Leslie Kish’s Impact on Survey Statistics

Graham Kalton

Abstract

Leslie Kish, one of the pioneers of survey sampling, died on October 7, 2000, at the age of 90. This paper reviews his impact on survey statistics, mainly in terms of his research but also in terms of his promotion of sound probability sampling methods around the world. Kish’s research was broad-ranging, covering sampling methods, variance estimation and design effects, nonsampling errors, small area estimation, survey designs across time and space, and observational studies. He promoted probability sampling designs through consultancies in many countries, his writings, and in particular through the highly effective intensive summer Sampling Program for Foreign Statisticians that he established at the Survey Research Center of the University of Michigan.

Key Words: Sample design; Variance estimation; Nonsampling errors; Rolling samples.

1. Introduction

Leslie Kish, one of the pioneers of survey sampling, died on October 7, 2000, at the age of 90. During his long and productive career, he had a major impact on the field, achieved both through his impressive research contributions and through his extremely successful promotion of the use of scientific probability sampling methods throughout the world, and especially in developing countries. His wide-ranging research always focused on issues of practical importance, and his innovations facilitated the use of effective probability sampling in diverse areas. He promoted the practice of probability sampling through his expository writings (particularly for sociologists and demographers), through his numerous consultancies and advisory services, and through his training of survey statisticians, particularly those from developing countries.

This paper reviews Kish’s impact on survey statistics, primarily with respect to his contributions to the advancement of survey sampling and survey research more generally. It is useful to start with a brief account of his career in order to place these contributions in a temporal context. The interview of Kish in 1994 by Frankel and King (1996) is recommended for those interested in more details of Kish’s fascinating life. Some of the material in this paper is drawn from that interview.

Kish was born in 1910 in Poprad, which was then part of the Austro-Hungarian Empire and is now in Slovakia. In 1926, he emigrated to the United States with his family. When his father died the following year, he became a laboratory assistant at the Rockefeller Institute for Medical Research, while attending Bay Ridge Evening High School. He graduated from high school in 1930 and enrolled in the College of the City of New York night school, while continuing to work for 54 hours a week at the Rockefeller Institute. His interest in statistics arose out of his work at the Institute, and he studied on his own books by Fisher, Yule, Wallace and Snedecor, Tippett, Pearl, and others. In 1937, he interrupted his education to join the International Brigade to fight for the Loyalist cause in the Spanish Civil War. He returned to the United States in 1939 and earned a B.S. in Mathematics, cum laude, in that year. He was then hired by the U.S. Census Bureau as a Section Head, and subsequently moved to be a Statistician at the United States Department of Agriculture (USDA) Division of Program Surveys. In 1942, he left the Division of Program Surveys for war service, returning there in 1945 after the war. In 1947, he moved with a group of USDA colleagues headed by Rensis Likert to set up the Survey Research Center at the University of Michigan. He remained at the Survey Research Center until his retirement in 1981, when he became a Professor Emeritus. He remained fully active professionally until his death in 2000.

2. Research

At the start of Kish’s career, survey sampling was in its infancy. Much survey research was based on nonprobability samples. Methods for probability sampling were under development and many problems remained to be resolved. While at the USDA, Kish identified three important problems that he pursued at the Survey Research Center (SRC) in developing sampling methods there.

One of these problems was how to have an interviewer randomly select an individual within a sampled household. At the time, probability sampling methods for sampling households had been developed and were being applied in the Current Population Survey, but the CPS collected data on all members of sampled households, so that no selection of persons within households was needed. Kish invented a method for objective respondent selection and wrote it up in a memorandum. He was urged by his colleague Angus Campbell to submit the work for publication, and it resulted...
in the famous paper that was his first published research (Kish 1949). The widely used method is now known as the Kish selection table.

