2010 World Population and Housing Census Programme

UNITED NATIONS SECRETARIAT

DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS

STATISTICS DIVISION

Post Enumeration Surveys

Operational guidelines

Technical Report

New York, April 2010

NOTE

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The term “country” as used in this publication also refers, as appropriate, to territories or areas.

The designations “developed regions” and “developing regions” are intended for statistical convenience and do not necessarily express a judgment about the stage reached by a particular country or area in the development process.

Symbols of United Nations documents are composed of capital letters combined with figures.

PREFACE

The quality of population and housing census data is very important for many reasons, building public trust and understanding in the national statistical system. The purpose of census evaluation is to provide users with a level of confidence when utilizing the data, and to explain errors in the census result. It is therefore important to choose an appropriate way of sending out these messages to the right group of people.

It is universally accepted that a population census is not perfect and that errors can and do occur at all stages of the census operation. Errors in the census results are classified into two general categories - coverage errors and content errors. Coverage errors are the errors that arise due to omissions or duplications of persons or housing units in the census enumeration. Content errors are errors that arise in the incorrect reporting or recording of the characteristics of persons, households and housing units enumerated in the census. Many countries have recognized the need to evaluate the overall quality of their census results and have employed various methods for evaluating census coverage as well as certain types of content error.

Numerous methods are available to estimate the coverage and content error of censuses. These include simple techniques of quality assurance such as internal consistency checks. Comparisons of results with other data sources including previous censuses, current household surveys and/or administrative records are also useful techniques. Demographic analysis and post enumeration surveys are two very important methods for evaluating census data.

Nevertheless, countries have frequently encountered problems in implementing evaluation methods, and especially in designing and implementation post enumeration surveys. In that regard, they have urged the United Nations Statistics Division to prepare guidelines on the design and implementation of Post Enumeration Surveys. In that context, the United Nations Statistics Division organized in 2009 and 2010 a series of workshops in different regions of the world. The objectives of these workshops were to present an overview of the various methods of evaluating censuses with a focus on the post enumeration survey (PES) methodology.

The present manual was developed to serve as a reference document for these workshops, as well as to provide operational guidelines for the preparation and implementation of a post enumeration survey. It defines the PES and enumerates its objectives. In addition, the manual covers elements of: sample design; questionnaire design; planning and implementation of a PES; matching; field reconciliation; the Dual System of Estimation (DSE); tabulations; the evaluation of content error. The conclusion highlights the usefulness of a PES and care that must be taken in its implementation. It additionally, presents challenges associated to a successful implementation of the PES and how they can be mitigated. Finally, the Annex presents selected examples from countries that conducted PES during the 2000 and 2010 round of censuses. This manual has been prepared by Dr. Jeremiah Banda, International Consultant and benefited from comments from Dr. Rajendrah Singh, former Chief of the Decennial Statistical Studies Division, U.S. Bureau of the Census.

Table of contents

Chapter 1. Overview of Census Evaluation and selected Methods

1. In this chapter, we briefly discuss census errors; census evaluation methods including the strengths and weaknesses of the methods.

1.1. Census errors

2. Errors in censuses can arise from many sources such as flawed data a collection and processing procedures. Census designs can also be a source of error by introducing measurement error through faulty questionnaires, instructions, training materials, and procedures. At the outset it must be pointed out that errors are inevitable in a large data collection exercise such as a census.

3. ***Coverage error*** refers to either an under-count or over-count of units owing to omissions of persons/ housing units or duplication/erroneous inclusion, respectively. ***Content error*** pertains to the error in the characteristics that are reported for the persons or housing units that are enumerated. Both types of error can affect the distribution of the population with respect to their characteristics.

4. There are three types of coverage error:

(a) Omissions
(b) Duplications and
(c) Erroneous inclusions

These are errors can also occur in the recorded characteristics with respect to enumerated persons such as age, family relations, marital status, etc.

1.1.1. Omissions

5. This results from missing housing units, households, or persons during census enumeration. In the case of missing the whole housing unit, it implies that all households and persons living in that housing unit will also be missed during the census enumeration.

6. The major causes of omission of housing units include; failure to cover the whole land area of the country in creating enumeration areas and mistakes made by enumerators in canvassing in assigned areas. The former problem can be caused by unclear boundaries of enumeration areas, faulty maps or coverage error made by field staff in the pre-census listing exercise. The enumerator canvassing errors can be caused by ambiguous definition of enumerator assignments, faulty enumeration areas (EA) maps, and laxity of an enumerator. Missing of persons within enumerated dwellings can result because all or some of the persons were not present on the census date or night. Furthermore, inadvertent or

deliberate omission of some household members by the respondent may occur. This often occurs for people loosely connected to dwelling, thus, who sleep there irregularly.

1.1.2. Duplications

7. Duplications occur when persons, households or housing units are enumerated more than once. They also occur owing to enumerators overlapping of assignments owing to, perhaps, errors done during pre-census listing and EA delineation including failure by enumerators to clearly identify boundaries of EAs on the ground. In practice, the number of omissions usually exceeds the number of duplications. Thus, it is common to have a net census undercount.

1.1.3. Erroneous inclusions

8. This includes housing units, households, and persons that are enumerated in the census while they should not have been or were enumerated in the wrong place. For example, including in a census of people who died before the census date and births after the census date.

1.1.4. Gross versus net error

9. Gross coverage error represents the sum of three components of coverage error namely, duplication, erroneous inclusions and omissions. Net coverage error is the difference between over-counts and under- counts. In most cases net coverage error is negative because under-counts exceed over-counts. With respect to coverage error, gross census coverage error would consist of:

All persons omitted + all erroneous enumerations.

10. In measuring *net census coverage error* we have to recognize that omissions result in underestimating the population while duplication and erroneous inclusions results in the overstatement of the total population, therefore this has to be taken into account in the analysis. This implies that the net census coverage error would be measured by the excess or deficit of errors resulting in population underestimates over the errors resulting from population overestimates. In this regard a net census under-count exists when the number of omissions exceeds the number of duplicates and erroneous enumerations. On the other hand the net census over-count exits when the number of duplicates plus erroneous enumerations exceeds the number of missed enumerations.

1. 2. Overview of census evaluation methods

11. There are a number of methods used to evaluate censuses including: demographic analysis; interpenetrating studies used in conjunction with a current census; record checks and comparison of census data with results of existing household surveys; post enumeration surveys (the focus of this manual). Evaluation methods differ with respect to technical sophistication, data requirements and quality of results. An analytical framework can be created by grouping methods on the basis of three

criteria: (i) whether the method uses a single source of data; (ii) methods requiring multiple sources of data for matching or those that do not; (iii) the type of error to be measured (e.g. coverage and content error).

12. A typology of census evaluation methods based upon the above criteria is shown in table 1 below.

Table 1. Classification of methods for evaluation Census Error

Data sources and Methods	Type of Error			
	Coverage error		Content error	
	Net	Gross	Net	Gross
Single source of data:				
-Demographic analysis of the census	x		x	
-Interpenetration studies (as part of the census)		x		x
Multiple sources of data				
(i)**Matching studies;**				
-Post-censal matching	x	x		
-Re interview surveys			x	x
-Record checks	x	x	x	x
-Comparison with existing household surveys	x	x	x	x
- Post enumeration survey	x	x	x	x
(ii)**Non-matching studies;**				
-Demographic analysis using previous censuses	x		x	
-Comparison with administrative statistics	x		x	
-Comparison with existing household surveys	x		x	

Source: U.S. Bureau of the Census, Evaluating Censuses of Population and Housing (1985).

13. If the information for evaluating the census is from the census itself, thus a *single source* of data, there are two methodologies which can be used, namely: (i) demographic analysis and (ii) interpenetrating subsamples which may be carried out as part of the census operations. As indicated later the evaluation methods on a single source, in this case, from the results of the census itself are limited.

14. If the information for evaluating the census comes from sources independent from the current census there are possibilities of a wide range of evaluation methodologies to use. Methods which depend on two or more sources of data are more powerful in assessing the different contributions of different types or error. The following methods fall into this category: post enumeration surveys; record

checks; comparisons using results of existing household surveys. These methods are in the category of matching studies. Non-matching studies will include: demographic analysis using previous censuses; comparison with administrative records; and comparison with existing households survey results.

1.2.1. Demographic Analysis of census results

15. By undertaking demographic analysis, results from a census may be compared with data from other demographic systems such as vital registration of births and deaths including net migration if such data are available. The cohort component method of demographic analysis uses data from successive censuses as well as life-table survival rates; age-specific fertility rates; and estimated levels of international migration between censuses. The population is projected forward to the reference date of the second census based on estimated levels and age schedules of fertility, mortality and migration and the expected population is compared with enumerated population in the second census.

