



Distr.: GENERAL

ECA/DISD/STAT/RPHC.WS/ 2/99/Doc 3.5
2 November 1999

UNITED NATIONS

ECONOMIC AND SOCIAL COUNCIL

Original: ENGLISH

ECONOMIC AND SOCIAL COUNCIL

Training workshop for national census personnel within
the framework of the 2000 round of population and
housing censuses

15-19 November 1999
Addis Ababa, Ethiopia

C. CENSUS POST ENUMERATION ACTIVITIES

- e. Data analysis and evaluation**
- f. Dissemination and utilization of census results**

**UNITED NATIONS POPULATION FUND
COUNTRY SUPPORT TEAM
ADDIS ABABA**

**SELECTED CENSUS TOPICS ON PILOT CENSUS, FINALIZATION
OF DOCUMENTS; AND DATA EVALUATION, ANALYSIS,
DISSEMINATION AND UTILIZATION**

By
Israel Sembajwe
Regional Adviser
UNFPA/CST Addis Ababa

Prepared for the Training Workshop for National Census Personnel within
the framework of the 2000 Round of Population and Housing Censuses,
from 15 to 19 November 1999

INTRODUCTION

The Population and Housing Census is the most extensive data collection exercise in any country. It is the major source of demographic and socio-economic data in developing countries since it provides the basic information for all sectors of planning.

By nature, this type of exercise puts extreme demands on Government in terms of resource provision (including financial and human resources). Because of this extreme demand, there is always a need for a comprehensive planning process that takes care of every stage of the census from the preparatory stage right to the stages of data dissemination and the promotion of utilization of that data and the results of the analysis. The comprehensive census plan would by nature take care of the resources required and incorporate these in a comprehensive budget which UNFPA (normally as the lead United Nations agency in the population field) would use to assist Government to mobilize resources both from its own treasury and from cooperating organizations including UNFPA, the British Department for International Development (DFID), and the United States Agency for International Development (USAID).

On the premise that the planning process is comprehensive and a budget is available to guide the commitment of resources right from the beginning of the process, this paper sets out to discuss only a few selected elements of this process. It deals with selected aspects of the pilot census, finalization of documents, data evaluation, analysis, dissemination and utilization of census results. Detailed discussion of these topics can be found in (Sembajwe, 1994; UNECA, 1996; and United Nations, 1998).

The paper is divided in three parts. The first part will cover the pilot census, finalization of questionnaires and other documents, and printing of census documents; the second part will cover data evaluation; and the third part will cover analysis, preparation of census reports, dissemination, and utilization of census results.

THE PILOT CENSUS, FINALIZATION OF QUESTIONNAIRES, MANUALS, OTHER DOCUMENTS; AND PRINTING OF CENSUS DOCUMENTS

1.1. Introduction

During the preparatory stages for the main census enumeration, there is a need to test the various aspects of the census plan. The testing is normally conducted during the pilot census. To be as comprehensive as possible, this stage is aimed at testing

- Field organization;
- Field instruments and related documents;
- Training programme;
- Publicity and advocacy;
- Enumeration;
- Data processing; and
- Other related aspects of the entire process

This section, therefore, deals with the different aspects of the pilot census.

1.2. The pilot census

The pilot census covers a limited number of selected enumeration areas with the aim of testing the entire census plan and organization. It is undertaken in conditions similar to those which will prevail during the main enumeration. Information gathered during the pilot survey is used for a number of purposes including those presented in the following sub-sections.

Questionnaire

The information collected brings out weaknesses in the questionnaire such as

- Suitability of the questions (in terms of their formulation, instructions about them and their sequence in the questionnaire);

- Time estimated for the enumeration;
- Staff requirements;
- Probable cost for the enumeration;
- Suitability of selected technology for data processing;
- Suitability of questions to the tabulation and analysis plans; and
- Appropriateness of the publicity/advocacy strategy adopted.

Field organization

With the assessment of the logistical flow of materials and activities, it becomes easier for the census planners to make adjustments where these are found necessary. Matters of transportation and quality control become paramount at this stage.

Field instruments and related documents

Apart from the questionnaire, which is the most important field instrument, other instruments and documents that require fine tuning are:

- Enumerators' Instruction Manual;
- Supervisors' Instruction Manual;
- Publicity/Advocacy Document (s);
- Fliers, Stickers and Posters;
- Summary forms for sub-national levels;
- Control forms (to control the movement of questionnaires); and
- Call back cards.

Training programme

The way the questionnaires are completed in the field will give an indication about the quality of enumerators and supervisors and the deficiencies that may have taken place during the training of these two groups of staff. The positive aspects will guide more emphasis in the training for the main enumeration, and the negative aspects will indicate where corrective measures will be necessary during the same training.

