UNITED NATIONS SECRETARIAT **ESA/STAT/AC.93/1**
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

Main sources of socio-demographic statistics [*]

by

Jeremiah P. Banda [**]

[*] 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

Introduction

1. Household surveys are among the major sources of social and demographic statistics in many countries. It is recognised that population and housing censuses which are another source of social statistics are conducted at long interval of about ten years while for most countries the administrative record systems for social statistics are not well developed and in many cases incomplete. Household surveys, therefore, provide a convenient avenue for collection of detailed and varied socio-demographic data pertaining to conditions under which people live, well-being, activities in which they engage, demographic characteristics and cultural factors which influence behaviour and social and economic change. This, however, does not preclude the complementary use of data generated through household surveys with data from other sources such as censuses and administrative records.

Objectives of the Handbook

2. While having a sampling background will be helpful in using the present handbook, other users with a general knowledge of statistical and mathematical concepts will also be able to use and apply the handbook with little or no assistance. This is because one of the key objectives of the handbook is to present material in a practical, hands-on format as opposed to stressing the theoretical aspects of sampling although the latter is not completely ignored.

3. The main purpose of the handbook is to include in one publication the main sample survey design issues that can conveniently be referred to by practicing national statisticians, researchers and analysts involved in sample survey work and activities in countries. Methodologically sound techniques that are grounded in statistical theory are used in the handbook, implying the use of probability sampling at each stage of the sample selection process. The contents of the handbook can also be used as a training material for introductory courses in sample survey design at various statistical training institutions, which offer courses in applied statistics.

4. In addition, the handbook has been prepared to complement other publications dealing with sample survey research issued by the United Nations, such as the series published under the National Household Survey Capability Programme (NHSCP).

More specifically, the objectives of the Handbook are to:

a. Provide, in one publication, basic concepts and procedures relevant to the principles of sample survey sampling and applied aspects of household sample design.
b. Serve as a practical guide for survey research professionals in designing and implementing efficient household samples surveys.
c. Illustrate the interrelationship of sample design, data collection, processing and analysis.
d. Highlight the importance of reducing nonsampling *errors* in household sample surveys.

Chapter One: Main sources of socio-demographic statistics

5. The main sources of social and demographic data are population and housing censuses, administrative records and household sample surveys. These three sources, if well planned and executed, can be complementary in an integrated programme of data collection and compilation. Social and demographic statistics are essential for planning and monitoring socio-economic development programmes. Population composition by age and sex including geographical distribution are among the most basic data necessary to describe a population or/and a sub-group of a population. These basic characteristics provide the context within which other important information on social phenomena, such as education, disability, labour, health, nutrition, crime, fertility, mortality and migration, can be studied.

6. We reiterate that a fully integrated system of social and demographic statistics depends on data from all the three sources and that none can be used to the exclusion of the other if the aim is to maximise the availability of social and demographic data in a country. For example, population censuses are the only feasible vehicle for obtaining small area statistics that are vital for local planning for social services such as the number and location of schools and hospitals. Sample surveys on the other hand are better suited to national and relatively large geographic domain level data on topics that need to be explored in depths such as the multi-dimensional aspects of disability, household expenditure, labour-force activity and criminal victimization. This is in contrast to censuses that collect relatively general information.

1.1. Population and housing censuses

7. A population census, hence forth referred to as census, is the total process of collecting, compiling, evaluating and disseminating demographic, social and other data at a specified time covering all persons in a country or in a well-delimited part(s) of a country. It is a major source of social statistics, with its obvious advantage of providing data for small geographic units, which are not prone to sampling errors. A census is an ideal method for providing information on size, composition and spatial distribution of the population in addition to socio-economic and demographic characteristics. In general the census collects information for each individual in households and each set of living quarters, usually for the whole country or well defined parts of the country.

1.1.1. Basic feature of a population and housing census

a. Individuals in the population and each set of living quarters are enumerated separately and the characteristics thereof are recorded separately.

b. Universality within a defined geographic area/territory. The population census potentially covers the whole population in a clearly defined territory. It should include every person present and/or usual residents depending on whether the type of population count is de facto or de jure. In the absence of comprehensive population or administrative registers, censuses are the only source that can provide small area statistics.

c. The enumeration has to be as simultaneous as possible. All persons and dwellings should be enumerated with respect to the same well-defined reference period.
d. Censuses are usually conducted at defined intervals. Most countries conduct censuses every 10 years while others every five years. This facilitates the availability of comparable information at fixed intervals.

