UNITED NATIONS SECRETARIAT **ESA/STAT/AC.93/4**
Statistics Division **03 November 2003**

English only

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

D R A F T

Documentation and evaluation of sample designs*

by

Anthony G. Turner**

* This document is being issued without formal editing.

** The views expressed in this paper are those of the author and do not imply the expression of any opinion on the part of the United Nations Secretariat.

Table of contents

Chapter Four: Documentation and evaluation of sample designs

1. In countries with little prior experience in conducting household surveys, metadata are often poorly documented in survey worksheets and reports. In this way errors creep into survey analysis. For example, probabilities of selection may not be fully known at the time of analysis without careful record-keeping. Consequently survey weights needed to inflate the data may be miscalculated. The handbook highlights the importance of keeping detailed records of metadata, so that analysis is conducted properly and the means for evaluating the sample design exist.

4.1. Need for documentation and evaluation

2. The sampling technician should take necessary steps to carefully document not only the sample plan for the particular survey undertaken but also its implementation. Sample designs often require adaptations at various stages of the field work because of unforeseen situations that arise in the conduct of the survey. It is important to record - step-by-step - all the procedures used in carrying out the sample plan to make sure the implementation is faithful to the design. When it is not it is even more important to document all the departures from the design, even minor ones. This information is necessary later at the analysis stage, in case any adjustments need to be made; but it is also indispensable for planning future surveys.

4.2. Labels for design variables

3. The identification of the units of selection at each stage must be clearly and uniquely labeled. In a multi-stage design this will mean establishing codes for the primary, secondary, tertiary and ultimate sampling units (depending upon how many stages are in the design). Normally a four-digit code will suffice for the first stage of selection and a three-digit code for the remaining stages. Geographic domains must also be properly labeled. In addition, the administrative codes identifying the geographic, administrative structure of the areas to which the sampling units belong should be part of the labeling process.

- ***Example***

Suppose a sample of 1200 *PSU*s, defined as census *EA*s, is selected for a two-stage design – 600 in each of two domains defined as urban and rural. A convenient way to code the *PSU*s is 0001 through 1200. Moreover, it is also useful to assign those codes in the same sequence that was used to select the *PSU*s for use in calculation of sampling variances. Thus if the rural *PSU*s were selected first they would be coded 0001 to 0600, while the urban ones would be coded 0601 to 1200. Such a coding scheme has two advantages. First, each *PSU* is uniquely numbered and identified and secondly, analysts can tell at a glance whether a *PSU* is urban or rural simply by its ID. In the second stage of the sample, each *PSU* is listed and 20 households are selected for interview. In this stage all listed households would be given a three-digit code (or four digits if some *EA*s were to contain more than 999 households), again in the sequence in which they are listed. The sample households would retain the code assigned in this manner, as opposed to assigning, say, codes 01-20 for the selected ones. Finally, administrative codes are assigned as necessary. Thus a sample household that might be coded as 09 003 008 0128 080 would identify it as the 80th household listed (and selected for interview) in *PSU* 0128, which belongs to civil

division 008 in district 003 of province 09. Moreover, the *PSU* number instantly identifies the household as belonging to the rural domain. If the survey obtains information about the members of the households, each one of them would also carry a unique code of two-digits, 01 to 99.

4. It is perhaps apparent why proper labeling is essential. One clear reason is for quality control. As assignments are made to interviewers and questionnaires are returned from the field, they can be checked off against a master list to make sure that all sample households are accounted for. Secondly, the unique numbering systems is invaluable to the data processing staff because it allows tabulations to be made by geographic location.

4.3. Selection probabilities

5. One item of information that is often overlooked in sample documentation is calculation of the probabilities of selection at the various stages. Where information does exist it is often confined to the overall sample weight (from which the overall probability can be readily calculated) for each sample case.

6. A particularly important detail for proper documentation occurs when sub-sampling is done in field. It may happen when a sample segment/cluster is too large. It also may occur when there is more than one household in a dwelling (when the dwelling is the listing unit). Careful recording of the sub-sample rate is essential so that the probability of selection for the affected segment or household can be accurately calculated by the sampling staff and the weight thereby properly adjusted.

7. It is also useful to record the probabilities of selection at each stage, however. For example, the probability of selecting each *PSU* is different whenever *pps* sampling is used. This is true even if the overall sample design is self-weighting. If the probabilities of selection of the *PSU*s are not recorded it is not possible to properly figure the weights if those *PSU*s should be sub-sampled for subsequent surveys.