Publicity and advocacy

Problems related to cooperation by national leaders and different community leaders as well as the civil society will provide indicators on where publicity/advocacy failed so that more effort may be put on those aspects during the publicity/advocacy campaign for the main enumeration. On the other hand, more positive aspects of the campaign may be highlighted so that more emphasis is put on those aspects during the next stage of the publicity/advocacy program.

Enumeration

The results of the pilot census will give a clear indication of the number of personnel required to complete the main enumeration in a given time; the amount of materials required; the logistical requirements implied; transportation needs; and so forth.

Data processing

The type of data processing approach selected will be tested when handling the pilot census results. For example, if the manual data entry approach utilizing microcomputers is selected, the length of time available for data processing, the number of data entry clerks required to finish the task in a given period, the cost implications of the approach, and hardware and software requirements will be adequately assessed. The issue of data entry errors manually generated will be critically assessed. Alternatively, if the use of more recent technology in data capture and processing is selected including the use of Optical Mark Readers (OMR) or Optical Character Readers (OCR) or Intelligent Character Recognition (ICR) scanners, all the above considerations, except manual data entry errors, will be examined including the financial implications of using scanners and related software. In using scanners, the quality of paper, design and printing of the questionnaire as well as the rejection rate of questionnaires during scanning will be given center stage.

The data processing tests will also lead to the adjustment of the tabulation plan, fine tuning of the data editing specifications, and the strengthening of training and adjustment of instructions for data entry clerks and their supervisors.

Other related aspects of the entire process

The tabulations and related analysis of the pilot census results provide invaluable insight on the suitability of the tabulation and analysis plans.

1.3. Finalization of Questionnaire, Manuals and other Documents

After fine tuning the questionnaires and related documents using the pilot census results, they are finalized for printing.

1.4. Printing of Census Documents

The printing of the census documents requires special planning. For example, the printing of the final questionnaire(s) requires putting into consideration the following aspects:

- Method of enumeration (for example will the approach be single individual, single household or single set of living quarters, etc.);
- Data capture and processing technology (for example, is data capture going to be done manually or by using optical mark readers (OMRs)/ optical character readers (OCRs)/ intelligent character recognition (ICR) scanners);
- Format;
- Exact wording;
- Arrangement of questions; and
- Quality of paper and type of print to be used.

The printing should be done in good time to facilitate timely binding and distribution (in adequate quantities) to the lowest levels of census enumeration administration.

DATA EVALUATION

2.1. Introduction

The exercise to evaluate demographic data is carried out mainly for three reasons:

- To identify the types of errors in order to know which sections of the data contain these errors and what procedures of data collection and processing led to these errors.
- To determine the accuracy of the data collected. Different indicators enable us to do this and to locate the interval of the provided accuracy range in which our observed data fall.
- To make appropriate adjustments to the data before we can use them confidently to estimate different demographic and related parameters.

In this section we shall outline ways of dealing with the first two reasons.

2.2. Types of errors

- *Coverage Errors*

These are errors which affect the completeness and quantitative accuracy of the census. For example, they arise from whether all enumeration areas were covered twice or not covered, and whether certain sections of the population were over-enumerated or under-enumerated.

- *Content Errors*

These are errors which affect the quantitative characteristics of the census such as quality of age and sex data. For example, if female children under the age of five are under-reported this will affect the quantity of the child population as well as the quality of the sex and age distribution data.

2.3. Methods of Data Evaluation

Direct Methods

Direct methods depend on at least two independent sources of information (such as the census and survey or the census and vital registration system) which have covered identical areas either at the same time or within a short period of time of each other to assess the coverage and content errors. The most common source of the second source of information is called a post enumeration survey (PES). The survey is carried out immediately after the census and its success depends on the selection of a representative sample and getting an independent repetition of the first enumeration exercise under similar conditions before comparing the results from the two sources. A limited number of variables would be selected for the re-interview process to facilitate the matching of events.

Let us suppose we have two independent sources of information and we can evaluate their coverage by matching the events. We can utilize the Chandrasekaran Method as follows:-

1. Assume that the probability of an event being recorded by one system is the same regardless of whether the event is recorded by the other system. The events from both systems are matched to establish C, the number of events recorded by both Systems; N_1 , the number recorded only by System 1 (e.g. Registration System); and N_2 , the number recorded only by System 2 (e.g. Survey). The Chandrasekaran-Deming Formula then estimates total events N^* as:

$$N^* = C + N_1 + N_2 + \frac{N_1 N_2}{C} \quad \text{from the following information.}$$

	System 1	Not in System 1	Total
System 2	C	N ₂	S
Not in System 2	N ₁	X*	
Total	R		N*

* Denotes Estimate.

