Expert Group Meeting to
Review the Draft Handbook on
Designing of Household Sample Surveys
3-5 December 2003

Guidelines to authors[*]

---

[*]   This document is being issued without formal editing.

Guidelines to authors

Introduction

Sample surveys are increasingly becoming one of the major sources of detailed social and demographic statistics in many countries. Such surveys are mostly based on multi-stage sample designs. For many developing countries, however, such surveys are not efficiently designed thereby rendering their results unreliable. There is a shortage of sampling statisticians and practical publications on sample survey design that could be easily accessible and used by statisticians in designing good probability samples in these countries. This publication is intended to satisfy this need. This chapter will include:

1. Purpose/objective of handbook
2. Organization of handbook
3. Acknowledgements

I. Sources of socio-demographic data

Experience has shown there are a number of sources of micro- data on social and other statistics. The main sources are censuses, administrative records and surveys. While there is no doubt that all these sources are essential, their complementarities as data sources are rarely brought out. This chapter will briefly present the three sources and highlight existing complementarities and limitations.

1.1. Population census
1.2. Administrative records
1.3. Differences and complementarities

II. Sampling strategies

Most sample designs pertaining to household surveys, carried out in developing countries, are complex because they are multi-stage stratified and clustered. It is for this reason that this handbook focuses on multi-stage sampling strategies. Basic review of sampling techniques will be presented in the annex.

2.1. Reasons for selecting samples at more than one stage (multi-stage sampling)
2.2 Selecting samples in two or more stages
2.3. Primary sampling units (PSUs) -characteristics
2.4. Stratification
2.5. Sample size determination
2.6. Probability proportional to size (PPS) sampling
2.7. Sampling of PSUs
2.8. Sampling of households (use of implicit stratification)
2.9. Estimating sampling errors

III. Sampling frames and master samples: features and maintenance

A. Sampling frames

Faulty frames are a common source of error in household surveys for instance coverage error. There is, therefore, a need to elaborate best practices in frame construction and usage taking into account various stages of sampling. This publication will cover issues on frames related to multi-stage sample design covering the following:

3.1. Survey population
3.2. Ideal properties of frames
3.3. Area frames
3.4. List frames
3.5. Multiple frames
3.6. Master sampling frames
3.7. Common problems of frames and suggested remedies

B. Master samples

Master samples can be cost effective and efficient, however if wrongly designed and maintained could generate unreliable results. It is against this background that issues listed below will be covered in the handbook:

3.8. Key features of a master sample
3.9. Ideal characteristics of a master sample of PSUs
3.10. Selecting of first stage sampling units as a master sample
3.11. Characteristics of PSUs ideal to support several household surveys
3.12. Allocation across domains (administrative regions, etc.)
3.13. Multiple use of master samples to support surveys
3.14. Maintenance and updating of master samples
3.15. Rotation of master samples of PSUs

IV. Documentation and evaluation of sample designs

Metadata are rarely documented in many survey worksheets and reports. In this way errors creep in into survey analysis because, for example probabilities of selection may not be fully known at the time of analysis. The handbook will, therefore, highlight the importance of keeping detailed records of metadata. Specifically it will cover:

4.1. Need for documentation and evaluation
4.2. Labels for design variables, strata, PSUs and case identification particulars
4.3. Selection probabilities, response rates and coverage rates at various stages of sample selection
4.4. Weights
4.5. Information on costs
4.6. Features of data files

V. Construction of sample weights

Biased results from many sample surveys are partly due to lack of weighting or inappropriate use of sampling weights. The handbook will address this problem by illustrating how weights are derived and used. This will include weights based on selection probabilities, none response and undercoverage. There will be detailed coverage of:

5.1. Need for weights
5.2. Base weights
5.3. Weighting for non-response
5.4. Weighting for non-coverage/post-stratification
5.5. Self-weighting designs
5.6. The use of weights in estimation
5.7. Domain estimation

VI. Estimation of sampling errors for complex survey data

Some developing countries do not calculate or publish sampling errors, because of faulty designs, and lack of awareness and knowledge of software packages that are available for use in calculating sampling errors. The handbook will cover the following topics:

6.1. Overview of variance estimation methods for complex samples
6.2. Software for sampling error estimation
6.3. Basic requirements of software
6.4. Comparison of sampling error estimation software

VII. Nonsampling errors in surveys

While sampling errors are measured and controlled through relevant sample designs non sampling errors are not given much attention in household surveys in many developing countries. This publication will address this problem by discussing, in appreciable detail issues related to:

7.1. Total survey error
7.2. Coverage errors
7.3. Nonresponse
7.4. Measurement/ response error

VIII. Data processing, analysis and dissemination

This chapter will describe how the "raw" data collected from a household survey should be processed before it can be subjected to any useful analysis. Data processing is presented as a multi-stage process involving pre-processing activities, coding and editing, consistency checks, preparation of data files and databases, and the evaluation of various errors committed during the

data collection process. Other topics to be covered include basic summarization, tabular analysis, presentation, and dissemination of survey data. Specific areas to be covered include:

    8.1.   Development of databases
    8.2.   Data editing, coding and data capture
    8.3.   Tabulation and production of basic tables
    8.4.   Data analysis and report preparation
    8.5.   Data dissemination.

IX.   Examples of surveys carried out in some developing countries

There is need to bring out the unique sample design features of each of the surveys listed below despite having many common features. Attempts will be made to collect materials from and for different countries with respect to:

    9.1.   Demographic and health surveys [cite technical report here]
    9.2.   Labor force surveys
    9.3.   Household income and expenditure surveys
    9.4.   Time use surveys
    9.5.   Informal sector surveys
    9.6.   LSMS (World Bank) surveys [cite technical report here]

Annexes

I.   Review of basic sampling techniques

    1.1   Probability and non-probability sampling
    1.2   Simple random sampling
    1.3   Systematic sampling
    1.4   Stratified Random sampling
    1.5   Cluster sampling
    1.6   Estimation based on random samples
    1.7   Estimation of sampling errors
    1.8   Design effect

II.   Planning and execution of surveys

Issues related to survey planning and implementation are important but in many surveys they are not properly handled. While this subject will not be covered in any appreciable detail, salient features will be briefly summarized in the annex under the following topics.

    2.1   Problem statement and organizational requirement
    2.2   Survey budget
    2.3   Data collection methods
    2.4   Survey instruments
    2.5   Selection and training of field staff
    2.6   Conducting interviews