the field beyond the borders of past accomplishments. Such progress has been evident in survey science also, at times.

A presence of more and more data, of different kinds, from all kinds of sources, is a certain expression of growth and expansion. But "more and more data" is not synonymous with "more and more scientific insight". It is not synonymous with "more accurate statistics" either, although deep in the popular image of statistics seems to lie a conviction that this must necessarily be so.

The discussants do not take a definite stand; one can sense on the one hand a certain discomfort with the "more and more data" syndrome, on the other hand an opportunity and a promise for a rewarding future. Time will undoubtedly see this take a more definite form.

Citro and Rancourt describe initiatives that I see as positive and constructive outgrowths of "more data": The increased uses of administrative data sources in national statistics production, while respecting the need to be observant and critical, and to develop standards in this work.

Brick holds, as already noted, that "the focus on more data and cheaper data only" leaves something to be desired.

Lehtonen asks: "Does the data revolution fit into the theory and methodology framework of survey science?" His optimistic response: "My opinion is 'yes'."

Thompson notes that "… other experts would say that we have already reached the point of having too much data".

Fay, finally, expresses an appreciation for past accomplishments as well as a trust in their continued importance. He believes that "the future will be much more the outgrowth of avenues of research that are

already being pursued. Alternative sources of data will require continued diligence to assess and monitor quality, the work of empirical science, and statistical frameworks to incorporate the information".

A positive final note: Nothing in the six discussions strikes me as contradicting that a finite population of identifiable units is "a sustainable concept", worthy of guiding the science in the future as well, with the aid of a "smallish" – and otherwise commendable – subset from it.

References

Basu, D. (1971). An essay on the logical foundations of survey sampling, Part 1. In *Foundations of Statistical Inference*, (Eds., V.P. Godambe and D.A. Sprott). Toronto: Holt, Rinehart and Winston, 203-242.

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

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

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

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

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons, Inc.

Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

Laudan, L. (1977). Progress and its problems; towards a theory of scientific growth. Univ. of California Press.

Wallgren, A., and Wallgren, B. (2022). *Register-based Statistics. Registers and the National Statistical System*. New York: John Wiley & Sons, Inc.