The second problem that Kish identified was counting nonresponse. He had to argue for counting and reporting nonresponse with probability samples against the concerns of colleagues who felt that to do so would put the SRC at a competitive disadvantage, particularly with organizations using nonprobability methods. He won his case and SRC adopted his approach, which is now fully accepted as standard practice.

The third problem was that of deep stratification. Standard stratification assumes independence of selections between strata, with the maximum number of strata possible being the number of selections. Particularly when the number of selections is small, as is often the case with the primary sampling units (PSUs) in a multistage design, it can be desirable to obtain greater balance in the sample than standard stratification permits. With Roe Goodman, Kish developed the technique of controlled selection that provides that greater balance by dropping the requirement of independence of selections between strata, while still retaining probability sampling (Goodman and Kish 1950).

Kish, who was always concerned to coin good names, preferred to call the technique ‘multiple stratification’, and he uses that term in his sampling text (Kish 1965a).

Kish’s subsequent research in survey statistics was wide-ranging, covering many aspects of survey sampling, nonsampling errors, small area estimation, survey designs across time and space, and observational studies. His many contributions have had a major impact on the development of the practice of survey sampling and of survey research more generally. The following paragraphs outline some of his contributions organized by topic.

Variance estimation. Before the 1970s, the analysis of survey data was severely limited by the analytic tools available, then mostly punch card equipment, such as counter-sorters and tabulators, and hand calculators. Thus, for example, weights – and particularly non-integer weights – were difficult to handle. For this reason Kish examined the use of uniform weights with the Kish selection table, even though unbiased estimation calls for weights proportional to the number of eligible household members.

As a result of the computational difficulties, prior to the 1970s sampling errors were rarely computed in a manner that reflected the complex sample designs typically employed in survey research. A widespread practice was to compute variances as if a simple random sample (SRS) had been drawn. Kish sought to promote the use of appropriate variance estimation methods by social researchers, which he did by illustrating the sizable underestimation that often arises when SRS formulas are applied to clustered samples (Kish 1957). Initially he developed and applied simple computational procedures, emphasizing the simplicity that can be obtained with a paired selection design in which two PSUs are sampled in each stratum (Kish and Hess 1959a; Kish 1968). He coined the term “design effect” for the ratio of the variance of a survey estimate for a given design to the variance of the same estimate obtained from a simple random sample of the same size. He made much use of this concept in his famous Survey Sampling book (Kish 1965a), which provides an encyclopedic treatment of practical survey sampling and is still widely read as a Wiley classic.

With the development of computers, Kish was quick to see their importance for variance estimation, and with SRC colleagues he developed an early Sampling Error Program Package (Kish, Frankel and Van Eck 1972). With his doctoral student Martin Frankel, he also extended the range of statistics for which sampling errors from complex sample designs could be computed (Kish and Frankel 1970, 1974). This highly influential research developed, applied, and evaluated balanced repeated replication (BRR) and jackknife repeated replication (JRR) methods of variance estimation. It also provided a definition of the population parameters estimated by analytical survey statistics in the finite population context.

Multipurpose surveys. The survey sampling literature deals mostly with an efficient sample design for estimating a single population parameter. Kish recognized the limitation of this approach since virtually all surveys are multipurpose in nature. He wrote several important papers dealing with multipurpose surveys, producing effective compromise designs that provide estimates not only for the population as a whole but also for various domains (Kish 1961b, 1969, 1976; Anderson, Kish and Cornell 1976; Kish and Anderson 1978; Kish 1980; Kish 1988). In recent years, he extended his interests to multipopulation surveys (e.g., Kish 1999, 2002).

Small area estimation. In considering the production of estimates for domains, Kish (1980, 1987a, 1987b) classified domains into major, minor, and mini domains and rare items. Estimates for major domains can be produced from a survey using standard sample-based estimators, particularly if the sample is designed to give sufficient domain sample sizes for this purpose. The sample sizes of most surveys preclude the production of estimates of adequate precision for minor or mini domains that comprise less than, say, one-tenth of the population. Yet, as Kish recognized early on, the demand for up-to-date estimates for small domains, particularly small geographical areas, would expand. This recognition led to his research in two related areas.

