The Survey Statistician, 2024, Vol. 89, 29-36

The first fifty years of the IASS, some thoughts

Carl-Erik Särndal

Professor emeritus, Statistics Sweden

Abstract

This article recalls and re-examines a number of ideas, concepts and lines of thought that influenced theory and practice in survey statistics, in particular during the fifty years of the IASS.

Keywords: randomization theory, types of inference, rigorous statistical treatment, modeling, theorization, the nineteen-seventies.

1 Introduction

The International Association of Survey Statisticians (IASS), created in 1973, marked its first quarter-century by publishing a Jubilee Commemorative Volume, subtitled Landmark Papers in Survey Statistics. I am reminded of it, in commenting now on two quarter-centuries in the life of the IASS. I shall refer to it as The Association.

The jubilee volume was an unusual and thought-provoking initiative of The Association. It paints a portrait of survey statistics, as it was seen twenty-five years ago. It bears testimony to an era in survey science.

This volume remains today as a witness to a period of growth in survey science, a document of a certain value in the history of the discipline. Its content reflects a subjectivity that “a selection of the best” will invariably bring. Many other excellent articles were published over the years.

The purpose of the present article is not to review the nine-teen selected articles. This may be of considerable interest, but is not the objective here. Nor is it to trace the steps in the progress in survey science that those articles brought.

The article offers a perspective on survey science taking the volume’s excellent preface as the point of departure. I note how it introduces and justifies the selected articles in terms of statistical ideas, concepts and expressions that were in vogue at the time, often mentioned in the literature of not so very long ago, but which may since have fallen more or less into desuetude. This article is thus an essay on ideas that influenced the discipline. It expresses the author’s personal opinions and impressions. It makes no claims to a complete coverage of the field, makes no attempt to write the history of the field. The result is of some educational value for a younger generation.

A creative period in survey science began in the late 1960's. It questioned the established state of the discipline. New ideas and approaches had a profound impact. I like to think that the birth of The Association in 1973 was to some degree a reaction to the new directions. The Association was a welcome addition to the scope and activities of the International Statistical Institute (ISI).

2 The selection

As the preface mentions, The Association had trusted a committee with the difficult task of selecting "landmark papers" from the vast stream of literature in survey statistics, beginning early in the twentieth century. The resulting volume presents nine-teen full-length papers with publication dates between 1934 and 1989. The stated objective was to choose among articles from roughly the last fifty years, emphasizing the significance of the contribution, rather than quality of exposition or direct usefulness.

Eleven out of the nine-teen had appeared in the years from 1969, an unbalanced selection, from a time perspective. Recent work tends to come more readily to mind. But it also reflects the fact that the times from around 1970 brought significant and in a sense revolutionary progress.

All nine-teen articles in the volume are framed, to varying degrees, in mathematical formulation and language. Although not a display of "hard mathematics", this nevertheless suggests that formal expression helps to bring about "a seminal contribution", which requires not only the recognition of a practically important survey question, but also a convincing mathematical formulation, treatment and resolution. It is the kind of article that is capable of generating a stream of further contributions.

3 One exceptional name

Altogether twenty-nine names get credit for authorship or co-authorship in this collection of seminal contributions. One name stands out, as first co-author of three of the chosen nine-teen articles. Dated 1943, 1961 and 1983, they point to a remarkable forty-year activity span. The name is Morris H. Hansen (1910-1990). He deserves to be recognized also because he was the first president of The Association.

While at the United States Census Bureau, 1935-1968, Morris Hansen was highly influential in the discipline, one of the first to develop methods for statistical sampling. He made important contributions in many areas of surveys and censuses. He was one of the principal builders of the "probability sampling paradigm"; he vigorously defended it, when needed. He and his colleagues pioneered in viewing survey quality under a broad umbrella, covering survey errors of different kinds.

4 The rise and the high point of randomization theory

The preface notes that "the randomization theory of sampling … proposed a logic of inference based on confidence intervals". This grew out of the important theoretical advances in the 1930's by J. Neyman and others. Design unbiased estimation was a key feature in this logic.