2. Estimate the total number of events (N*) as follows:-

$$N^* = C + N_1 + N_2 + X^*$$

If X* is estimated as $N_1 N_2 / C$ then

$$N^* = C + N_1 + N_2 + N_1 N_2 / C$$

Which is the same as

$$N^* = \frac{(C + N_1)(C + N_2)}{C} = \frac{RS}{C}$$

Where R denotes the number of events recorded by System 1;

S denotes the number of events recorded by System 2.

3. Estimate the completeness of registration as the match rate of the survey as follows:-

$$\frac{R}{N^*} \approx \frac{C}{S}$$

4. Estimate the completeness of the survey as the match rate of registration as follows:-

$$\frac{S}{N^*} = \frac{C}{R}$$

The following conditions are necessary for the procedure to work:

1. That all and only events which took place are identified and matched;
2. That all identified events took place in the population and reference period.
3. That the two Systems are independent.

Example (Based on Live Births)

Suppose

R	=	240 are System 1 records
S	=	260 are System 2 records
C	=	180 are records in both System 1 and System 2
N ₁	=	60 are records only in System 1
N ₂	=	80 are records only in System 2

Then

$$X^* = \frac{60 \times 80}{180} = 27$$

Hence $N^* = 180 + 60 + 80 + 27 = 347$

This leads to the following table:-

	System 1	Not in System 1	Total
System 2	180	80	260
Not in System 2	60	27*	87
Total	240	107*	347

* Denotes Estimate.

Then the completeness of registration of births is:

$$\frac{240}{347} \approx \frac{180}{260} = 69.2\%$$

and the completeness of recording of births in the survey is:

$$\frac{260}{347} \approx \frac{180}{260} = 75.0\%$$

Indirect Methods

We have three groups of methods under this category and these are:

- Those based on external consistency checks;
- Those based on internal consistency checks; and
- Other methods

These are discussed subsequently.

External Consistency Checks

Under this type of checks, we compare the census results with evidence from other sources such as the vital registration systems; international migration data recorded at border posts, airports, ports etc; and administrative records such as those on school enrolment, registration of voters, and health statistics records. One of the methods used, given just as an example, is the **balancing equation**.

This equation is of the form $P_1 = P_0 + B - D + I - E$

Where P_1 represents population at the second census;

- P_0 represents population at the first census;
- B represents births recorded within the intercensal period;
- D represents deaths recorded within the intercensal period;
- I represents immigrants recorded within the intercensal period; and
- E represents emigrants recorded within the intercensal period.

When detailed data are not available as is the case in the African countries, this equation can be reduced to:

$$P_1 = P_0 + NC + NM$$

OR

$$P_1 = P_0 + NC$$

Where P_1 and P_0 are as defined earlier;

NC represents natural change; and

NM represents net migration.

The latter reduced equation applies where migration is negligible.

Example (Based on Lesotho)

- a) Carry out an evaluation exercise using the balancing equation with the following information: Population in 1986 = 1595096; Population in 1976 = 1216815; 553800 births registered for 1976-1986; 184600 deaths registered for the same period; zero migration.
 - b) Re-evaluate the information given that the registered births for the period were 526140 and deaths were 85200 with zero migration.
 - c) In both situations above incorporate the following additional information: 21500 people left the country and 19200 came in.
- i) Relating this problem to $P_1 = P_0 + B - D + I - E$

We have

- a) Population in 1986 = 1595096 as compared to $P_0 + B - D = 1216815 + 553800 = 1586015$.

The two figures are close and reflect somewhat acceptable results both from the census and registration system although there may be an under-registration of 9081 births/deaths or an over-enumeration of the same magnitude in the 1986 census/under-enumeration of the same magnitude in the 1976 census. Further checks are required.

- b) Population in 1986 = 1595096 as compared to $1216815 + 526140 - 85200 = 1657755$.

There is a high difference in the two results of 62,659. This suggests serious problems in the registration system especially with regard to registration of deaths (the registered deaths reflect an average CDR of about 6 per 1000 over the 10-year period which is too low) and other problems as suggested above.

- c) Population in 1986 = 1595096 as compared to $1216815 + 553800 - 184600 + 21500 - 19200 = 1588315$ in (a); and $1216815 + 526140 - 85200 + 21500 + 19200 = 1660055$ in (b)

- ii) This information provides the possibility of adjusting the results for migration on the assumption that migration records are accurate. Otherwise the problems suggested above require more checks to resolve.

Other methods used in this category include those on

- Intercensal growth rates; and
- Comparison of specific groups with external sources.

Internal Consistency Checks and Other Methods

In this group we include such methods as

- Those based on single year age data;
- Those based on grouped age data;
- Those based on simple demographic techniques;
- Those based on stable population models;
- Those based on survival ratio models;

- Those based on reverse and forward projection techniques; and
- Those devoted to specifically evaluating fertility, mortality and migration data.

The discussion of these methods is presented elsewhere.