When a survey’s sample size is too small to produce small area sample-based estimates of adequate precision,
reliance may be placed on statistical models to produce indirect estimates. Much research on small area estimation techniques using this model-dependent approach has been conducted in recent years. In the 1970’s, Kish contributed to the development of the field through his direction of three doctoral dissertations at the University of Michigan (Ericksen 1973; Kalsbeek 1973; Purcell and Kish 1979, 1980).

Direct, or sample-based, estimates for small domains are sometimes possible. One obvious source of estimates for domains of any size is a population census, and indeed censuses are a major source of small domain estimates. However, data from a decennial census become out-of-date as the decade progresses. To address this problem, Kish proposed replacing the census by a rotating or rolling sample so that, by spreading the data collection over time, more up-to-date estimates can be produced. He first proposed such a procedure in 1979 (Kish 1979a,b), and wrote many papers on this topic after that (Kish 1981, 1983, 1986, 1990, 1997, 1998, 2002; Kish and Verma 1986), including the issue of how to cumulate sample data over time (Kish 1999). In another paper in this volume, Charles Alexander (2002) provides a detailed review of Kish’s work on this topic and its influence on the large-scale continuous survey, the American Community Survey, that the U.S. Census Bureau plans to introduce to replace the long form in the 2010 Census.

Special sample design problems. During the course of his work, Kish encountered a number of specialized sampling problems that often occur and he offered some efficient solutions. The areas to which he contributed include the following:

- **Sampling rare and elusive populations.** One of the most challenging design tasks faced by sampling statisticians is constructing an efficient sample design for a rare or elusive population (such as persons with a rare illness or the homeless). Kish (1965b, 1991) provides insightful reviews of methods for tackling this type of problem.

- **Maximizing overlap.** When a population is sampled repeatedly over time, the issue arises of how to control the sample overlap between one round and the next. A particular example occurs when a master sample of PSUs is used and needs to be updated when new census data become available. Frequently it is desirable to maximize the overlap in the sample of PSUs, while updating measures of size and changing the stratification to reflect current data. Kish and Scott (1971) provide a relatively simple and effective method of satisfying these requirements.

- **Sampling organizations of unequal size.** Some surveys are designed to produce estimates for units at different levels, for instance, for hospitals and for patients. When hospitals vary considerably in their numbers of patients, a design conflict arises between the production of efficient hospital- and patient-level estimates. Kish (1965c) examines this problem and clarifies the issues involved.

**Nonsampling errors.** Kish clearly recognized the harmful effects that nonsampling errors can have on the quality of survey estimates. Early in his career he collaborated with Jack Lansing to investigate the response errors in respondents’ reports of the values of their homes by comparing these reports with estimates made by professional appraisers (Kish and Lansing 1954). In his studies of interviewer variance, he took advantage of the theory on cluster sampling, measuring interviewer variance with the intra-class correlation coefficient, and determining the optimum number of interviews per interviewer based on a simple cluster sample cost model (Kish 1962). With Irene Hess, he conducted a study of noncoverage in area samples of dwelling units. The study was stimulated by a 10 percent noncoverage rate in SRC surveys at that time, and led to improvements that reduced this rate to about 3 percent (Kish and Hess 1958). Also with Irene Hess, he introduced an imaginative replacement procedure for noncontacts in one survey by substituting noncontacts from a previous, similar, survey (Kish and Hess 1959b). For stochastic imputation schemes, Kish was an early proponent of replicating the imputations to reduce imputation variance, in what he termed a repeated replication imputation procedure (RRIP) and what is now known as fractional imputation (Kalton and Kish 1984).

