

TRAINING NEEDS FOR SURVEY STATISTICIANS IN DEVELOPED AND DEVELOPING COUNTRIES

Graham Kalton, Westat

Westat, 1650 Research Boulevard, Rockville, Maryland 20850

Key Words: Survey sampling, Master's degrees, short courses, continuing education

1. Introduction

The session in which this paper is presented is entitled "Training of Survey Statisticians: A Session Dedicated to the Memory of Leslie Kish." The topic of the session is highly appropriate given Kish's lifelong dedication to training survey statisticians, especially those from developing countries. It is also appropriate that the session is cosponsored by the International Association of Survey Statisticians (IASS) and by four ASA sections—Health Policy Statistics, Survey Research Methods, Government Statistics, and Social Statistics. Kish was a key member of the committee that founded the IASS, and he served as President of both IASS and ASA.

Kish established a major program for training survey statisticians from developing countries at the Survey Research Center of the University of Michigan in 1961. That highly successful two-month summer Sampling Program for Survey Statisticians (SPSS) continues to this day and has now trained more than 500 survey statisticians from more than 100 countries. Kish's choice of "Developing Samplers for Developing Countries" as his topic for the fourth Morris H. Hansen Memorial Lecture is an indicator of the importance that he attached to this activity (Kish, 1996). His paper contains many valuable observations on the training needs of survey statisticians in developing countries.

This paper considers the training needs for survey statisticians in both developed and developing countries. The term "survey statistician" refers in this paper to those engaged in the statistical design and analysis of surveys, primarily survey samplers, rather than to the broader interpretation of the term that encompasses those engaged in any aspect of survey methodology. For those in developed countries, the focus here is on the education needed to prepare them for careers as survey statisticians, mainly in terms of a master's degree. For those from developing countries, the focus is on short-term programs that can provide applied training in survey sampling, programs like the SPSS and shorter programs. Throughout the emphasis is on training in applied sampling rather than in sampling theory.

Educational needs for survey statisticians in developed countries are discussed in Section 3, and the

training needs for survey statisticians in developing countries are discussed in Section 4. As preparation for the discussion in those sections, Section 2 contains observations about general training needs in survey research.

2. Training in Survey Research Methods

Survey research is by its nature an interdisciplinary activity. It is generally conducted as a team operation, with the team members providing expertise in the survey's subject matter, questionnaire design, survey sampling, survey operations, data collection, computing, data processing, and statistical analysis. In large-scale survey organizations, the survey team may be composed of specialists in each of the separate components of the survey procedures. In small-scale survey organizations, however, team members may need to cover more than one component. For example, the team member who designs the questionnaire may also manage the data collection, and the team member who selects the sample may also direct the data processing and survey analysis. Regardless of whether team members have responsibility for one or more than one of the survey components, they need to have a basic understanding of all components. Designing and implementing an effective survey requires a careful blending of its various components. Thus the members of the survey team need to be able to work together to formulate a survey design that can be implemented successfully. They should fashion their individual components to be the best in the context of an overall design, rather than optimal in some narrower sense. To interact effectively with the other members of the survey team, each member needs to appreciate the issues involved in the overall design of the survey.

An alternative way of demonstrating this need for survey team members to understand the whole survey process is through the widely used concept of total survey error. Kish (1965, Chapter 13) describes the diverse sources of error in survey estimates, including sampling error and the nonsampling errors that arise through nonobservation (noncoverage and nonresponse) and through observation (data collection and processing errors). Survey design often requires trade-offs to be made between different sources of error. For instance, a larger sample size will reduce sampling error but perhaps at the cost of more error from nonresponse and data collection, or a more burdensome data collection protocol may lower

measurement error at the cost of greater nonresponse error. All members of a survey team need to appreciate such interrelationships. Thus, an important part of training for all survey researchers should be to show them how their specialist area fits into the big picture.