16. Another method of analysis involves comparing age distributions of successive censuses. This method is widely used because it requires little data. Yet another method, in use, is the cohort survival regression method which uses population counts by age from two censuses and deaths by age during the inter-censal period to estimate coverage rate.

17. In general, population characteristics from the current census are compared to characteristics from other independent sources such as the vital statistics register, if a country has one, or other external models such as the projections from the previous censuses. For an overall assessment of census quality, data on sex, and age-groups or age cohorts can be used. The age pyramid is another example, of a standard method demographic analysis. Stable population analysis can also be undertaken as long certain assumptions are met, such as constant fertility and mortality rates and a closed population with no migration into and out of the population. In countries where mortality has been declining a quasi-stable model may be appropriate.

1.2.2. Interpenetration studies on current census

18. This method involves drawing subsamples, selected in an identical manner, from the census frame, with each subsample capable of providing a valid estimate of the population parameter. Assignment of census personnel (enumerators, coders, data entry staff, etc.) is also done randomly. This method helps to provide an appraisal of the quality of the census information, as the interpenetrating subsamples can, for example, be used to secure information on content error. In censuses and surveys, non-sampling errors, for instance, arise from differential interviewer bias, different methods of eliciting information, etc. After the subsamples have been enumerated by different groups of interviewers and processed by different teams of workers at the tabulation stage, comparison of estimates based on the subsamples provides a broad check on the quality of the census results. The results, from such studies, could be useful in improving the operations of future censuses and large-scale sample surveys.

1.2.3. Record checks

19. In this type of analysis, census records are matched with a sample of records from the vital registration or other identification systems, where the relevant respondents to the census questionnaire are traced to the time synchronized with the census. Such sources include previous censuses; birth registrations; school enrolment registries; voter registration list; health and social security records; immigration registers; national or citizen registration cards; etc. Both coverage and content errors could be measured through such comparisons. For coverage evaluation purposes, the following pre-conditions are necessary: (i) a large proportion of the census target population should be covered in the record system; (ii) the census and record system should be independent from each other; and (iii) there should be sufficient information in the records so that accurate matching is possible. For content evaluation purposes, the record system should contain some relevant items covered in the census such as age, sex, education, income etc. It is important to ensure that the definitions of items are the same. Countries that have used record checks include: Denmark, Finland, Guatemala, Honduras, Israel, Italy, the Netherlands, Norway, Sweden, Taiwan and Canada.

1.2.4. Comparison with existing household surveys

20. Theoretically, any probability sample of households or individuals can be used to evaluate coverage and content errors in a census if they have some identical items with the same concepts and definitions. However, the post enumeration survey discussed, in detail, below is specifically designed and most ideal to do so. In the absence of a post enumeration survey, other households survey results can be used to evaluate census results provided the principle of independence from the census and closeness to the census date are upheld. In addition, there should be sufficient identical information to perform accurate matching. For content evaluation, it is essential that several of the same data items are collected. It is obvious that household surveys conducted in many developing countries may not fulfil most of the above pre-requisites.

1.2.5. Overview of Post Enumeration Surveys

21. As earlier stated the primary objective of a census evaluation programme is to determine the sources and magnitude of coverage error and content error (for some selected variables). For many developing countries the post enumeration survey (PES) has become a plausible independent evaluation programme. This is partly because other independent sources of data with relevant, comprehensive and reliable information are not that common (ECA, 1999).

22. According to the *United Nations Principles and Recommendations for Population and Housing Censuses, Rev.2,* the post enumeration survey (PES) is a complete re-numeration of a representative sample of a census population followed by matching each individual enumerated in the PES with information from the census enumeration (UN, 2008). The results of the comparison are mainly used to measure coverage and content error in the context of the census. Some countries only confine the PES to evaluating coverage error. *Coverage error refers to housing units and people missed in the census or those erroneously included.* On the other hand *content errors evaluate the response quality of selected*

questions in a census and are also a basis for evaluating reliability of some characteristics reported in the census.

23. In general, an evaluation of the magnitude and direction of errors in a census is necessary in order to present to users the extent of reliability and accuracy of some characteristics reported. The evaluation, therefore, allows for better interpretation of census results by presenting limitations to users by quantitatively evaluating the accuracy of census results with respect to coverage or/and quality of responses to questions on selected variables. For some countries the results of the PES results are used to adjust census results if, for instance, there is evidence of major coverage errors.

24. In conclusion, it is should be pointed out that for many developing countries basic data, to facilitate census evaluation, are lacking or insufficient. For example, to undertake a demographic analysis there is need for very reliable data to make it possible to calculate the demographic components of the population, namely, fertility, mortality and migration. In some developing countries this data is not available. In addition, many developing countries do not have comprehensive vital registration systems; therefore, sophisticated demographic analysis to evaluate the census may not be feasible.

25. The use of interpenetrating studies and results from household surveys assumes that a country has a dynamic and well developed survey programme whose results can be used to evaluate the census. For many countries this is not the case. It is against this background that the manual focuses on post enumeration surveys as a way to go, in evaluating coverage and content error in censuses especially in developing countries. It should, however, be emphasized that the PES can only generate reliable and accurate results if the sample is well and efficiently designed; its implementation is of high standard; the matching exercise is meticulously done; and the analysis of results and estimation are correctly executed. We shall explore some of these aspects as we discuss, below, the technical aspects of post enumeration surveys.

1.2.6. Strengths and weaknesses of evaluation methods

26. Methods based upon a single source of data (refer to table 1) provide less insight into the magnitude and types of errors in the census data than the methods based upon comparison of two or more sources of data. Examples are age and sex distribution analyses, which provide a general impression of the quality of the census results, but provide little insight on relative contributions to coverage and content error. The advantage is that such methods do not require additional data to be collected for evaluation purposes and, in general, there is no need for sophisticated matching operations. Such methods can, however, complement other methods of evaluation such as post enumeration surveys.

27. Demographic analysis has the advantage that no additional data is needed to be collected to perform this analysis. Information is already available, therefore, it s less costly and where the national statistical/census office has demographers there may be no need for additional staff to carry out the

technical analysis. The limitation of single-data source methods is that they provide less insight into the type and magnitude of errors present in the sources of data.

28. Results of well designed and implemented interpenetrating subsamples can give good insights into different contributions of component errors to total error. This type of evaluation helps in the identification of operational stages that contribute to census error. However, such studies are relatively costly involving many field staff, intensive training and close supervision.

29. The PES is an independent evaluation method of a census. It demands adequate financial, human and other resources. A successful PES calls for a good sample design and survey implementation. Mention should be made that the matching exercise can be somewhat complex.

30. The major advantage of matching over non-matching studies (analysis that does not require matching censuses records with another source) hinges on their ability to provide separate estimates of coverage and content error. On the other hand the non-matching studies, because they review census results at the aggregate rather than unit level, for example, housing units, households, or persons, provide only estimates of net census error. The characteristics that can be evaluated from matching studies are much more than those for non-matching studies, which are usually limited to age and sex distributions. Matching evaluation methods, however, require high level technical skills, managerial and financial resources. In addition, for matching studies the extent to which theoretical assumptions are not satisfied can introduce a variety of biases in the results, including: (i) response correlation biases (due to the fact that the two sources may not be fully independent), (ii) matching biases (caused by faulty matches and imperfect matching rules), and (iii) out-of-scope biases.

31. In summary, since the demographic analyses are, in general, undertaken irrespective of a PES being conducted, the critical decision as to whether or not to conduct a PES lies in the quality and variety of demographic data available. Accurate data on fertility, mortality and migration levels and trends are needed. In general, there is paucity of international migration data in many developing countries, which makes the use of only demographic analyses questionable. Demographic analysis often depends on previous census data which may also be flawed. In such situations, therefore, the PES approach, though relatively complex, may be the only reliable way of evaluating census error. This, however, does not preclude complementing the PES with demographic analysis approaches in situations where requisite data are available.

Chapter 2. Post Enumeration Surveys

2.1. Objectives

32. The PES has the following specific objectives:

(a) To measure under-coverage and over-coverage of persons and in some cases households/housing units. It is possible to design the survey so that reliable estimates of under-count or over-count are obtained for geographic areas. In addition, some estimates can be made through post stratification by forming subpopulation groups such as sex and well determined age groups. Thus, the PES can indicate to census data user's specific coverage problems inherent in the census data and such errors can be quantified.