**Observational studies.** Early in his career, Kish (1959) wrote a widely cited paper on the design of studies to investigate causal relationships, particularly nonrandomized studies. In his writing about this topic he made use of his survey sampling expertise as, for instance, in the relationship between stratification and matching (Anderson, Kish and Cornell 1980). His work developed into his book *Statistical Design for Research* (Kish 1987a) in which he compared surveys, experiments, and observational studies for investigating causal effects in terms of the three R’s: realism, randomization and representativeness (see also Kish 1975). He also made clear the importance of assessing both bias and variance in assessing the ability of different study designs to measure causal effects, rather than concentrating on bias as had been common in the literature on this topic.

### 3. Other Contributions

Kish’s seminal and wide-ranging contributions to the methodology of survey statistics are of great importance. Yet of possibly even greater importance are his contributions to the promotion of the use of sound probability sampling methods around the world.
Kish’s writings, of course, contributed to the current widespread use of probability sampling methods by emphasizing good practical methods. His three books Survey Sampling (Kish 1965a), Statistical Design for Research (Kish 1987a), and Sampling Methods for Agricultural Surveys (Kish 1989) are all extremely valuable in this respect, as are his expository writings for social scientists.

Kish had a long-standing dedication to assisting developing and transition countries, and that can be seen in many of his activities. He was a sampling consultant to the World Fertility Survey from 1973 to 1983 and he consulted in many countries, he ran a training program for foreign statisticians, and he wrote specifically for statisticians in developing countries. Sampling Methods for Agricultural Surveys was, for instance, written for the FAO, particularly for use in developing countries. He contributed a Questions/Answers column for the Survey Statistician, the newsletter of the International Association of Survey Statisticians, from 1978 to 1994. In that column he provided sound advice on many practical sampling problems that frequently arise but that are not well addressed in the literature. The column was considered so useful that the IASS published the full set of questions and answers in a special volume (Kish 1995b).

Kish was rightly particularly proud of the intensive two-month summer Sampling Program for Foreign Statisticians that he established at the Survey Research Center in 1961. The SPFS has now trained more than 500 survey statisticians from 105 countries. It is significant that Kish chose “Developing samplers for developing countries” as the topic for his 1994 Morris Hansen Memorial Lecture (Kish 1996). To help maintain this important program, the Leslie Kish International Fellows Fund was established at the University of Michigan at a celebration of Kish’s 90th birthday. Of all his accomplishments, the SPFS was the one that gave him greatest pleasure.

4. Concluding Remarks

Leslie Kish is a giant in the field of survey sampling. His contributions were enormous and recognized by many honors. These honors included, among others, President of the International Association of Survey Statisticians in 1983-85, President of the American Statistical Association in 1978 (see Kish 1978, for his Presidential address on “Chance, Statistics and Statisticians”), Honorary Fellow of the International Statistical Institute, Honorary Fellow of the Royal Statistical Society, Honorary Member of the Hungarian Academy of Sciences, Fellow of the American Association for the Advancement of Science, Fellow of the American Academy of Arts and Sciences, recipient of the American Statistical Association’s Samuel L. Wilks Award in 1997, recipient of the Mindel Shep Award from the Population Association of America in 1998, recipient of the Methodology Award from the American Sociological Association in 1989, and honorary degrees from the University of Bologna, the Athens University of Economics and Business, and the Eotvos Lorand University in Budapest.

Yet Kish remained down-to-earth, approachable by all. He had a great enthusiasm for many subjects including sport, art, literature, politics, philosophy, and science. He was always concerned with improving the conditions of the world’s population. He was particularly interested in young people and one of his favorite sayings was “Keep young by being curious, and have young friends”. Undoubtedly his endearing personality played an important part in his great success in promoting sound sampling methods around the world. Ivan Fellegi’s excellent obituary in Survey Methodology was aptly titled “Leslie Kish – A Life of Giving” (Fellegi 2000). Kish gave so much personally to so many people and so much professionally to the development of survey statistics.

References