The decades that followed saw randomization inference developed to perfection, in other words, the inference built on the randomization feature of the probability sampling design. Known selection probabilities of the identifiable population units was the key to this. The selection could be stratified, in two or more stages, in two or more phases, by probability proportional to size, and yet others. A vast literature from that era bears witness to a panoply of methods; although in essence just variations of one single method: probability sampling. Randomization inference became popularly known as design-based inference.

Probability sampling with design-based inference progressed rapidly and convincingly, approved by the general public and authorities. It became a recognized tool, capable of handling a great variety of situations. Some form of probability sampling could usually be designed and carried out to deliver accurate information needed about some aspect of society.

Prior to 1970 there was no need to qualify inferences about the finite population - confidence intervals and other forms – as “design-based”. It was the golden rule. But in short time after 1970 came a need for making distinctions.

Survey statistics was seen by some as a field closed in itself, unaware or ignorant of reasoning and methods in “general statistical theory”. The topic was typically taught as “sampling” in university courses, if at all offered. It had a reputation of a rather special, somewhat marginal, field of study within statistical science. Some survey statisticians felt this distinctly, and hoped for a change.

In those days, some senior stake holders no doubt considered survey statistics as an essentially complete and saturated field. Probability sampling was the uncontested methodology. The national statistical agencies applied it. The probability sampling paradigm had tremendous power behind it. It was a formidable task to challenge this bulwark.

Survey statistics and “sampling” had built its reputation on the privileged setting of the finite universe, composed of *N* identifiable objects, or units, ideally listed in a frame, perhaps together with known properties linked to individual units, such as membership in population groups of potential interest.

Survey statistics is a scientific field in its own right. As such, it uses concepts and ideas of statistical science. This became much more evident from the late 1960’s and on, when general statistical theory came to influence survey statistics as rarely before.

5 A period of change

In the new scientific discourse, “models” and “modeling” became key concepts. Was “modeling” something new? Not at all, as some defenders of “the traditional thinking” liked to remind. Modeling was in fact present much earlier. For example, planners of a new survey could rightfully claim that their choice of an unusual probability sampling plan, say, a complex stratification, or a complex sampling design in several stages, was the result of a conscious – although perhaps not declared – modeling effort, in the interest of accurate estimation, unbiased by virtue of the randomization theory.

However, from around 1970, models became more apparent and explicitly declared, both in regard to the sampling design and in the construction of estimators.

Ample warnings were voiced. To “rely on models” – possibly wrongful or misleading – might expose the estimates to severe bias. Despite a privileged position of design-based inference, the new forms of reasoning not only survived; they thrived and set their important mark on developments in survey statistics in the decades until today.

6 Theoreticians

The preface notes that in the 1950’s and 1960’s “theoreticians addressed the foundations of randomization inference”, in attempts to “integrate randomization inference into mainstream statistical inference”.

Who are “the theoreticians”? What is their role in this practical field of survey statistics? Were theoreticians less prominent, less influential, in those early path-breaking decades of survey sampling?

“Theoretician” may refer to someone highly knowledgeable in “general statistical theory”, especially advanced estimation theory, someone with a pronounced mathematical/statistical training and orientation, but with possibly little experience of the practical work in a national statistical agency.

Important conferences were devoted to survey statistics, especially to “its foundations”. There was introspection, attempts to place of survey statistics into “the mainstream” of statistical science. Two such occasions, in 1968 and in 1977, were at the University of North Carolina at Chapel Hill. A 1970 symposium at the University of Waterloo on “the foundations of statistical inference” had an important portion on survey statistics theory.

However, the preface strikes a somewhat negative note: some of the resulting theoretical work "seemed too abstract to practitioners and may have resulted in the perceived divide between theory and practice". Nonetheless, the decade that saw the creation of The Association was one of lively exchange and debate; a march in new directions.

7 A new era

From around 1970, new theories emerged. Terms were coined for proposed new types of inference for the finite population. They were integrated into the scientific language. Survey statisticians began to communicate in a language and with a terminology hardly needed before. New terms enriched the scientific language, to inform readers of articles and participants at conferences on the nature of a contribution.