(b) To measure levels of agreement for responses to questions on selected characteristics, such as sex, age, marital status, relationship to reference person or head of household;

(c) Evaluates the comprehensiveness and definition of area primary sampling units (PSUs), such as EAs, as suitable units to be used in frame(s) of inter-censal surveys. One of the main objectives of the census is to develop a complete, accurate and up to date frame for use in inter-censal surveys, therefore through the PES the selected EAs can be evaluated;

(d) Offers the opportunity to learn from procedural and conceptual limitations in the census which need improvement in future censuses and large-scale surveys. A PES can identify erroneous procedures used in a census. In conducting subsequent censuses, some lessons learnt from the PES would be used to improve implementation and methods of future censuses. For example, in both PES and census, in some developing countries, age derived from date of birth rather than reported age is preferred, which implies that the former would be the way to go, if possible, in the next census. In some countries matching of information from the PES and the census can be problematic because respondents can have alternate names, therefore, there can be instances where persons use different names during the census and PES respectively. To improve the situation in the next census, it may be necessary to record alternate names;

(e) Provides, for some countries, a statistical basis for adjustment of census results. On the basis of net coverage rates adjustments may be made to population census results. Using results from a carefully designed and implemented PES, under or over counts can be converted into adjustment factors and the census population increased or decreased accordingly by such factors. The adjustment is done at national and at times at other large domain levels. We should point out that there are many countries that do not adjust census figures, for various reasons, on the basis of PES results.

33. It is clear from the above specific objectives that the PES responds to the needs of variety stakeholders. In general, for informed *data users, socio-economic planners* and *decision makers*, PES results of coverage error would help them to make judicious interpretation and use of census results. In addition, the measurement of content error helps *planners* of future censuses to improve the design and implementation of future censuses and large-scale surveys. Finally, the national statistical office benefits from the updated frame of EAs for use in inter-censal surveys.

2.2 Countries which have conducted and are planning to conduct post enumeration surveys

34. Table 2 below gives list countries which have conducted PESs during the 2000 and 2010 round of censuses and are planning to conduct PESs in the 2010 round of censuses 1995 -2004 and 2005 – 2014, respectively. This illustrates the interest of countries to have objectively evaluated censuses through the PES.

Table 2. Census dates during the 2000 and 2010 round of population and housing censuses

Region/Country	Year of previous census	Year of current census	PES Previous census round	PES Current census round
Afghanistan	-	2010		Yes
Albania	2001	2011		Yes
Algeria	1998	2008	Yes	
Angola	-	2010/2014		Yes
Antigua and Barbuda	2001	2011	Yes	
Argentina	2001	2010	Yes	Yes
Armenia	2001	2011	No	Yes
Australia	2006	2011	Yes	Yes
Austria	2001	2011	No	No
Azerbaijan	1999	2009	Yes	Yes
Bahamas	2000	2010	Yes	Yes
Bangladesh	2001	2011	Yes	Yes
Barbados	2000	2010	No	
Belarus	1999	2009	No	Yes
Belgium	2001	2011		No
Belize	2000	2010	Yes	
Benin	2002	2012		
Bermuda	2000	2010	No	No
Bhutan	2005	2015	No	Yes
Bolivia	2001	2011	Yes	No
Bosnia and Herzegovina	-	2011		Yes
Botswana	2001	2011	No	Yes
Brazil	2000	2010	Yes	Yes
British Virgin Islands	2001	2010	No	
Brunei Darussalam	2001	2011	No	Yes

Region/Country	Year of previous census	Year of current census	PES Previous census round	PES Current census round
Bulgaria	2001	2011	Yes	Yes
Burkina Faso	1996	2006		Yes
Burundi	-	2008		Yes
Cambodia	1998	2008	Yes	Yes
Cameroon	-	2005		Yes
Canada	2006	2011	Yes	Yes
Cape Verde	2000	2010	Yes	Yes
Cayman Islands	1999	2010		No
Central African Republic	2003	2013	Yes	
Chile	2002	2012	No	Yes
China	2000	2010	Yes	Yes
China, Hong Kong SAR	2006	2011	Yes	Yes
China, Macao SAR	2001	2011	Yes	Yes
Colombia	2005-2006	2013/2014	No	No
Comoros	2003	2013	Yes	
Cook Islands	2006	2011	Yes	No
Costa Rica	2000	2010	No	No
Côte d'Ivoire	1998	2010		Yes
Croatia	2001	2011	Yes	Yes
Cuba	2002	2012	Yes	Yes
Cyprus	2001	2011	Yes	Yes
Czech Republic	2001	2011	No	Yes
Democratic Republic of the Congo	-	2011		Yes
Denmark	2001	2011	Yes	No
Dominica	2001	2010/2011	Yes	Yes
Dominican Republic	2002	2010		Yes
Ecuador	2001	2010	No	Yes
Egypt	1996	2006	No	Yes
El Salvador	-	2007		
Estonia	2000	2011		Yes
Ethiopia	-	2007		No
Finland	2000	2010		No
France	1999	2006	No	No
Gambia	2003	2013	No	Yes
Georgia	2002	2012		Yes
Germany	1995	2011		Yes
Ghana	2000	2010	Yes	Yes
Greece	2001	2011		Yes
Greenland		2010		No
Grenada	2001	2010		No
Guatemala	2002	2012		Yes
Honduras	2001	2012	Yes	No
Hungary	2001	2011		No
Iceland		2011		No

Region/Country	Year of previous census	Year of current census	PES Previous census round	PES Current census round
India	2001	2011	Yes	Yes
Indonesia	2000	2010	Yes	Yes
Iran		2011		Yes
Iraq	1997	2010		No
Ireland	2006	2011		No
Israel	1995	2008		No
Italy	2001	2011		Yes
Jamaica	2001	2011		Yes
Japan	2005	2010	Yes	Yes
Kazakhstan	1999	2009		Yes
Kenya	1999	2009	Yes	Yes
Kuwait	2005	2010		No
Kyrgyzstan	1999	2009	Yes	Yes
Latvia	2000	2011		Yes
Lesotho	1996	2006	Yes	Yes
Liberia	-	2008	Yes	Yes
Lithuania	2001	2011		Yes
Luxembourg	2001	2011		No
Malawi	1998	2008		No
Malaysia	2000	2010	Yes	Yes
Maldives	2006	2011	Yes	Yes
Mali	1998	2009		Yes
Malta	-	2011		Yes
Mauritius	2000	2010	No	Yes
Mexico	2005	2010	No	No
Marshall Islands	1999	2010	Yes	Yes
Mongolia	2000	2010	Yes	Yes
Montenegro	2003	2011		Yes
Morocco	2004	2014		Yes
Mozambique	1997	2007	Yes	Yes
Namibia	2001	2011	Yes	Yes
Nauru	2002	2012		No
Nepal	2001	2011	Yes	Yes
Netherlands	2001	2011		No
New Zealand	2006	2011	Yes	Yes
Nicaragua	1995	2005		No
Niger	2001	2011		Yes
Nigeria		2006		No
Niue	2006	2011	Yes	Yes
Norway	2001	2011		Yes
Occupied Palestinian Territory	1997	2007		Yes
Pakistan	1998	2010	No	Yes
Panama	2000	2010	No	No
Paraguay	2002	2012	Yes	Yes

Region/Country	Year of previous census	Year of current census	PES Previous census round	PES Current census round
Peru	-	2007	-	No
Philippines	2007	2010	No	Yes
Poland	2002	2011		Yes
Portugal	2001	2011		Yes
Qatar	2004	2010		No
Republic of Korea	2005	2010	Yes	Yes
Republic of Macedonia	2002	2011	Yes	Yes
Republic of Moldova	2004	2012		No
Romania	2002	2011		Yes
Russian Federation	2002	2010	Yes	Yes
Saint Vincent and the Grenadines	2001	2011		Yes
Samoa	2006	2011	No	Yes
São Tomé and Príncipe	2001	2011		No
Saudi Arabia	2004	2010		No
Serbia	2002	2011		Yes
Seychelles	1997	2010	No	Yes
Singapore	2000	2010	No	No
Slovakia	2001	2011		No
Slovenia	2002	2010		No
South Africa	2001	2011	Yes	Yes
Spain	2001	2011		No
Sri Lanka	2001	2011	Yes	Yes
Sudan	-	2008	No	No
Suriname	2004	2011	Yes	Yes
Sweden	-	2011		No
Switzerland	2000	2010		No
Tajikistan	2000	2010	Yes	Yes
Thailand	2000	2010	Yes	Yes
Timor-Leste	2004	2010		Yes
Tokelau	2006	2011		No
Turkey	2000	2011		No
Turkmenistan	1995	2009	Yes	Yes
Uganda	2002	2012	Yes	Yes
Ukraine	2001	2011		Yes
United Arab Emirates	2005	2010		No
United Kingdom	2001	2011	Yes	Yes
United Republic of Tanzania	2002	2012	Yes	Yes
United States of America	2000	2010	Yes	Yes
Uruguay	2004	2010	No	No
Venezuela	2001	2011	No	Yes
Viet Nam	1999	2009	Yes	Yes
Yemen	2004	2014	Yes	No