There is a notable lack of educational and training programs in survey research around the world. Few universities provide, or are equipped to provide, these programs. They seldom have the full range of faculty needed, and most lack instructors with the wealth of experience required to teach the applied aspects of survey research effectively. In recognition of the shortage of such programs in the United States, in 1992 the Federal statistical agency heads, the head of the Office of Management and Budget Statistical Policy Office, and the chair of the Council of Economic Advisors initiated the establishment of a center for survey methods that would provide graduate education and research to serve the needs of the Federal agencies. As part of the development of that initiative, the Committee on National Statistics (CNSTAT) at the National Research Council convened a workshop on the proposed center. The summary of the workshop (Wolf, 1992) provides many important ideas about the training needs for survey researchers from the perspectives both of those working in survey research in Federal agencies and of faculty involved in teaching survey research. Here are a few of those ideas:

- There should be a focus on the interdisciplinary nature of survey research and training for it, along the lines discussed earlier.
- The center should engage in both degree and nondegree programs.
- Training could include core courses, specialty courses, and electives.
- Apprenticeships or practicums are important in providing real-world experience.
- A mechanism to respond to new training needs will be necessary.
- A wide variety of courses might be part of the center's curricula.

The center was established at the University of Maryland in 1992 after an open competitive process. It involves a collaboration of the University of Maryland, the University of Michigan, and Westat and is known as the Joint Program in Survey Methodology (JPSM). The ways that JPSM has instituted the ideas given earlier are on its web site: <http://www.jpsm.umd.edu>. The JPSM has set up PhD and MS degree programs, Citation and Graduate Certificate nondegree programs, and noncredit short courses. The JPSM MS degree program requires 46 credit hours. It has two areas of

concentration, statistical science and social science, with all students taking certain core courses.

It is instructive to make note of the collaboration involved in JPSM. The three-way collaboration brings together the range of faculty and experience needed to run the program and, in addition, the program is supported by teaching contributions from staff in the Federal agencies. Collaborations in training in survey research between a university, a survey research organization, and government also occur in the United Kingdom. The Department of Social Statistics at the University of Southampton offers Diploma/MSc programs in Social Statistics and in Official Statistics, the latter offered in collaboration with the Office for National Statistics. The department also collaborates with a survey research organization—the National Centre for Social Research—and the University of Surrey in the Centre for Applied Social Surveys (CASS), which offers short courses in survey methods. See <http://www.socstats.soton.ac.uk> for details.

3. Education for Survey Statisticians

This section focuses on graduate-level education for survey statisticians, mainly survey samplers and, in particular, on the issues involved in setting a curriculum for a master's level program. Like Kish (1996), the focus here is on survey sampling and sample design, not the foundations of sampling theory.

In general, also like Kish, I favor a flexible program to accommodate the needs of a variety of students. Some may want to concentrate specifically on survey sampling, others may want to take a more general program in survey methodology with a special emphasis on sampling, while yet others may want to combine survey sampling with a particular subject-matter area (e.g., agriculture, education, health, economics). Some may plan to move into managerial positions as their careers progress, whereas others may want an education that prepares them for a lifetime career in survey sampling. The MS education for this latter group needs to provide a firm foundation in the subject on which they can build as new developments occur.

My criteria for a master's program for survey statistics are the following:

- Strong foundation in statistical theory and methods,
- Specialized training in the theory and practice of survey sampling,
- Broad knowledge of the whole survey process,
- Development of good computing skills, and

- Training in oral and written communication.

Statistical theory and methods are listed first because they provide the basic foundation for survey sampling. Students need a strong foundation in this area to be able to read the current sampling literature. They will no doubt have even more need for this foundation to upgrade their skills as advances in survey sampling occur in the future. As an illustration, consider the recent developments in small area estimation and in variance estimation with data sets containing imputed values. Both of these areas illustrate how new methods in survey sampling are using more high-powered statistical methods. A strong foundation in statistical theory and methods today requires at least four courses (two in probability and statistical theory and two in statistical methods), and that already seems inadequate to provide the foundation needed to cover the newer techniques in survey sampling. To prepare for the future, more courses are likely to be needed.

Survey sampling is now a highly developed field with a rich literature and a wide range of methods. A practicing survey statistician needs at least three basic courses in this area (e.g., applied sampling methods, basic theory, and variance estimation), and then there are many special topics that should ideally be covered in some fashion. The following is a partial list of important topics that cannot be adequately covered, if mentioned at all, in the basic courses:

- Nonresponse and calibration weighting adjustments
- Edit and imputation methods
- Design and estimation for surveys over time
- Measurement error models
- Sampling rare populations
- Small area estimation
- Sampling for establishment surveys
- Design and analysis of observational studies
- Disclosure avoidance
- Treatment of outliers
- Sample coordination over time

A selection of these topics can appear in a fourth sampling course, but not all of them can be covered, and those that are covered cannot be treated in depth. Several of these topics warrant one (or even more) three-credit courses for a full treatment, and others warrant one or two credits. Additional courses could be considered, but they are likely to run up against the constraints on the acceptable maximum number of credit hours for an MS program. With only one course on these topics, a difficult choice has to be made between whether to attempt to make the students aware of all the areas of survey sampling methods that they

may need in their work, or whether to let them choose a small number that they will study in somewhat more depth.

