Introduction

Vijay Verma
University of Siena
Siena, Italy

1. Section A of this publication provided a comprehensive introduction to major technical issues in the design and implementation of household surveys. Apart from questionnaire design, it gave an overview of survey implementation and sample design issues. The present section addresses, in more specific terms, selected issues related to the design of samples for household surveys in the context of developing and transition countries. It contains three chapters, one chapter on the design of master sampling frames and master samples for household surveys, and two chapters concerning the estimation of design effects and their use in the design of samples.

2. The objective of a sample survey is to make estimates or inferences of general applicability for a study population, derived from observations made on a limited number (a sample) of units in the population. This process is subject to various types of errors arising from diverse sources. Usually a distinction is made between sampling and non-sampling errors. However, from the perspective of the whole survey process, a more fundamental categorization distinguishes between "errors in measurement" and "errors in estimation". Errors in measurement, which arise when what is measured on the units included in the survey depart from the actual (true) values for those units, concern the accuracy of measurement at the level of individual units enumerated in the survey, and centre on the substantive content of the survey. They are distinguished from errors in estimation which arise in the process of extrapolation from the particular units enumerated to the entire study population for which estimates or inferences are required. Errors in estimation, which concern generalizability from the units observed to the target population, centre on the process of sample design and implementation. These errors include, apart from sampling variability, various biases associated with sample selection and with survey implementation, such as coverage and non-response errors. All these errors are of basic concern to the sampling statistician. Often, several surveys or survey rounds share a common sampling frame, master sample, sample design, and sometimes even a common sample of units. In such situations, errors relating to the sampling process tend to be common to these surveys, and less dependent on the subject matter.

3. It is this distinction between measurement and estimation that informs the selection of the issues covered in this section. The chapters in section B address two important aspects of estimation: the sampling frame, which determines how well the population of interest is covered and influences the cost and efficiency of the sampling designs that can be constructed; and design effect, which provides a quantitative measure of that efficiency and can help in relating the structure of the design to survey costs. There are of course other aspects of the design and it would, therefore, be useful to study the chapters of this section with reference to the framework developed in the preceding section, in particular the discussion of basic principles and methods of sample design presented in chapter II.

4. Chapter V discusses in great practical detail the concepts of a master sample and a master sampling frame. The definition of the population to which the sample results are to be

generalized is a fundamental aspect of survey planning and design. The population to be surveyed then has to be represented in a physical form from which samples of the required type can be selected. A sampling frame is such a representation. In the simplest case, the frame is merely an explicit list of all units in the population; with more complex designs, the representation in the frame may be partly implicit, but still accounts for all the units. In practice, the required frame is defined in relation to the required structure of the samples and the procedure for selecting them. In multistage frames, which for household surveys are mostly area-based, the durability of the frame declines as we move down the hierarchy of the units. At one end, the primary sampling frame represents a major investment for long-term use. At the other end, the lists of ultimate units (such as addresses, households and, especially, persons) require frequent updating.

5. The frame for the first stage of sampling (called the primary sampling frame) has to cover the entire population of primary sampling units (PSUs). Following the first stage of selection, the list of units at any lower stage is required only within the higher-stage units selected at the preceding stage. For economy and convenience, one or more stages of this task may be combined or shared among a number of surveys. The sample resulting from the shared stages is called a master sample. The objective is to provide a common sample of units down to a certain stage, from which further sampling can be carried out to serve individual surveys. The objectives in using a master sample include the following:

(a) To economise, by sharing between different surveys, on costs of developing and maintaining sampling frames and materials;

(b) To reduce the cost of sample design and selection;

(c) To simplify the technical process of drawing individual samples;

(d) To facilitate substantive as well as operational linkages between different surveys, in particular successive rounds of a continuing survey;

(e) To facilitate, as well as restrict and control as necessary, the drawing of multiple samples for various surveys from the same frame.

6. It is also important to recognize that, in practice, master samples also have their limitations:

(a) The saving in cost can be small when the master sample concept cannot be extended to lower stages of sampling, where the units involved are less stable and the corresponding frames or lists need frequent updating;

(b) Reasonable saving can be obtained only if the master sample is used for more than one, and preferably many, surveys;

(c) The effective use of a master sample requires long-term planning, which is not easily achieved in the circumstances of developing countries;

(d) The lack of flexibility in designing individual surveys to fit a common master sample can be a problem;

(e) There can be increased technical complexity involved in drawing individual samples; in any case, there is need for detailed and accurate maintenance of documentation on a master sample.

7. It is also possible to extend the idea of a master sample to include not a sample, but the entire population, of PSUs. This is the concept of a master sampling frame discussed in chapter V. The investment in a master sampling frame is worthwhile when available frame(s) do not cover the population of interest fully and/or do not contain information for the selection of samples efficiently and easily. The use of a master sampling frame also ameliorates the constraints on the type and size of samples that can be selected from a more restricted master sample.

8. Chapters VI and VII deal with the important concept of the design effect. The design effect (or its square root, which is sometimes called the design factor) is a comprehensive summary measure of the effect on the variance of an estimate, of various complexities in the design. It is computed, for a given statistic, as the ratio of its variance under the actual design, to what that variance would have been under a simple random sample (SRS) of the same size. In this manner, it provides a measure of efficiency of the design. By taking the ratio of the actual to the SRS variance, the design effect also removes the effect of factors common to both, such as size of the estimate and scale of measurement, population variance and overall sample size. This makes the measure more "portable" from one situation (survey, design) to another. These two characteristics of the design effect -- as a summary measure and as a portable measure of design efficiency -- contribute to the great usefulness and widespread use of the measure in practical survey work. Computing and analysing design effects for many statistics, as well as for estimates over diverse subpopulations, are invaluable for the evaluation of the present designs and for the design of new samples.

9. Although it does remove some important sources of variation in the magnitude of sampling error mentioned above, the magnitude of the design effect is still dependent on other features of the design such as the number and manner of selection of households or persons within sample areas. Above all, it is important to remember that design effects are specific to the variable or statistic concerned. There is no single design effect describing the sampling efficiency of "the" design. For the same design, different types of variables and statistics may (and often do) have very different values of design effect, as do different estimates of the same variable over different subpopulations. Such diversity of design effect values across and within surveys is illustrated from the range of empirical results, covering different types of variables from 10 surveys in 6 countries, presented in chapter VII.