1.1.2. Uses of census results

a. Censuses provide information on size, composition and spatial distribution of the population together with demographic and social characteristics.
b. They are also used as sources of sampling frames with some auxiliary information, which can be useful for sample selection and estimation in household survey.
c. Censuses are a source of small area statistics.

8. Population censuses have been carried out in many countries during the past decades. For example, about 184 countries and areas have conducted or are planning to conduct censuses during the 2000 round (1995-2004).

1.2. Administrative records

9. Many types of social statistics are compiled from various administrative records as by products of the administrative processes. Examples include health statistics compiled from hospital records, employment statistics from employment exchange services vital statistics compiled from the civil registration system and education statistics from enrolment reports of the ministry of education.

10. The reliability of statistics from administrative records depends on the completeness of the administrative records and the consistency of definitions and concepts. It is therefore necessary to continuously improve and update the systems of recording, compiling and analysing such data. Wherever possible, it is advisable to use the same definitions and concepts used for other data sources. This would facilitate the comparability of data.

11. While administrative records can be very cost-effective sources of data, such systems are not well developed in most developing countries. This implies that in a majority of cases such data are unreliable. Even if the administrative recording processes are continuous for purposes of administration, the compilation of statistics is, in most cases, secondary. Statistical requirements that need to be maintained such as standardisation of concepts and definitions, adhering to timeliness and complete coverage are not usually considered or adhered to.

12. For most countries, information from administrative records is often limited in content as their uses are more for legal or administrative purposes. Civil registration systems are examples of administrative systems that many countries have developed. However, not all countries have been successful in this effort. Countries with complete vital registration systems are able to produce periodic reports on vital events, such as number of live births by sex; date and place of births; number of deaths by age; sex; place of deaths and cause of death; marriages and divorces; etc.

13. A population register maintains life databases for every person and household in a country. The register is updated on a continuous basis when there are changes in the characteristics of an individuals and a household. If such registers are combined with other social registers they can be a source of rich information. Countries, which have developed such systems include Denmark, Norway, the Netherlands, Germany and Sweden. For most of these countries censuses are based on the registration system.

14. In many developing countries, while administrative records for various social programmes can be cost-effective data source and an attractive proposition, they are not well developed. In this case their complementary use with other sources is a big challenge because of lack of standardized concepts, classification systems coupled with selective coverage and under coverage.

1.3. Household surveys

15. Household sample surveys are key source of data on social phenomena. They are among the most flexible methods of data collection. In theory almost any subject can be investigated through household surveys. It is common for households to be used as second stage sampling units (SSUs) in most area based sampling strategies (see chapter 3 of this handbook). In sample surveys, part of the population is selected from which observations are made or data are collected and then inferences are made to the whole population. Because in sample surveys there are smaller workloads for interviewers and a longer time period assigned to data collection, there are possibilities of covering most subject matters in greater detail than in censuses. In addition, there is scope for training field staff more intensively. The reality is that not all the data needs of a country can be met through census taking, therefore, household surveys provide a mechanism for meeting the additional and emerging needs on a continuous basis. The flexibility of household surveys, therefore, makes them excellent choices for meeting data users' needs for statistical information which otherwise would not be available and insufficient.

16. Many countries have in place household survey programmes, which include periodic and ad-hoc surveys. It is advisable that the household survey programme should be part of an integrated statistical data collection system of a country. In the area of social and demographic statistics intercensal household surveys can constitute part of this system.

17. The National Household Survey Capability Programme (NHSCP) was a major effort to help developing countries to obtain requisite information on the household sector. The NHSCP was implemented for nearly 14 years from 1979 to 1992. By the time of its conclusion, 50 countries had participated in the programme. Its major achievement was the promotion and adoption by countries of continuous multi-subject integrated household surveys. In addition, the programme fostered sample survey capacity building, especially in Africa.

18. There are different types of household surveys that can be conducted to collect data on social and demographic statistics such as specialised surveys, multi-phase surveys, multi-subject surveys, longitudinal surveys. The selection of a specific type of survey will depend on a number of factors including, subject matter requirements, resources and logistical considerations.

19. Specialised surveys cover single subjects or issues such as time-use or nutrition. The surveys may be periodic or ad-hoc.

20. Multi-phase surveys entail collecting statistical information in succeeding phases with one phase serving as a precursor to the next. The initial phase usually constitutes a larger sample than subsequent phases. It is used to screen sample units based on certain characteristics to ascertain the eligibility of such units to be used in the subsequent phases. These surveys are a cost-effective way of reaching the target population in the latter phases to collect detailed information on a subject of interest. Demographic and Health surveys are among those suited to this approach.