"Superpopulation" was an exotic new term for an imagined infinite universe with certain assumed features, and assumed to have generated the actual finite population, from which sample selection then took place.

As the preface further notes, "theoreticians indicated that a modeling approach could be adapted to complex finite population structures and sampling schemes". A few visionary statisticians – theoreticians, in the eyes of some – had, in the years around 1970, the audacity to use postulated assumptions – more or less trustworthy relationships among variables – as the basis for addressing the substantive issues: the sample selection, the properties of estimators, such as their unbiasedness, variance and mean square error.

Model-based theory of inference for a finite population saw the light of day, holding that inference could well be based entirely on a modeled relationship between variables, notably that of the auxiliary variables with the survey variable(s).

In its pure form, this theory is model dependent. The validity of the resulting estimates depends on "the truth" of the model that the survey statistician ventured to assume. It was both a vulnerable theory, because the truth of the model can never be taken for granted, and a revolutionary theory, because it challenged the classical randomization theory, which had taken pride in delivering trustworthy inferences without any assumptions, valid whatever the form of the finite population.

Not surprisingly, the new ideas were at first controversial, albeit received by some as refreshing and vitalizing. Defenders of the classical design-based school received the new theory with a good deal of suspicion. Practitioners were at first hesitant to use methods that appeal more or less directly to "modeling".

Then came the theory called "model assisted (design-based) survey sampling". Capitalizing on advanced forms of modeled relationship among variables, it nevertheless preserved the precious design-based nature of the inferences. Model assisted design-based theory and methods became widely accepted and used in national statistical offices, especially fruitful in countries, such as Scandinavia, where reliable registers give ample supply of explanatory variables – those called "auxiliary" – for the model fit. The theory of calibration estimation of recent decades is a further outgrowth of this thinking.

8 Non-sampling aspects

The preface notes, somewhat apologetically: "Although non-sampling aspects of our subject, such as response errors, editing and imputation, are well recognized as of prime importance in the practice of survey work they have not always received the rigorous statistical treatment of topics such as sample design and estimation." A couple of the nine-teen articles do deal with non-sampling aspects.

A classical distinction in survey statistics was that between sampling error and non-sampling error. One can claim that a disproportionate part of research and published work focused on the former type, on methods to reduce that error with the aid of efficient sample design and advanced estimation theory. Much of this theoretical work was set in ideal conditions: the absence of nonresponse, measurement error and other imperfections labelled as non-sampling error. As some critics

maintained, it was a focus on “finding a (marginally) better estimator” under ideal conditions, with results often of limited use in practice; a display of “rigorous statistical treatment” in unrealistic settings.

9 Rigorous statistical treatment

When the preface uses this concept, it sounds as a self-evident obligation, a high ideal, for survey statistics to live up to. What does it require, today or in the future? How rigorous must the treatment be? Is it a question of mathematical rigor, or some other kind? For example, are all of the competing theories in survey science commensurate with the concept?

In one interpretation, the concept asks for a discourse where powerful theoretical tools can be brought to bear on the practical question, in a formal language and structure - as the word “rigorous” begs - rather than just a fleeting verbal discussion.

Also, “rigorous statistical treatment” obliges producers of statistics to keep users informed, in appropriate statistical measures, on the reliability, trustworthiness, and probable error of estimates. A part of this should be probability statements on accuracy - such as 95% confidence intervals - interpretable in one of the acclaimed theory frameworks, design-based, model-based, Bayesian or yet other.

But, as the preface hints, to genuinely accomplish this in the presence of the various non-sampling errors proved difficult. It is, somewhat paradoxically, an unresolved dilemma for the discipline.

Research in recent decades did try to make up for a perceived lack of “rigorous statistical treatment” with respect to non-sampling errors, but without any complete or decisive result.

In particular, much attention was devoted to one of those imperfections, the rapidly growing problem of survey nonresponse, and the bias it leads to in the estimates. An improved understanding of respondent motivation and behavior did help to reduce nonresponse at the data collection stage. At the estimation stage, advanced nonresponse bias adjustment methods helped to improve the quality of statistics produced.