Region/Country	Year of previous census	Year of current census	PES Previous census round	PES Current census round
Zambia	2000	2010	Yes	Yes

Note: The list of countries may not be exhaustive but represent those countries for which UNSD had information at the time of writing the manual

Chapter 3. Planning and Implementation of a Post Enumeration Survey

35. We begin by presenting a general discussion of planning and implementation of a post enumeration survey. In order for the PES to yield desired results, there is need to pay particular attention to the preparations that precede field work. The PES, like any other survey requires careful and judicious preparations if it has to generate accurate and reliable results. In general, the methodology of a sound PES is relatively complex; therefore, advance planning is necessary if not critical.

3.1. Planning of a PES

36. The planning of a PES should be preceded by a clear and unambiguous statement of objectives of the evaluation. This facilitates articulate planning. Planning for a PES should, to the extent possible, be synchronized with the planning for the census. It should start early and adequate resources should be devoted to it as part of the overall census programme. The success of the PES depends mainly on the availability of qualified human and other adequate resources and it has to be independent from the census operations. For some developing countries the decision to conduct a PES, is an afterthought, made just a few months before the census enumeration. At times the PES is not conducted because there are no funds allocated for its implementation. In some cases, insufficient resources are at the disposal of PES planners to support its thorough conduct. Without adequate resources, the quality of the PES results would seriously be compromised. It is therefore necessary for the organization responsible for the conduct of the PES to develop a plausible survey plan with adequate budgetary and manpower requirements clearly spelled out.

3.2. Cost

37. Cost is a determining factor as to whether a PES should be undertaken or not. There is need to have adequate financial allotment to ensure availability of qualified enumerators and supervisors; competent matching clerks; qualified data processors; adequate training for all involved; and effective operational and quality assurance in the whole PES process. The sample size in turn will depend on whether only national estimates are required. In this case the sample size can be relatively smaller compared to an overall sample size aimed at getting reliable results for many different domains. The latter will require independent estimates, which can only be reliable if the sample size is reasonably large, implying that adequate sample sizes are obtained for each domain with its specified reliability levels. Common domains include rural/urban, regional provincial or other sub-regional domains. Large samples, for example, will demand recruitment of a large pool of enumerator’s, supervisors, data entry clerks etc. This increases the cost of the PES.

3.3. Other Resources

38. The availability of resources other than finances is equally important in the planning for a successful PES. Such resources may entail assembling additional staff including statisticians with some specialization in survey methodology and sampling; demography; data processing and computer programmers; and system analysts. One way of resolving the shortage of qualified personnel is to seek, secondment of personnel from other agencies. This assumes a country subscribes to the concept of a national statistical system and treats census and PES activities as national priorities. Some countries augment their skilled professionals with technical assistance supported by cooperating partners. In addition, planning must be done for procurement or leasing of equipment such as vehicles for transporting materials countrywide and for use by some supervisory staff. Data processing equipment and software should also be planned for in advance. Detailed examination, by type, of the overall resource requirement is a must for a sound PES.

3.4. Organizational structure

39. One of the prerequisites of a successful PES is to secure full commitment to the evaluation of the census from the highest levels of management in the national statistical/population office. This is critical in view of the fact that after the census operation, in some countries, there is ambivalence to conduct a PES. The commitment, therefore, should be made to the philosophy of the PES. Managers should be convinced of the objectives of the PES and the implementation of the survey in all its stages. If there is lack of commitment the exercise is likely to fail due to cut back in resources originally allotted to the PES. In addition, it would be difficult to overcome fatigue resulting from the census exercise and other negative perceptions associated with the PES if there was no commitment. It should be noted that for most countries the same organization which conducts the census is responsible for the PES.

40. Independence between the PES and the census is a prime requirement for the dual system of estimation. Thus, as earlier stated, the validity of the PES estimates hinges on the assumption of independence between the census and PES. Efforts should, therefore, be made to separate the two to the extent possible. This would imply assigning the technical responsibility for the PES to a unit independent from the Census. This unit would ideally devote its full time to the planning and implementation of the PES. For the ideal practical actions planned for maintaining independence between the census and PES refer to Section 12.1 (c).

41. Independence between the two operations should be attained by maintaining operational independence. It is advisable to have a unit specifically designated to the PES headed by a director/manager who should guide all aspects of the PES effort. The director should devote all his/her attention to the PES activities and should not have any responsibilities related to the census. Like the director of the PES, personnel assigned to the evaluation unit should concentrate on the PES effort with no other operational responsibilities associated with the main census. We may add that the planning for training of personnel should be done early in the process.

3.5 Technical planning

42. When initiating technical planning it is important to take into account the objectives of the PES. As indicated in Section 3.1 these objectives can be many and would require stratified plans which have to be coordinated and integrated. There is need to devise detailed plans that can be assigned to designated technical planning subgroups that would handle different components of the PES. Some of the technical groups would include sample design and execution group; Methodological technical group related to the dual system estimation; and reconciliation visits. Each group would develop its own plan. Planning independently from the census should not preclude the coordination with census operations in areas such as choice of methodology and development of an evaluation design (U. S. Bureau of the Census, 1985).

43. We reiterate that a statement on clear objects is crucial to technical planning as it makes it easier to formulate the PES technical plans when the objectives are clearly understood by members of the technical group. Evaluation planners have, at least, to take into account the plans from the census. While in an ideal situation the sampling frame of the PES is supposed to be independent of the census frame, as is the case with the Australian PES, for most countries EA maps drawn for the census are used in the PES. In addition, timing of field work and data processing for the census are to be taken into account in planning for the PES. It is against this background that coordination with the census operations should be taken into account by PES planners. The evaluation programme should be planned in such a way that they fit with the census operations. However, we underscore that the philosophy of independence between the two has to be maintained.

44. The technical planning group has a responsibility to choose the suitable methodology for the PES. In doing so due consideration must be made, taking into account the prevailing conditions of statistical development of a country. New improved methods may be evolving , for example for estimating, net coverage error and imputations, however, a careful review of the methods is essential as some of developments may not easily be implemented in some countries. For example, in the U.S. Bureau of Census there were firm indications that logistic regression rather than post-stratification methodology will be used to produce dual system estimates in the 2010 Census round (U. S. Bureau of the Census, 2009). While these approaches may end up producing more accurate results, they may be too complex and new for some developing countries to immediately adopt them. Planners should, therefore, choose methods that are robust but practical taking into account the objectives of the PES.

45. A research plan is also critical to overall planning of the PES. It highlights the parameters to be included in the plan, which usually include : (i) the date set for the start of field work of the PES; (ii) the type of data to be collected; (iii) trade off between costs and levels of precision and accuracy of some characteristics in the PES; (iv) the sample design to be adopted; (v) the universe to be covered including domains, etc., in addition, planning for alternative designs should not be ignored as modification(s) of some parameters in the initial design may be necessary. For instance, while planning to use procedure C

in data collection it may be necessary to also plan for alternative procedures A or B (refer to Section 6.3).

46. The list below gives some elements of planning for PES. It is necessary to produce a detailed calendar of events with pragmatic timelines for each activity.