The justification for the need for all survey researchers to acquire a broad knowledge of the whole survey process has been made in the previous section. That justification clearly applies to survey statisticians, who must design samples that meld with the other survey design features in an effective manner. A broad knowledge of the survey process requires at least two or three 3-credit courses (e.g., data collection methods, sources of error in surveys), and one or two more courses if a survey practicum is included. Carefully chosen apprenticeships that give exposure to the whole survey process may also serve the important requirement of giving students an appreciation of the overall design considerations that apply in survey practice.

The need for computing skills, particularly the use of statistical software packages, is clear. Computing features in most modern courses in statistical methods, and it can also be included in relevant sampling courses. That should in general suffice, provided that attention is given to computing in course development.

A widely acknowledged weakness of much statistical training is a lack of attention to oral and written communication. Strong communication skills are essential for applied statisticians in general, and this is certainly so for survey statisticians. In a paper on preparing statisticians for careers in the Federal Government, an ASA Committee cited the need for communication skills: "The need for writing skills cannot be overemphasized. As statisticians move into managerial positions, communication—both written and oral—becomes a vital ingredient of a successful performance." (Eldridge et al., 1982, p. 76). A paper entitled "Preparing Statisticians for a Career at Statistics Canada," by Denis et al. (2002) contains recommendations for universities in preparing students to work in a statistical agency that are in line with the criteria proposed earlier, including again a call for writing skills:

- "Provide students with a solid grounding in statistics";
- "Give students the opportunity to work on "real" problems";
- "Generate interest in learning to use statistical software"; and
- "Put more emphasis on writing skills".

Training in oral and written communication can best be achieved by requiring presentations and reports in many courses. Assignments of 15-minute presentations supported by well-prepared slides in

several classes can help students develop their oral communication skills and prepare them for making presentations at professional meetings. Training in report writing can be obtained through course assignments, but to be effective, instructors or writing experts must spend adequate time on the writing rather than on the technical aspects of the reports. In some cases, a separate course on the principles of report writing may be advisable.

The challenge in developing an MS program for survey statisticians is to accommodate all these objectives within a reasonable course load. Given the unavoidable constraints on the length of an MS program, some priorities need to be established. The initial set of priorities noted earlier is a solid grounding in statistical theory and methods (at least four to six courses), an extensive coverage of sampling methods (at least four courses), and a basic overview of the whole survey process (at least two courses). After that, I favor flexibility to allow choices of additional topics in any of these areas or in other relevant areas, such as a substantive area of application. As a rule, students wishing to embark on careers as specialist samplers and those wishing to pursue PhD degrees in survey sampling would be advised to elect more courses in statistical theory and methods and in topics in survey sampling.

However an MS course is structured, it cannot be expected to provide a detailed coverage of all the survey sampling topics that a survey statistician will encounter at work. Also, as the field of survey sampling continues to advance, new methods will emerge. As a consequence, I see a strong need for continuing education programs for survey statisticians to develop their range of skills. Advanced short courses are helpful in exposing participants to specialized techniques, but they lack the means to ensure that the participants undertake the supplementary study needed to fully master the material. I believe that longer term formal courses, with assignments and examinations, are required, together with some form of certification for successful performance. This kind of continuing education, which is required in some other disciplines, is now needed for survey sampling. Hopefully employers will recognize the benefits of such continuing education, encourage staff to participate, and reward those who complete the courses successfully. It is of interest to note that Statistics Canada invests a good deal in professional training. It has established its own Training Institute to offer courses in-house, and it expects supervisors and employees to discuss a training plan for the coming year during the annual performance meeting (Denis et al., 2002).

4. Training Courses for Survey Statisticians from Developing Countries

This section addresses the training needs for survey statisticians from developing countries. The MS programs described earlier can serve the educational needs for survey statisticians from developing countries well. However, it is expensive and time-consuming to send persons from developing countries to MS programs abroad. Alternative shorter training programs are, therefore, much needed.