21. In multi-subject survey different subjects are covered in a single survey. This approach is generally more cost-effective than conducting a series of single subject surveys.

22. In longitudinal surveys, data is collected from the same sample units over a period of time. The interval can be monthly, quarterly or annually. The purpose for conducting such surveys is to measure changes in some characteristics over a period of time. The major problem with this type of surveys is the high attrition rate of respondents. There is also the problem of conditioning effect.

23. While household surveys are not as expensive as censuses they tend to be costly especially if results have to be produced for relatively lower administrative domains. Because of the cost constraint of large household sample surveys, in practice most household surveys are limited in sample size. It would therefore be inappropriate to make inferences, based on such samples, for relatively small domains because standard errors would be very large rendering the results less precise.

Some advantages of household surveys compared to censuses:

a. The overall cost of a survey is generally lower compared to a census as the latter requires large amounts of manpower, financial, logistical and material resources. From a properly selected representative and implemented sample, accurate and reliable results can be a basis for making inferences on the total population. Consequently for some estimates such as total fertility rate, there is no compelling need for a census.
b. In general sample surveys produce statistical information of better quality because, as stated earlier, it is more feasible to engage better and well-trained enumerators. It is also easier to provide better supervision because supervisors are usually well trained and the supervisor/ enumerator ratio can be as high as 1 to 4. In addition it is possible to use better measuring equipment in surveys if need arises. In a census data quality is, in some cases, compromised because of the massive nature of the exercise, which is prone to lapses in, and neglect of, quality assurance at various stages, resulting in high nonsampling errors.
c. There is a greater scope and flexibility in a sample survey than in a census with respect to the depth of investigation and number of items in the questionnaire.

Information of a more specialised type may not be collected in a census because of the prohibitive number of specialists or equipment necessary to carry out the study. An example is the weighing of food and other measurements in a nutrition study. It is likewise, not feasible to subject every person in the population to a medical examination to collect data necessary for estimating HIV/AIDS infection. On the other hand, it is possible to add items in a household sample survey that would be relatively complex for the census.

1.4. Complementarities of the three data sources

24. The subject of combining information from different data sources has gained considerable attention in recent years. The interest in this area is driven by the necessity to limit census and survey costs and to lower response burden, to provide data at lower domains, which may not be covered by survey data for instance, and to maximise the use of available data in the country.

25. Because censuses cannot be repeated frequently, household surveys provide a basis for updating some census information especially at national and other large domain levels. In most cases only relatively simple topics are investigated in a census and the number of questions is usually limited. Census information can therefore be complemented by detailed information on complex topics from the household surveys, taking advantage of their small size and potential flexibility.

26. Censuses and household surveys have, in many instances, been complementary. For examples, sampling has been applied at different stages of census activities such as pre-testing, enumeration and tabulation stages.

27. Collecting information on additional topics from a sample of the population, or households during the census is a cost-effective way to broaden the scope of the census to meet the expanding demands of social statistics. The use of sampling methods and techniques makes it feasible to produce urgently needed data with acceptable precision when time and cost constraints would make it impractical to obtain such data through complete enumeration.

28. Sample enumeration carried out within the census operation is able to collect data on additional topics and ask more detailed information.

29. On the other hand the census, among other things, provides a sampling frame, statistical infrastructure, statistical capacity (it is part of a fundamental statistical base of a country) and benchmark statistics that are needed in conducting household surveys. It is common to draw a sample of households within a census context, to collect information on more complex topics such as, disability, maternal mortality, economic activity and fertility.

30. Censuses support household surveys by providing frames and some auxiliary information for use in efficient design of surveys. Apart from provision of sampling frames, census information can be used to construct master samples. Further more auxiliary information from

censuses can be used improve sample estimates through regression and ratio estimates, thereby improving the precision of survey estimates.

31. In many countries, sampling frames for household surveys are based on population censuses. In general, the census provides an explicit list of all area units, such as enumeration areas, commonly used as first stage units in household sample surveys selection process.

32. In order to achieve integration of data sources there is need to clearly identify units of enumeration and adopt consistent geographic units in collecting and reporting statistics through the various sources. In addition, it is essential to adopt common definitions, concepts and classifications across different sources of data.

33. Data from household surveys can also be used to check, for instance census coverage and content. The aim is to determine the size and direction of such errors. Post enumeration surveys were used for this purpose during the 2000 round of censuses in Zambia and to evaluate coverage errors, and in Cambodia, just to mention a few. Likewise census data can be used to evaluate some survey results.