(a) Preparatory activities
- (i) Compilation of sample EA maps for the PES
- (ii) Plan for sample design and estimation procedures
- (iii) Development of a questionnaire
- (iv) Drafting of enumerators and Supervisors manuals
- (v) Development of matching rules
- (vi) Production of a tabulation and analytical plans
- (vii) Conducting of a Pilot- test
- (viii) Taking into consideration Pilot –test results, finalization of questionnaires, enumerators and supervisors manual and any other modifications to the design of the PES

(b) Data collection related activities
- (i) The conduct of publicity campaigns
- (ii) Recruitment and training of enumerators and supervisors
- (iii) The actual PES data collection

(c) Matching
- (i) Development of computer programme(s) for matching (if computer matching is to be used)
- (ii) Development of matching manual
- (iii) Training of clerical staff to perform matching as required (computer matching still require clerical confirmation of uncertain links)
- (iv) Initial matching
- (v) Final matching after having reconciliation results

(d) Reconciliation
- (i) Developing instructions for field reconciliation visits
- (ii) Training of enumerators and supervisors
- (iii) Carrying out the reconciliation visits
- (iv) Final decisions of un-resolved cases, to establish matching status of cases

(e) Data Processing
- (i) Selection or development of computer programmes for data entry, tabulation and estimation

(ii) Recruitment and training of data processing staff
(iii) Carrying out data entry, editing and tabulation

(f) Estimation of coverage and content errors
(i) Estimation of coverage error
(ii) Estimation content error

(g) Estimation of sampling errors of key variables

(h) Report preparation and dissemination
(i) Analysis of PES results
(ii) Preparation of PES analytical report
(iii) Dissemination of PES results

3.6. Budgeting

47. The budget for the PES indicates the financial requirements necessary for it to be conducted. It supports and guides the implementation of the PES and the construction of the timetable for producing the results. The cost estimates must be as detailed as possible. This demands, for the planners of the PES, to understand all the detailed steps envisaged for the survey operation. The budget should show cost of personnel, equipment and all other items of expense. In general, the budget will largely depend on the PES design, the precision required and geographical coverage (UN, 2008), including, to the extent possible, person-hours to be spent on each activity.

48. The budget of the PES should be done at the same time as the census one. Realizing that the census attracts considerable financial resources, budget for the PES together with that of the census makes much sense as the PES resource requirement will be considered, by those who provide resources, as part of the census programme. By the way this does not defeat the fundamental methodological philosophy of independence between PES and census. It is more of a strategic move to get adequate resources for the PES. It is advisable that detailed budget lines should be developed.

49. The following are some of the items included in a budget:

(a) Professional Staff
(i) PES director/manager
(ii) Survey methodologist
(iii) Sampling statistician
(iv) Data processor
(v) Data analyst

(b) Field staff
(i) Enumerators and supervisors (including reconciliation field staff)

(ii) Drivers

(c) Office staff (those linked to the PES)
- (i) Matching clerks
- (ii) Matching supervisors and professionals
- (iii) Data entry staff and their supervisors

(d) Equipment and other material
- (i) Microcomputers
- (ii) Field supplies
- (iii) Office supplies

(e) Transport
- (i) Vehicles
- (ii) Petrol
- (iii) Bicycles (for use by enumerators in rural areas of some developing countries)

(f) Printing and reproduction of material
- (i) Questionnaires
- (ii) Manuals and other quality assurance forms
- (iii) EA maps
- (iv) Publicity posters and other literature
- (v) Publication of PES reports

3.7. Publicity

50. In order to encourage active participation in the PES by as many respondents as possible, it is advisable to plan for and mount publicity campaigns. Advance publicity is necessary because it prepares potential respondents for the PES by soliciting their cooperation. In this way response rates may be increased. Different approaches, to publicity, may be adopted depending on prevailing circumstances in different parts of a country. For example, in urban areas in some countries television, radio and newspaper messages could complement posters, while in rural areas radio messages and posters could be used. At times it may be necessary to arrange meetings with local opinion leaders. At such meetings people would be briefed about the objectives of the PES. Further, leaders may be asked to persuade people in the selected EAs to cooperate with enumerators and provide them requisite information.

Chapter 4. Implementation of PES

51. A number of factors contribute to errors in executing a PES, among them: use of faulty maps defining enumeration areas and unclear addresses especially in rural areas; poor publicity; shortage of transport facilities and limited communication during the data collection exercise; poor planning for data collection and data processing activities coupled with resource constraints (Onsembe, 2003). For a successful PES the above mentioned causes of error need to be avoided or mitigated.

4.1. Pilot test

52. A comprehensive test of all PES procedures is advisable. This can be a dress rehearsal of the actual PES just as the pilot census is a dress rehearsal of the census. The pilot test can cover some selected administrative divisions. The aim is to test the adequacy of the entire PES plan and the PES organization. The PES pilot test should preferably be undertaken in conditions similar to the actual enumeration. This implies that the Pilot PES should immediately follow the census pilot test. The purpose of the pilot PES is to prepare for the main PES; however, while it is not a source of usable substantive data, it provides lessons pertaining to the operational aspects of data collection that can be implemented in the current census. For example, if it is found that some EAs are not clearly demarcated or defined to act as good elements of the area frame, the EAs updating exercise can be mounted before the actual census and PES. Ideally, the pilot should be taken a year before the actual PES just as the Pilot census is taken one year before the planned census (*United Nations Principles and Recommendations for Population and Housing Censuses, Rev. 2, 200*8).

53. With respect to content re-interview survey the questionnaire can be different from the main census questionnaire. It is therefore, essential to pre-test such questions to determine possible modifications. The PES pilot may also reveal issues with the Census questions that could not otherwise be identified. In general, the pilot PES provides an opportunity to test the overall field methodology before mounting the full scale PES. It also offers a chance to test the overall matching and analytical procedures.

54. Matching of records between a PES and the census is one the main features of the evaluation exercise. It should, however, be stated that it is one of the complex and challenging undertaking in a PES programme. It has to be done well for the PES results to be useful. Thus the PES pilot results are necessary inputs into the planning requirements for the matching operations. The results of a pilot contribute to the establishment of matching rules, reconciliation procedures, and logistical flow of documents between the PES and census. It may be possible to make broad estimates of precision and accuracy on the pilot PES results, such as sampling errors and certain bias components of the total mean square error.

4.2. Data collection

55. The method commonly used in data collection with respect to the PES is the personal interview method. The method entails enumerators going to households, in selected EAs, and interviewing the respondents, thereby collecting information by asking questions from the PES questionnaire. The main advantage of this approach is that the enumerator has the potential to ask probing questions. This is, in most cases, necessary in a PES. Additionally, enumerators are in an interactive mode with respondents such that they can explain to respondents the objectives of the PES when asked.

4.2.1. Questionnaire design

56. Questionnaires for the PES should be designed based on the final census questionnaire in order to facilitate an objective evaluation of the census. The PES questionnaire plays a central role in the survey process in which information is transferred from the respondents to the survey analysts. It is the vehicle through which the information needs of users are expressed in operational terms as well as the main basis of input into the data-processing system. The format and size of the questionnaire demands serious consideration. As earlier stated, it is prudent to design a questionnaire at the time of planning for the PES. It may be worth noting that if the questionnaire is to be used for recording responses by enumerators, in the field, it should be sturdy enough to survive handling. It is also advisable that the questionnaire should be designed to facilitate the collection of accurate information. Ideally, the questionnaire should be well spread out to allow for easy reading of questions and must have clear instructions. This calls for precise definitions of data to be collected including precise specifications with respect to translation of data requirements and related concepts into operational questions. In this connection, pretesting of PES questionnaire is imperative as results of the test may contribute to the final validation of the questionnaire prior to PES field work. A sample of PES country questionnaire is provided in Annex 2.

57. A good questionnaire should have the following qualities:

(a) Enable the collection of accurate data to meet the needs of potential data users in a timely manner;
(b) Facilitate the work of data collection, data processing and tabulations;
(c) Ensure economy in data collection avoiding the collection of non-essential information;
(d) Permit comprehensive and meaningful analysis and purposeful utilization of collected data.

58. The PES questionnaire, therefore, must be developed so as to yield data of the highest quality possible with emphasis on relevance, accuracy and timeliness.

4.2.2. Selection and training of enumerators

59. As earlier stated, enumerators are at the interface with the respondents. Their work is critical to the success of the PES field work. It against this background that the selection of an enumerator should be given great attention and should be undertaken with care and objectivity. In general, an enumerator should be able to effectively communicate with respondents and should have qualities needed to collect accurate information from respondents in a timely manner. An enumerator for the PES must have an adequate level of education and should be able to record information honestly. It is important that the selected enumerators should follow instructions and use definitions and concepts as provided in the enumerators' manual.

60. The selected enumerators should be thoroughly trained before being assigned to do field work. It should be noted that the main objective of the training programme is to enhance uniformity and minimize measurement error, in interviewing procedures of the PES. This is important, in order, to avoid differing interpretations, by enumerators, of the definitions, concepts and objectives of the PES, thereby minimizing enumerator bias. Qualified instructors, who are well versed in the objectives of a PES, should be responsible for training. It is advisable that the trainers should be part of the PES planning and implementing team. In addition to following lectures, trainees should take turns in explaining to others the various items in the questionnaire. In addition, practical sessions should be arranged both in the classroom environment and actual field situation. The training programme should result in decision by the PES director of which trainees may require additional training and whether any of them are entirely unsuited for the assignment.