4.4. Response rates and coverage rates at various stages of sample selection

8. As part of the evaluation process to examine the implementation of the sample survey, it is essential to provide information to users on response rates and coverage rates. It is useful to make as much detail as possible available. Thus it is important to provide not only the rate of response (or its complement, rate of non-response), but also a tabulation of the reasons for non-response. Categories of non-response would likely include the following:

- No one at home.
- Vacant dwelling unit.
- Demolished or uninhabitable dwelling unit.
- Refusal.
- Away temporarily (holiday, etc.).

9. Often, whole clusters are not interviewed for various reasons including issues of security such as civil strife or disorder and lack of accessibility due to the terrain or weather. Frequently when such problems occur substitute clusters are selected, a procedure that is seriously biased because the inhabitants of the substitute clusters are almost always likely to differ in very significant ways from those in the replaced clusters. Nevertheless, when such substitutions are made it is incumbent upon the survey team to record the number and location of such clusters. Moreover, it is also important to provide some information on under-coverage in such cases. This might be done by estimating, to the extent possible, the number of persons in the target population(s) thought to reside in the areas that the replaced clusters represent.

10. It is useful to note that problems of the type mentioned in the preceding paragraph can be reduced somewhat by identifying in advance of sample selection the areas of the country that are "out of scope" for survey interviewing due to security or accessibility concerns. Those identified should be excluded from the survey universe before sampling, and the survey reports should mention clearly that these areas are not "represented" by the sample.

4.5. Weighting: base weights, non-response and other adjustments

11. Calculation of survey weights is presented in the next chapter of the handbook. Here we emphasize the importance of documenting those calculations.

12. Weighting for household surveys generally involves up to three operations – calculation of the base, or *design*, weights, adjustments for non-response and adjustments for post-stratification. In many applications only the design weights are used, while in others the design weights may be adjusted by an additional factor to reflect non-response. In comparatively few applications the weighting may reflect another factor, either with or without non-response adjustments, intended to adjust the population distribution obtained from the sample to agree with the distribution from an independent source of data such as a recent census. The latter is often referred to as post-stratified weighting. In some applications no weighting is done at all; this would occur only when two conditions are met: the sample is completely self-weighting and the data generated are restricted to percentage distributions, proportions and ratios, as opposed to estimated totals or absolutes.

13. When weighting is used it is necessary of course to carefully record the calculations. As mentioned previously the weights (or probabilities) at each stage of selection should be calculated and recorded. Also, separate weights at each phase of data operations should be recorded, that is, (1) design weights, (2) design weights after multiplication by the non-response adjustment factor(s) and (3) the latter after adjustment factors for post-stratification have been applied. It is important to note that weights will naturally be different when the sample design includes domain estimation. Each domain will have its own distinct weight if the sample is self-weighting within domains or set of weights if it is not self-weighting within domains. In addition, it should be noted that non-response adjustments are often applied separately by important geographic sub-areas such as major regions, irrespective of whether domain estimation is present in the design.

4.6. Information on costs

14. While household surveys are usually budgeted very carefully it is equally important to keep records of actual expenditures of its various operations. Regarding the sampling operations this is especially useful for master sample designs as well as for planning the sampling aspects of future surveys.

15. Sampling operations that ought to be carefully monitored with respect to costs include the following:

a. Salaries for sample design including fees for any outside consultant.
b. Field costs for up-dating the sample frame including personnel and preparation of auxiliary materials such as maps.
c. Computer costs to prepare the sample frame for selecting the sample of *PSU*s.
d. Personnel costs to select the sample of *PSU*s (if not done by computer).
e. Field costs to conduct the listing operation in the penultimate-stage sampling units including personnel and the preparation of materials such as cluster folders.
f. Personnel costs to select the sample of households within the sample clusters.

4.7. Evaluation

16. Much of the documentation mentioned in the preceding subsections is useful for evaluating the sample design and survey implementation as well as for processing the survey results. Information on response rates are used to evaluate the survey results, while sampling costs may be used to evaluate the effectiveness of the sample design and its utility for future surveys.

17. An important component of sample evaluation is estimation of sampling errors for the key survey estimates. As mentioned previously, one of the distinguishing characteristics of a probability sample is that the sample itself can be used to estimate standard errors. Methods of variance estimation are discussed in detail in chapter 6. Generally, estimates of standard errors are prepared for the key characteristics of interest in the survey, since it is neither practical nor necessary to calculate them for all the items. The standard errors of course provide the means for users to evaluate the reliability of the survey estimates and to construct confidence intervals around the point estimates.