PART III

ANALYSIS, PREPARATION OF CENSUS REPORTS, AND DISSEMINATION AND PROMOTION OF UTILIZATION OF RESULTS

3.1. Introduction

The analysis of the census data facilitates the dissemination and optimal utilization of the census results. But for a comprehensive analysis process to be initiated, there is a need for the following steps to be taken:

- The preparation of a comprehensive analysis plan (during the preparatory stages of the census);
- The setting up of a team of analysts (composed of members from the Census Office, other Government departments, training and research institutions, as well as the non-Governmental Organizations and the private sector); and
- A plan for disseminating the results and for promoting the utilization of the census results.

The next sub-sections will examine each of these steps in more detail.

3.2. The preparation of a comprehensive analysis plan

The analysis plan should include the following:

- A list of topics for analysis;
- A list of analysts responsible for each topic;
- The type of reports to be produced (i.e. whether preliminary report or main reports);
and
- The content of each report as well as the time frame for producing the report.

It is expected that the analysis topics would cover such areas as age, sex and socio-economic characteristics of the population; fertility; mortality; migration; housing and housing amenities; the youth; gender issues; and other priority issues specific to each national situation. For example, the HIV/AIDS pandemic is an emerging problem with yet unexplored impacts on the socio-economic, health and demographic conditions of the region. An analysis topic or even a separate detailed report may be devoted to this issue.

It has already been suggested that the selected topics would be handled by consultants from the training and research institutions, public and private sectors (specialized in their areas of competence) who would work in collaboration with members of the Census Offices/Statistics Office so that capacity building in data analysis for the Statistics Office is generated during this exercise.

After producing the analysis drafts, the Census Office would normally put all the drafts together and synthesize them before hiring a professional editor to refine them for publication.

The Preliminary Report

The Census Office would plan to publish the preliminary results within the shortest time possible, preferably within six months after the main enumeration. These results may include population distribution by sex and geographical areas.

Main Reports

The main reports may consist of two or more volumes. In the case of two volumes, for example, you may have the following:

- Volume I. This volume may deal with the general issues of population distribution by age, sex and geographical areas; literacy, school attendance and educational status; labour force participation; and other issues of general interest.
- Volume II. This volume may be divided into several volumes. It should be devoted to the publication of the detailed results of the census. As stated before, the analysis topics will cover such areas as the socio-economic characteristics of the population; fertility; mortality; migration; housing and housing amenities; the youth; gender issues; and HIV/AIDS.

3.3. The setting up of a team of analysts

It has already been suggested that a team of analysts should be set up composed of experts from training and research institutions, public and private sectors to work in collaboration with members of Census Offices/Statistics Office so that capacity building in data analysis for the Statistics Office is generated during this exercise. The consultants to be used in the analysis should be identified during the preparatory stages of the census and involved in earlier activities of the census such as the training of trainers, senior supervisors, supervisors and enumerators so that they are not new to the census activities and will have already developed a high degree of ownership for the census process before the analysis. In this regard, they would appreciate better the value of the data collected and the likely inadequacies that may exist in them.

3.4. A plan for disseminating the results and for promoting the utilization of the census results

The publication of the first volume should be followed by the conducting of publicity/advocacy seminars at all levels to disseminate the results. Subsequently, when more detailed results are produced, they should also be followed by more intensified publicity/advocacy activities aimed at disseminating the results to policy makers, planners, and experts in non-governmental organizations (NGOs) and the private sector at national and sub-national levels, and at sensitizing the same group of people in the utilization of data for planning in different sectors such as education, agriculture, health,

the provision of specialized services, and business and marketing. In addition, dissemination should be extended to the civil society using appropriate media such as radio, TV, newspapers, booklets, fliers and posters.

SUMMARY AND CONCLUSION

Summary

This paper has touched on a number of elements that are vital to the census process. These have included selected aspects of the pilot census, finalization of documents, data evaluation, analysis, dissemination and utilization of census results. These components are part and parcel of a larger comprehensive process that is vital to national planning and development.

Conclusion

The discussion provided in this paper has placed selected aspects of the pilot census, finalization of documents, data evaluation, analysis, dissemination and utilization of census results in the right context of the entire census process. It should be emphasized that none of these components is enough in itself to lead to a successful census process. All stages of the census process (from preparatory activities to dissemination and the promotion of utilization of the census results) taken together and handled efficiently lead to the achievement of a successful census.

REFERENCES

- Sembajwe, I. 1994. Evaluation of Demographic Data. United Nations Regional Institute for Population Studies (RIPS). Monograph Series No. 8. Accra, Ghana.
- UNECA. 1996. Making Population Censuses Cost-Effective And Pertinent for African Development. RAF/96/P02-UNFPA/ECA-TSS. Addis Ababa, Ethiopia.
- United Nations. 1998. Principles and Recommendations for Population and Housing Censuses. Revision 1. New York, USA.