4.2.3. The role of supervisors

61. It is recognized that training is a prerequisite of effective and successful PES field work. Notwithstanding the foregoing, training without proper supervision may not yield accurate results. In reality the success of PES fieldwork demands dedicated and effective supervision by supervisors that are supposed to be more experienced and better qualified than the enumerators. Like the enumerators, supervisors should undergo extensive training in all aspects of the PES. It should be underscored that a supervisor is an important link between the PES planners or management and the enumerator. The supervisor is supposed to organize work for enumerators, by determining field assignments. They review completed work and maintain a high level of commitment, by enumerators, to the PES programme. In order to achieve the above it is suggested that about five enumerators, at the maximum, should be assigned to one supervisor. A supervisor can play a central role in making follow up of non-respondents. Non-response is a common phenomenon during a PES field work, just like in any other surveys. Some would be respondents may refuse to cooperate and there may also be item non-response. Because supervisors are supposed to have better qualifications and more experience than enumerators, there are therefore in the best position to contact the non-respondents and try to collect the requisite information.

4.2.4. Field data collection

62. For the procedure C (refer to Section 6.3) which is commonly used and suggested here, the PES questionnaire should allow for the classification of each listed person in a particular household as non-mover; out-move, in-movers or out-of scope.

Table 3. PES enumeration status

Serial number	Status
1	Non-mover
2	Out-mover
3	In-mover
4	Out-of-scope

63. The above population parameters are defined or described as follows:

(a) ***Non-mover :*** a person who resided in a particular household as of the census date and still do at the PES time;

(b) ***Out-movers***: refer to persons who lived in the household on the census date but did not live in the household by the PES date;

(c) ***In- mover***: a person who lived in the household on PES date but did not live in the household on the census date. Such a person arrived in the household after the Census;

(d) ***Out-of-scope***: a person who does not belong to the target population as of the Census date, for example, a child born after the census date and/or in a de facto census a person who lived outside the country on census date but lived in the household on PES date.

64. Using a well designed questionnaire it is imperative to thoroughly probe during the interview, to make sure that the *correct enumeration status* of each person, in household, is recorded. For movers, when in doubt, it may be helpful to ask for the exact date of arrival or departure of a household member, in order to determine if the person is an in-mover or out-mover, respectively.

65. The following are some of the socio-demographic variables included in a census questionnaire and repeated in the PES questionnaire for matching content error. The listed variables below are relatively easy to measure and are considered important demographic and social variables worth measuring response error if any. They include:

(a) Age
(b) Sex
(c) Relationship to head of household or reference person
(d) Marital status
(e) Education level
(f) Type of housing unit

66. Probing questions, more often than not, are useful for identifying persons that a proxy respondent unintentionally may have omitted from the household listing. This is common in the case of babies, visitors, persons travelling, etc.

67. Below are some illustrative questions included in the PES questionnaire for the 2001 PES in South Africa.

NAME (P-00 question on household listings)
• Please tell me the names of all persons who spent the night between 6 and 7 November in this household. • In addition, tell me the names of any person who did not spend the night between 6 and 7 November, but who did spend the night between 9 and 10 October in this household.

In an attempt to determine presence for each person the questionnaires allowed for the classification of each listed person as non-mover, out-mover, in-mover or out-of-scope with regard to their household presence status on the census night. The questionnaire, therefore, sought to ascertain the whereabouts of each person listed for each of the two reference nights, whether present in the household or not present (elsewhere, unborn, deceased).

PRESENT ON PES NIGHT (P-02)	PRESENT ON CENSUS NIGHT (P-03)
Based on P-00 write down where the person spent the night between 6 and 7 November • In this household • Elsewhere • Deceased	Ask: Where did (person) spend the night between 9 and 10 October? • In this household • Elsewhere • Unborn

4.2.5. Interviews

68. As earlier stated, the method of face-to-face interview is currently the most common method in post enumeration surveys. As indicated above the questionnaire solicits for information on persons who resided at a sample address on census day/night and further asks questions to ensure that the persons should have been enumerated at the current address. Subsequently, it searches for them at that address in the search area surrounding it (Whitford and Banda, 2001). Because people move, it is advisable that the PES interview should take place soon after the census. In particular, getting

information about out-movers, people who moved out of the sample address between census and the PES dates, is usually difficult.

69. The classic and basic set of questions include:

(a) What are the names of all people living here on the census day/night? Be sure to include babies, older persons, and persons away on holiday or in hospital;
(b) What is the relationship to the head of household or reference person?
(c) What is each person's age and sex?
(d) Is each person still residing here? If not what is the person's current address?
(e) What are the names and relationships of other people who lived here on census day/night?
(f) Has anyone who is currently living in the household or lived in the household during the census been missed?

70. To ensure that the person's did not live elsewhere on census day/night or missed, the interview questions like the last three questions above are used to probe (U.S. Bureau of the Census, 1985).

4.2.6. Attributes of a good PES data collection programme

71. With reference to procedure C the collection of data in a P Sample (refer to Section 6.1) focuses on identifying all residents at the time of the census in addition to those as of the date of the PES.

72. The PES data collection should take place as soon as possible after the Census to minimize matching difficulties related to changes in the composition of the household between the census and PES dates. On the other hand enough time should be given to allow for retrieval of all census materials from the field to avoid possible contamination between the census and PES operations.

4.2.6.1. Independence

73. The Dual System Estimation methodology is based on the assumption that the PES is an independent collection from the Census. The method is modelled on the technique of capture-recapture commonly used to estimate the population of wildlife. The methodology assumes a closed population. The assumption being that the population remains unchanged during the period of the study. Independence, therefore, requires that the PES must not be influenced by what took place in the census. For example, in theory, the frame should be independent from that of the census; the planning of a PES must be done by and independent group of people; in addition the PES implementation should be independent from that of the census. This implies that different enumerators and supervisors are supposed to be selected and deployed for the PES from the census. However, what is maintained is operational independence, from the census at every stage of the PES such as enumeration, data

processing and administering the survey (refer to Section 12.1 (c)). In order to maximize the independence between the two exercises, in some countries, the enumeration in the census is through the self-completed questionnaire, while the PES is conducted face-to-face and the frame for PES is independent from the census enumeration areas. In addition, the estimation procedure is based on the case-by-case matching of two different and independent sources describing the same event.

4.2.6.2. Quality assurance

74. Just like in a census and any good sample survey quality control measures are essential to ensure the quality of the operations. This should ensure the completeness of enumeration, control of response error and non-response when classifying enumerated cases. It should be mentioned that pre-tests help identify control measures necessary to ensure quality enumerations at the time of the PES. A concerted effort must be made to identify all households; and it is necessary to make follow-ups of non respondents and obtain complete interviews without, preferably, any item non response.

4 .2.6.3. Control of non-sampling error

75. In all surveys there are bound to be non-sampling errors, the focus should be to minimize them. Both sampling and non-sampling errors should be controlled and reduced to a level at which their presence does not compromise the usefulness of the PES results. Non-sampling errors are particularly harmful when they are non-random, because they introduce bias in PES estimates. Such bias is complicated or difficult to measure. The best way to control non-sampling error is to follow the right procedures, in all PES activities including planning, sample design, and processing and analyzing results. Emphasis is laid on careful and intensive training of field staff.

76. The following are some of the factors which contribute to non-sampling errors in PES:

(a) Vague objectives resulting in inadequate and/or inconsistent specifications with respect to objectives;
(b) Duplication or omission of units in the PES due to imprecise definition of the boundaries of the EAs;
(c) Inappropriate methods of interviewing, observation using ambiguous questionnaires, definitions or enumerator or supervisors instructions;
(d) Lack of trained and experienced field interviewers including lack of good-quality field supervisors;
(e) Incomplete identification particulars of sampling units or faulty methods of interviewing;
(f) Errors occurring in data-processing;
(g) PES respondent doesn't know census occupants or remembers incorrectly;
(h) Adequate information for movers is inadequate or not available;
(i) Other identifying information is poorly remembered for out-movers.