Was the “rigorous treatment” of non-sampling error too much to ask? Will users do without this assurance in the future, and just accept declarations that “the numbers were produced with the best methodology we know”, without any further assurance of closeness of estimates to the truth.

The roots of the failure are traceable to the rigor imposed by randomization theory: its implicit obligation of one hundred percent response rate, and from precisely those in the designated probability sample. Unless “precisely those” are obtained, the theory is strictly speaking transgressed and compromised; estimates are biased.

It is a vulnerable theory, hardly made for the tough conditions of today’s survey climate. That the theory worked well, with some amendments, for as long as it did may surprise us now. Certainly, in the theory’s youth, some eighty or more years ago, nonresponse was low, or negligible, hardly worth worrying about. But today, the theory sidesteps a reality that surveys now face.

10 Theorization

The preface sees a “rigorous statistical treatment” as desirable, yet regrets that some published theoretical work may be seen as abstract, causing a “divide between theory and practice”. Realistic and convincing theorization can indeed pave the way for further advances, on non-sampling error as on sampling error.

How should research in survey science accomplish a proper balance between purely theoretical progress and a practical utility? It is a difficult question. Theorization can certainly be pursued for its own sake, without much chance of being applicable in practice, yet qualify as high caliber research from a technical point of view, as evidenced by hundreds of articles published in the last eighty years, many them “run-of-the-mill” papers, as will happen in what philosophers of science call “a period of normal science”.

The degree of theorization depends on the nature of the topic at hand. Some topics in survey statistics seem to strike researchers as more inviting, and more directly suited, for theorization and a certain abstraction. Small area estimation is of this kind. Other topics, although important for practice, seem to resist, or discourage, a desirable theorization. Non-sampling errors tended to be of this kind.

However, much valuable work on new important themes was addressed in the literature with little mathematical formalization, in a discourse framed in theory and concepts proper to other sciences, notably the behavioral sciences.

11 Subpopulations

"Small area estimation" became a prominent topic in recent decades, in a sense a unique topic. Such estimation has always resided in the realm of survey statistics, but only in the 1970's did the width of the question catch the attention of the theoreticians and other stakeholders in survey statistics.

An immediate challenge lies in the title: The words "small area" warn about a possible shortage of data from within the area itself, under any realistic cost frame for the entire survey. To overcome this dilemma by the development of new theory was the answer. Advanced theorization came quickly and is continuing, a beneficial playground for "the theoreticians".

The topic was a welcome and attractive avenue for survey science, which at times seemed rather short of challenges that could inspire both advanced theorization and deliver results of prime importance for practice.

"Small area" proliferates the idea of "a set of identifiable objects" as a target of inference. The set is not only finite, but smallish finite. "Small area" is to be understood in a wide sense, as the estimation for subpopulations, also called domains, contained within the greater population concept. Focus is usually not on one single domain, but rather on estimation for many, all contained in the large entire finite population, also a target of inference in the same survey.

Domains of interest are often placed in the geographical or administrative context of a country. Who can deny the need for accurate information for smallish but politically critical regions of a country? More generally, domains of interest can be any subgrouping of the well-defined population at large.

12 Recent trends and future prospects

Certain topics of importance did not set a clear mark on the jubilee volume of twenty-five years ago, understandably, since much of the work on those is relatively recent. Among them are *survey quality*, *total survey error*, *survey cost*, and *mixed data inputs* for statistics production. They deserve to be mentioned in these concluding sections.

Survey conditions changed, "for the worse" in the view of some, during the second twenty-five year period of The Association. The demise of the probability sampling paradigm has far-reaching consequences that will take time for the national statistical agencies to comply with and adjust to.

One must recognize that research in survey statistics is steeped in two different lines of scientific discourse. One is held in more or less formal mathematical language. The other progresses rather as a verbal discourse, framed in concepts and ideas proper to "survey generalists". To say that "the theoreticians" shape only the former type of discourse is not correct. The second type also draws on theory, but coming more likely from the behavioral sciences. Survey statistics is indeed interdisciplinary in character.