34. Small area estimation, which has received a lot of attention due to growing demand for reliable small area estimators, is an area where data from surveys and administrative records are concurrently to produce estimates. Traditional area-specific direct estimators do not provide an adequate precision because sample sizes in small areas are seldom large enough. Small area estimation is based on a range of statistical techniques used to produce estimates for areas when traditional survey estimates for such areas are unreliable or cannot be calculated. The techniques involve models that borrow strength over space and time or from auxiliary information from administrative records or censuses. The basic idea of small area procedures is, therefore, to borrow and combine the relative strength of different sources of data in an effort to produce more accurate and reliable estimates.

35. In countries with well-developed civil registration systems, census and survey data can be successfully used together with data from administrative records. For example in the 1990 population census in Singapore, enumerators had pre-filled basic information, from administrative records, for every member of the household. This approach reduced interviewing time and enumeration costs. Since the register-based census provides only the total count of the population and basic characteristics of the population, detailed socio-economic characteristics are collected on a sample basis.

36. Data from administrative records can be used to check and evaluate results from surveys and censuses. For instance in countries with complete vital registration systems, data on fertility and mortality from censuses can be crosschecked with that from the registration system.

1.5. Conclusion

37. In conclusion, censuses, household surveys and administrative sources should be viewed as complementary. This implies that, whenever possible, in planning for censuses and surveys

common concepts and definitions should be used. Administrative procedures should also be checked periodically to make sure that common concepts and definitions are being used.

38. The household survey programme should be part of an integrated statistical data collection system within a country, including censuses and administrative records so that the overall needs for socio-demographic statistics can be adequately met.

References and further reading

Ambler, R. et al. (2001) Combining unemployment benefits data and LFS to estimate ILO unemployment for small areas: An application of modified Fay-Herriot method, Invited paper, International Statistical Institute, 34th Session, Seoul, August 22-29, pp.128-131.

Biemer, P.P. and Lyberg, L. E. (2003). Introduction *to Survey Quality*: Wiley Series in Survey Methodology, Wiley, Hoboken.

Biemer, P. P. et al. (editors) (1991). *Measurement Errors in Surveys:* Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Bee-Geok, L. and Eng-Chuan (2001). Combining Survey and Administrative Data for the Singapore's Census of Population 2000.Invited paper, International Statistics Institute, 3—the session, Seoul, August 22-29, 2001, pp. 139-155.

Groves, R.M. and Couper, M. P. (1995), Theoretical motivation for Post-Survey Nonresponse Adjustment in Household Surveys: Journal of Official Statistics. Vol.11, N0.1, pp. 93-106.

Hansen M. H., Hurwitz, W. N. and Bershad, M. A. (2003). Measurement Errors in Censuses and Surveys: Landmark Papers in Survey Statistics IASS Jubiliee Commmorative Volume.

Kalton, G. and Heeringa, S. (2003), *Leslie Kish: Selected papers*, Wiley Series in Survey Methodology, Hoboken.

Kiregyera, B. (1999). Sample Surveys: with Special reference to Africa, PHIDAM Enterprises, and Kampala.

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

Murthy, M.N. (1967). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.

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

Raj, D and Chandhok, P (1998) *Sample Survey Theory*, Narosa Publishing House, London.

Rao, J. N. K. (1999) *Some recent advances in model-based small area estimation*, in *Survey Methodology*, Vol. 25, No. 2, pp. 175-186, Statistics Canada

Som, R. K. (1996), *Practical Sampling Techniques*, Marcel Dekker Inc., New York.Cochran, W.G. (1963), *Sampling Techniques*, Wiley, New York.

Suharto, S. (2001), Complementary Sources of Demographic and Social Statistics, United Nations Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid –Decade Assessment and Future Prospects, New York, 7-10 August.

Sukhatme, P.V. et al (1984), Sampling Theory of Surveys with Applications, Iowa State University Press and Indian Society of Agricultural Statistics, Ames and New Delhi.

United Nations (1981), The National Household Survey Capability Programme: Prospectus, New York.

_____ (1982), National Household Survey Capability Programme: Non-sampling Errors in Household Surveys: Sources Assessment and Control, New York.

_____ (1982), Nonnsampling errors in Household Surveys: Sources, Assessment and Control: National Household Survey Capability Programme, New York.

_____ (1984). Handbook of Household Surveys (Revised edition), Sales No. E.83.XVII.13, New York.

_____ (2002), Technical Report on Collection of Economic Characteristics in Population Censuses, New York.

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

Zanutto, E. and Zaslavsky, A. (2002), Using Administrative Records to improve Small Area Estimation: An Example from the U.S. Decennial Census, Journal of Official Statistics, Statistics Sweden, Vol. 18, No.4, pp.559-576.