18. The standard errors may also be used to evaluate the sample design itself. A particularly useful statistic for doing this is the sample design effect, *deff*, or more precisely, *deft*, the square root of *deff*. It is fairly straightforward to calculate *deft* for every data item for which the standard error is estimated. It only entails dividing the estimated standard error, for a given item, by the standard error from a simple random sample of the same sample size, namely, pq/n, where p is the estimated proportion; q is $1-p$ and n is the sample size. The exercise serves to confirm or refute the design effects that were assumed when the sample was being designed, since the actual *deff*s or *deft*s cannot be known until after the survey has been conducted, the data processed and the standard errors estimated.

19. The sampling statistician can use the calculated design effects to evaluate whether the cluster sizes are of reasonable size for key data items and take corrective action if necessary. For example, if *deft* is much larger than anticipated for certain key items, the sample for a future survey may be designed to use smaller cluster sizes.

References and further reading

Bennett, S. (1993), "The EPI Cluster Sampling Method: A Critical Appraisal," Invited Paper, International Statistical Institute's 49th Session, Florence.

Cochran, W.G. (1977), *Sampling Techniques*, third edition, Wiley, New York.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953), *Sample Survey Methods and Theory*, Wiley, New York.

Hussmans, R., Mehran, F. and Verma. V. (1990), *Surveys of Economically Active Population, Employment, Unemployment and Underemployment, An ILO Manual on Concepts and Methods*, Chapter 11, "Sample Design," International Labour Office, Geneva.

International Statistical Institute (1975), *Manual on Sample Design*. World Fertility Survey Basic Documentation, Voorburg, Netherlands.

Kalton, G. (1983), *Introduction to Survey Sampling*, Sage, Beverly Hills.

_____ (1987), "An Assessment of the WHO Simplified Cluster Sampling Method for Estimating Immunization Coverage," report to UNICEF, New York.

_____ (1993), *Sampling Rare and Elusive Populations, National Household Survey Capability Programme*, United Nations Statistics Division, New York.

Kasprzyk, D. et al. editors (1989), *Panel Surveys*, Chapter 1, John Wiley & Sons, New York.

Kish, L. (1965), *Survey Sampling*, Wiley, New York.

Krewski, D., Platek, R., and Rao, J.N.K. editors (1981), *Current Topics in Survey Sampling*, Academic Press, New York.

League of Arab States (1990), *Sampling Manual, Arab Maternal and Child Health Survey*, Basic Documentation 5, Pan Arab Project for Child Development (PAPCHILD), Cairo.

Macro International Inc. (1996), *Sampling Manual*, DHS-III Basic Documentation No. 6. Calverton, Maryland.

Namboodiri, N.K. editor (1978), *Survey Sampling and Measurement*, Academic Press, New York.

Raj, D. (1972), *Design of Sample Surveys*, McGraw-Hill, New York.

Scott, C. (1993), Discussant comments for session on "Inexpensive Survey Methods for Developing Countries," Invited Paper Session, International Statistical Institute's 49th Session, Florence.

Som, R. (1966), *Practical Sampling Techniques*, second edition, Marcel Dekker, Inc., New York.

Turner, A., Magnani, R., and M. Shuaib (1996), "A Not Quite as Quick but Much Cleaner Alternative to the Expanded Programme on Immunization (EPI) Cluster Survey Design," *International Journal of Epidemiology*, Vol.25, No.1, Liverpool.

United Nations Children's Fund (2000), *End-Decade Multiple Indicator Survey Manual*, Chapter 4, "Designing and Selecting the Sample," UNICEF, New York.

United Nations Statistics Division (1984), *Handbook of Household Surveys*, revised edition, United Nations, New York.

_____ (1986), *Sampling Frames and Sample Designs for Integrated Household Survey Programmes, National Household Survey Capability Programme*, United Nations, New York.

United States Bureau of the Census (1978), *Current Population Survey Design and Methodology*, Technical Paper 40, Bureau of the Census, Washington.

Verma, V. (1991), *Sampling Methods*. Training Handbook, Statistical Institute for Asia and the Pacific, Tokyo.

Waksberg, J. (1978), "Sampling Methods for Random Digit Dialing," The Journal of the American Statistical Association, 73, 40-46.

World Bank (1999), *Core Welfare Indicators Questionnaire (CWIQ) Handbook*, Chapter 4 "Preparing the CWIQ Sample Design," World Bank, Washington.

World Health Organization (1991), *Expanded Programme on Immunization, Training for Mid-level Managers: Coverage Survey*, WHO/EPI/MLM91.10, Geneva.