Chapter 5. Frame and Sample design

5.1. Frames

77. For most countries the post enumeration survey is essentially one-stage stratified cluster household survey. The frame of area clusters is the basis for the first stage sample selection. A sampling frame is a set of source material from which a sample is selected. It provides a means for choosing units of the target population to be included in the sample. In multi-stage sample surveys more than one set of materials are necessary. Generally, in household surveys, the first stage selection is typically drawn from geographical area frames and the ultimate or last stage the selection is a list of households in selected EAs. It is important to note that in multi-stage designs the frame for each stage must be regarded as a separate component. A specific frame is, therefore, different at each s*tage*. In the case of the PES, a one stage design is adopted for many countries where EAs are selected at the first stage and all households in the selected EAs are enumerated. This approach is necessary because, in order, to estimate coverage error the total population in the household must be enumerated in a PES.

78. In any survey, including the PES it is always advisable to clearly and unambiguously identify the target population which is defined as the entire list of units/elements from which we would want to get some information/data. For example, a target population in a PES at the first stage would be well demarcated and defined enumeration areas (EAs) covering the whole country. At the next stage the target population would be total number of households in the selected EAs. After the target population has been defined it is necessary to determine and develop means of accessing it. This is made possible by developing a frame which is a list of all units/elements comprising the target population. Frames commonly used in a PES are constructed at various stages. As earlier stated, an area frame, for example, in this case would comprise of a list of EAs, forming part of the multi-stage sampling frames. In the selected EAs, lists of say housing units, households, persons etc., would form the next frame. The EA is a conveniently sized geographical unit (cluster). In some countries, EAs are constructed to contain roughly equal numbers of households, say 100 households, in order to provide manageable workloads for survey enumerators.

79. One of the preparatory activities for a population and housing census is to prepare and update maps. The drawing of EAs is partly justified by the subsequent use of maps for other purposes and particularly as frame for inter-censal sample surveys. The frame has to have an up to date listing of localities and addresses. The listings are needed, among other reasons, for the coding of place names. These listings can be updated before the full enumeration of the PES. According to Kish, 1965, a key function of a frame is to facilitate statistical inference, through a chain of links from statistical inference to the frame as follows: *statistical inference* $\Rightarrow$ *measurability* $\Rightarrow$ *probability sampling* $\Rightarrow$ *mechanical selection* $\Rightarrow$ *frame.*

80. The following are some of the characteristics of a good frame:

(a) Each unit/element should be included in the universe. If some units are excluded you will introduce bias in the measurements especially if the characteristics of the excluded units are different from those included;

(b) There should be no duplicate units. Having duplicates in the selection process would introduce bias as some units will have unintended high probability/ chances of selection;

(c) Each unit should be well defined and distinguished from the other units. This ensures that if the unit is selected it will be identified and accessed;

(d) The frame should be updated with current information of boundaries if the unit is an EA. All the key information should be current in order to facilitate easy collection of information, for example names, addresses, roads/streets should be updated. In rural areas of some developing countries if EA maps were not updated, say between two censuses, some villages would disband, some split, new ones formed and names change. Therefore, it is always a good strategy to pay special attention to the updating of frames.

81. We list below the good attributes of an EA (i) they should preferably cover the country's entire land area; (ii) the boundaries must be well delineated with clear physical boundaries; (iii) they should be mapped; (iii) preferably they should have measures of size such as household listings and population figures.

82. One of the main objectives of the census is to develop an area frame of EAs which are assigned to enumerators to collect household and housing unit information, commonly, on a complete enumeration basis. In addition, these EAS are found to be useful as elements of area frames for inter-censal surveys, for many countries, including the PES. The sampling frame, at each stage must, capture the target population; it should be complete, accurate and update. The mapping and its updating exercise, which is a prelude to the conduct of a census is the basis for developing an area sampling frame of EAs. As earlier stated, they form the first-stage sampling frame, which is a collection of all EAs in a country.

83. For detailed discussion on frames, sample design methodology , estimation and other relevant topics on household survey methodology refer, among others, to two United Nations Statistics Division's handbooks on *(i) Designing Household Survey Samples: Practical Guidelines* (UN, 2008*) and (ii) Household Sample Surveys in Developing and Transition Countries (UN, 2005).*

5.2. Stratification

84. In stratified sampling the sampling units in the frame (population) are categorized into *groups* called strata. The population is subdivided into heterogeneous groups that are internally homogeneous. In general, when sampling units are homogeneous with respect to the auxiliary variable, termed the stratification variable, the variability of strata estimators is usually reduced, translating to a reduced overall variability of the total sample. In order, therefore, to improve the efficiency of the PES sample design the frame is divided into homogeneous strata. Variables that are correlated with the coverage

error are chosen; among them, geographical area e.g. rural and urban, provinces, districts etc. Stratification may also be necessary in order to obtain separate domain estimates.

85. For the ultimate stage stratification, advantage can be taken of other variables correlated with the extent of coverage as for example, geopolitical subdivisions since they are delineated areas and usually posses a high degree of internal homogeneity with regard to socio-demographic variables. In urban areas, for example, stratification can be achieved through consideration of density and socio-economic status. It is common, in many countries, to clearly define, in urban areas, low, medium and high income groups by identifying residential areas.

86. Steps to be followed in developing strata are as follows:

(a) The entire population of say EAs is divided into internally homogeneous but externally heterogeneous subpopulations, for example, rural and urban;
(b) Within each stratum, a separate sample is selected from all sampling units in the stratum;
(c) From the sample obtained in each stratum, estimates such as mean, standard errors, coverage rates, Index of inconsistency etc., can be obtained. It is important that such stratum estimates can be properly weighted to obtain overall population estimates;
(d) In general, with respect to a PES proportionate sampling within strata is used when national and domain estimates are the objective of the survey.

5.3 Sample design

87. There are two sampling methods, namely, probability and non-probability samples. Probability sampling requires that each element/unit in the frame has a ***known*** and ***nonzero*** probability of selection. This property justifies the calculation of sampling errors and population estimates. On the contrary the non-probability sampling methods do not rely on probability theory. We strongly recommend that PES samples are randomly selected probability samples.

88. Before delving into the practical aspects of PES sample design we discuss some general sampling issues pertaining to sample selection; elements which go into the determination of sample size; and sampling strategies.

5.3.1. Sample selection

89. As earlier stated above, in a probability sample selection process, each of the units/elements in the frame must have a known and nonzero probability of selection. This is cardinal because it is only under these requisite properties that one can objectively generalize the population estimates from the sample to the target population. It makes it possible to estimate say sampling errors of a number of variables for which information was collected in the sample. To do a sample selection it is imperative to randomize the selection process. Random numbers can be generated by computer or a selection can be done using a book or table of random numbers. In the systematic sampling option every a k^{th} unit from

the frame is selected, a random start is chosen between 1 and k to determine the first unit to be selected.

90. The reciprocals of selection probabilities are used to calculate design weights. It should be noted that the method of sample selection depends to a greater extent on the sampling design. The more complex the design the more challenging is the sample selection.

5.3.2. Sample size

91. The decision on an appropriate sample size has many facets, among them, availability of resources ; required precision; the proportion of population subgroups with the attributes being measured; type and size of domains; variability of the attribute being measured; expected level of non-response; and the adopted sample design.

92. Resources which impact on the determination of sample size include, finance, personnel and equipment, such as, vehicles for use in supervising field work and transportation of material; bicycles for enumerators, especially in rural areas of some developing countries and computers (refer to Section 4.6). The availability of critical resources would, therefore, dictate the extent of the sample size.

93. Survey results are desired with various tolerance levels. It is common in surveys to establish acceptable margin of error and levels of confidence required. The sample size has an impact on the above. The expressions for calculating sample sizes are based on probabilistic statements, that the true population parameter is contained in an interval with a given confidence level. The precision of the interval depends on the population variance, on the degree of confidence and on the sample size. In general, the greater the desired confidence level, the wider the interval. On the other hand the width of the interval decreases as the sample size increases. The following is a classic expression of a confidence interval of a population mean $\overline{Y}$ taking into account the estimator of the population mean (sample mean) $\hat{\overline{Y}}$ assuming simple random selection.

$$P\left(\hat{\overline{Y}} \pm t_{1-\alpha}\sqrt{\left(1-\frac{n}{N}\right)\frac{s^2(y)}{n}}\right)=1-\alpha$$

Where $1-\alpha$ = the confidence coefficient for the interval

$t_{1-\alpha}$ value of t that cuts α% area in tails of a normal distribution curve

$n=$ the sample size

$N=$ total number of elements in the population

$s^2(y)=$ Estimated variance of the population parameter (refer to any survey sampling text book and handbooks in particular Kish, 1995 for its significance and how it can be estimated)

Assuming $1-\frac{n}{N} \approx 1$, it gives $n = t^2_{1-\alpha} \frac{s^2(y)}{m^2}$