Behind this division lies the different educational backgrounds of survey statisticians. One stream is trained to work comfortably in a formalized mathematical idiom. Another is trained in concepts and ideas from fields with a less formalized structure. Both kinds contribute. The Association embraces both categories and encourages a fruitful co-operation between the two.

Survey Quality and Total Survey Error are areas that attracted considerable attention in the survey statistics literature of recent years. The latter has taken on the role of a conceptual framework, a

central organizing structure of the field of survey methodology. As such, it has filled an important need.

The elusive concept “quality”, especially “quality of survey statistics”, came to the fore and inspired much thought and drew much attention. Statisticians asked themselves: Survey quality, what is this? Quality, as subsequently argued, is a multi-faceted concept. Several national statistics offices spent considerable time to identify and elaborate their own vision of the essential dimensions, as many as six or more, of survey quality, so as to back up their mandate to provide valid numbers for the nation. Although no doubt helpful to some, those dimensions remain little more than “just names”, of limited value to many. They seem to escape attempts at synthesis, proper measurement and “rigorous statistical treatment”.

13 Challenges

One can claim that survey science is a science that is not free to act on its own behalf. It is always driven by extraneous conditions, notably the cost consideration. Statistics should be accurate and timely and relevant, but must not cost too much to produce. Quality statistics is a goal, naturally, but always subject to survey cost.

The jubilee volume does not dwell directly on the critical role of survey cost. Nevertheless, cost continues to be a driving force for survey theory and development.

The cost aspect is not new. It was important already a hundred years ago, when theoretical progress and empirical evidence finally convinced the statistical community that a complete enumeration of the country´s entire population was not necessary. National statistics of excellent – or at least sufficient – accuracy were obtainable at a reasonable cost with “just a sample”, more specifically a probability sample, just a modest fraction of the population. Public trust in such “low cost but accurate statistics” gradually developed.

For a long time, the statistical profession took pride in and thrived on this trust, on behalf of authorities and the general public, until, some decades later, high cost, high non-response and other non-sampling survey errors darkened the outlook.

High nonresponse rates in recent times made data collection from the units in the designated probability sample cumbersome, time consuming and delaying. Multiple attempts at contact with those particular units drove survey cost up. And even after a costly effort, remaining nonresponse bias plagues the estimates and has to be “adjusted for”, by a variety of suggested methods.

“The high cost syndrome” is to a degree responsible for the demise of the probability sampling paradigm. Meanwhile, data input from less expensive “alternative data inputs”, or “mixed data inputs”, have become ingredients that one cannot afford to disregard in building future theories for survey science, for “statistics at a reasonable cost”.

If probability sampling shall vanish from the survey science scene, the nonresponse problem in its original sense – the failure to get response from precisely those in the designated probability sample – will also disappear. On a positive note, this will liberate the discipline from burdensome or embarrassing chains. But how well the future data inputs shall represent the finite population is a matter that has to be addressed through other terms and concepts.

An open question is the validity of the statistics produced; how do we guarantee it? A “rigorous statistical treatment” seems mandatory. But what this attractive but elusive concept will require in the future is not clear. It is evolving over time, in tune with an evolving society and changing survey conditions. Nevertheless, the concept is an indispensable guideline for the discipline and the scientific community.

“Theorization” will take on new forms and directions. The sharp distinction between sampling error and non-sampling is expected to lose its contours. The fixation on theory development within the probability sampling paradigm is likely to disappear. Powerful new determinants will come from non-statistical considerations, such as survey cost.

Some prestigious sciences – such as physics – are capable of presenting to the world an admirable image of cumulative progress: On the basis of truths so far established, we find out more and more, we establish piece by piece further insights into a fixed – but highly complex – component of the universe, such as the atom. The universe that survey science is addressing is not unalterable, but a changing and evolving world.

14 Conclusion

The Association's Jubilee Volume, from about twenty-five years ago, was my guiding light in these notes. I commented on issues and lines of development that seemed important. Times are changing. If The Association were to now paint a portrait of the field that we call survey statistics, what would a new commemorative volume contain? Which contributions stand out today as the landmarks and breakthroughs that will guide the discipline into the future?

The author acknowledges the contribution of two anonymous reviewers of this article.