One model for such a training program is the SPSS that Kish developed over 40 years ago, as mentioned earlier. Another long-term program for training staff from national statistical offices is that run by the International Program Center at the U.S. Census Bureau. The center now runs workshops from two to six weeks in length on a variety of topics, including survey sampling (<http://www.census.gov/ipc>).

The IASS has successfully run short courses in survey sampling and other aspects of survey research in association with the biennial sessions of the International Statistical Institute for many years. Kish initiated short half-day meetings in 1973, and these were expanded into longer workshops in 1987. At the ISI session in Seoul in 2001, the IASS offered the following courses: Workshop on Survey Sampling, Variance Estimation in Complex Surveys, Introduction to Small Area Estimation, Nonsampling Error Research, and Edit and Imputation of Survey Data. The workshop on survey sampling is specifically designed for survey statisticians from developing countries. Together with the course on variance estimation, it provides an overall introduction to the subject.

Courses like those offered by the IASS serve a valuable role both by providing training in survey sampling and also by bringing together survey statisticians from different countries to share their experiences. Such courses should emphasize survey sampling practice and introduce basic theory only to the extent necessary. They should cover basic sampling techniques (e.g., systematic sampling, stratification, multistage sampling, PPS selection) and also nonresponse, noncoverage, weighting, and the use of software for variance estimation. They should also include illustrations from developing countries. Simple exercises can be helpful in reinforcing key points being made. Where possible, hands-on use of computer software should be included.

The key problem with providing training courses for survey statisticians from developing countries, whether longer courses like the SPSS or shorter courses like those of the IASS, is that of funding. The SPSS was started with funds from the Ford Foundation and now is

supported in part by the Leslie Kish International Fellows Fund. Many statisticians from developing countries have been able to attend the IASS courses with fellowship support from the UN Statistical Division. In both cases these sources of funding support have provided the core basis for maintaining the courses, with other support coming from various other sources.

When training persons from different countries, language becomes an important issue. The above courses are generally given in English and require participants to understand English. However, at least when there is only one other major language in addition to the language of instruction, a translation of presentation slides—or at least key terms—into the other language can be valuable. The use of two projectors for the two languages can then help those who are not so fluent in the language of instruction.

A type of training program that seems particularly desirable is one that is given in a developing country for statisticians in that country and neighboring countries. A one- or two-week regional program could be a very effective use of limited resources. In his invited discussion of this paper, Hermann Habermann suggested an enhancement to this scheme that would extend the program to a series of training activities over a period of several years. I thoroughly endorse this extension, which I see as a way to develop the internal training resources in the region, with local instructors taking over in later years.

In earlier years the World Fertility Survey directed by Sir Maurice Kendall for the International Statistical Institute and the UN Statistical Division's National Household Survey Capability Program made major contributions to establishing survey infrastructures in many developing countries. Since that time, a sizable number of international survey programs

have begun, but unfortunately none of them has a training mission in survey statistics or survey methodology. Much would be gained if they could be persuaded to devote small proportions of their budgets to training and capacity building.

The programs listed in this section play a very valuable role in training survey statisticians in developing countries. Yet much more needs to be done. Might ASA and IASS collaborate in fostering such training programs? Both associations are concerned with improving the use of statistics, and efforts in this area would serve that objective well.

5. References

- Denis, J., Dolson, D., Dufour, J., and Whitridge, P. (2002), "Preparing Statisticians for a Career at Statistics Canada," Working Paper no. HSMD-2001-006E/F, Ottawa: Statistics Canada.
- Eldridge, M.D., Wallman, K.K., Wulfsberg, R.M., Bailar, B.A., Bishop, Y.M., Kibler, W.E., Orleans, B.S., Rice, D.P., Schaible, W., Selig, S.M., and Sirken, M.G. (1982), "Preparing Statisticians for Careers in the Federal Government: Report of the ASA Section on Statistical Education Committee on Training of Statisticians for Government," *American Statistician*, 36, 2, 69-89.
- Kish, L. (1965), *Survey Sampling*, New York: Wiley.
- Kish, L. (1996), "Developing Samplers for Developing Countries", *International Statistical Review*, 64, 143-162.
- Wolf, F.E. (ed.) (1992), *Center for Survey Methods: Summary of a Workshop*, Washington D.C.: National Academy Press.