94. To calculate the sample size a choice has to be made among many estimates to be measured in the PES. You base the calculation based on the most important variable. When they are many key indicators the convention, at times, is to calculate the sample size needed for each and then use the one that yields the largest sample. The sample size determination depends on the degree of precision desired for the indicators. The more precise or reliable the survey estimates must be, the bigger the sample size. The PES director must be cognizant of the impact that overly stringent precision requirements have on sample size and hence on the cost of the PES. Conversely, the director must be careful not to choose a sample size so small that the main indicator, in the domains, will be too unreliable for informative analysis of results of the PES. To recap, the sample size increases as the degree of statistical confidence increases in order to maintain a given precision. It should be noted that if auxiliary information, which is correlated to a study variable about the population is used, it is possible to have a smaller sample size for the same level of accuracy. The 95 Per cent confidence level is taken as a standard. For detailed discussion of determination of sample size read Section 4.3 in the *UN publication Designing Household Survey Samples: Practical guideline (2008)* and Sections *D and E the UN publication on Household Sample Surveys in Developing Countries (2005).*

95. In distributing the sample between the main strata, for example rural and urban in most developing countries consideration must be given to maintaining the required precision in these strata. For example, in the Zambian 1990 PES, the objective was to maintain complete levels of precision in rural and urban strata, taking into account the greater socio-demographic variability in the urban area and greater field costs in the rural area, the sample was, therefore, distributed taking into account the above realities. This implied taking a relatively high proportion of sample EAs in urban areas compared to the rural area where socio-economic characteristics were less variable.

5.3. 3. Sampling strategies

96. If PES results are to be inferred for small subgroup domains of the sample, a larger sample will be necessary to maintain the accuracy level for domain estimates. Types of probability sampling designs include: simple random sampling (SRS); Systematic sampling; stratified sampling; sampling with probability proportional to size (PPS); Cluster sampling and Multi-stage sampling. As will be seen below for the PES, a multi-stage sample design is usually adopted. In this strategy the higher-level units such as enumeration areas contain more than one elementary unit. In this case the primary sampling unit is a cluster. One would wish that a simple method of selecting a random sample of the households in the country would suffice. In practice it would be very expensive to cover a representative simple random sample of households in a country. Just having a complete frame of all households in a country may not be feasible and if available would not be updated. In order to avoid this problem clusters in the form of EAs are formed. A sample of EAs is selected; subsequently lists of households in the selected EAs are developed and enumerated.

97. For example, it may not be possible to obtain a comprehensive list of all households at the time of the PES. Conversely it is usually possible to find a complete list of EAs. In this case a sample of EAs can be selected. This represents a single-stage cluster sampling design because after selecting EAs all households or persons in those EAs are usually enumerated. Thus, in many countries the sample design for the PES would be a one-stage stratified cluster sample design. A single -stage design is common for PES because all units, for household population, are enumerated to facilitate matching with the census records. The unit of analysis in most post enumeration surveys is household members. Under the usual household survey design, however, it is advisable to select more EAs at the first stage and subsample households in the selected EAs (this decreases the clustering effect).

98. Here are some reasons justifying the adoption of cluster sampling (UN, 2008):

(a) Clustering reduces travel and other costs related to data collection;
(b) The strategy can improve supervision, follow-up coverage and other aspects that have an impact on quality of the collected data;
(c) The construction of the frame is less costly as the strategy is conducted in stages. As will be seen below, in a PES a frame requiring all units is needed only for selecting the primary sampling units (EAs). These clusters are selected at the first stage. At any lower stage, a frame is required only within the units selected at the preceding stage;
(d) Frames of larger units, for example, EAs tend to be more durable and therefore more usable at longer periods of time. While lists of small units such as households and persons tend to become obsolete within a short period of time.

99. In order to introduce efficiency in the design, EAs can be selected with Probability Proportional to size (PPS), if they have very varied measures of size, within each stratum. The number of households in each EA could be designated as measures of size. Before the selection of EAs in each stratum they could be geographically ordered in a serpentine manner to ensure an efficient systematic sample selection that provides implicit stratification within each stratum. Here is how systematic sampling can be used. In systematic sampling within a stratum, a number is taken at random from 1 to k_h inclusive using random numbers. Where k_h is the sampling interval. Assuming the random start is i then the sample will contain n_h units with serial numbers $i_h, i_h + k_h, i_h + 2k_h, -----, (n_h - 1)k + 1$. The sample consists of the first unit selected at random and every k_h^{th} unit thereafter. It is obvious that the first unit in the sample determines the whole sample. A practical example is given below.

100. An illustration of a systematic selection of EAs with probability proportionate to size will suffice. Let us assume a stratum, has a total of 32 EAs, and it has been decided to select eight (8) sample EAs from this stratum. In selecting a sample using PPS, first you determine the interval (I) using the figures in table 3 below the interval would be $\frac{4{,}493}{8} \approx 561.2$ where, 4,493 is a cumulative measure of size of all EAs and 8 is the number of EAs to be selected in the stratum. The random start is selected between 1

and 561.2 or 562 using random numbers. The random start in this case is 355.2 which falls under EA number 4 between cumulative measures ranging from 350-575. To obtain the other seven sample elements the interval figure is successively added first to 355.2. In this hypothetical example the sample would include EAs 004, 008, 011, 015, 021, 024, 027 and 031. The selection can also be done by computer.

Table 4. Illustration is of systematic selection of Clusters (EAs) with probability to size

E A number	Measure of size (MOS) (number of households)	Cumulative (MOS)	Selected EAs
001	200	200	
002	70	270	
003	80	350	
004	225	575	355.2
005	120	695	
006	55	750	
007	100	850	
008	165	1,015	916.8
009	193	1,208	
010	203	1,411	
011	75	1,486	1,478.4
012	60	1,546	
013	245	1,791	
014	170	1,961	
015	97	2,058	2,039.6
016	88	2,146	
017	124	2,270	
018	78	2,348	
019	86	2,434	
020	60	2,494	
021	220	2,714	2,601.2
022	137	2,851	
023	199	3,050	
024	210	3,260	3,162.8
025	165	3,425	
026	272	3,697	
027	208	3,905	3, 724.4
028	230	4,135	
029	68	4,203	

E A number	Measure of size (MOS) (number of households)	Cumulative (MOS)	Selected EAs
030	72	4,275	
031	108	4,383	4,285.6
32	110	4,493	

101. The following were steps followed in selecting EAs in the above example:

(a) We recorded for each EA the measure of size , in this case the number of persons;

(b) Cumulated the measures of size down the list of EAs, in the stratum. The last cumulated number is equal to the total number of persons in stratum h;

(c) Decided on the number of EAs (n_h) to be selected

(d) $k_h = \frac{N_h}{n_h}$

(e) Then we obtained a random start i_h between 1 and k_h, inclusively;

(f) Determined the selected EAs as follows using cumulative measures:
$M_{hi} = i_h + (j-1)k_h$ for j = 1, -----, n_h

(g) The j_{th} selected EA is the one for which cumulated measure of size is closes to M_{hi} without exceeding it. For example, 4,285.6 is close to 4,383 in between the range 4,275 and 4,383.

Chapter 6. Overview of methodologies and procedures for evaluating coverage and content error

102. As earlier stated, no matter how carefully a census is planned and executed; errors especially of coverage and content are inevitable. Two types of error are of special interest in the analytical framework, namely, omissions and erroneous enumerations. Where under-coverage results from erroneously omitting persons, housing units and/or households in a census. If the magnitude of under-coverage is considerably high, census data users could question the validity of the census results. In addition, if the rate of under - coverage is significant among major population subgroups, the statistics of such subgroups could be distorted.

103. Erroneous enumerations, as earlier defined includes duplicate or multiple enumerations such as enumerations that should have not occurred because housing units or persons do not exist or enumerations were wrongly assigned according to geographic or demographic subgroups. The former will lead to incorrect census totals at all levels while erroneous enumerations that are incorrectly assigned according to geographic or other subgroups will be correctly represented at the highest total level but may introduce biases into estimates of census coverage error.

104. The goal of a PES interview is to establish carefully who lived in a particular housing unit on the day the census was officially taken. The next step is to match the results from the interview to appropriate census forms in a well defined area of the particular housing unit**.**

6.1. The P and E Samples

105. The PES basically involves two samples, namely, P sample and the E sample. *The population (P) sample*: consists of a sample of EAs (clusters) drawn from the same target population but independent from the census, for the purpose of estimating ***census omissions*** when compared to census records. The estimate of erroneous inclusion provides a correction factor needed in ***the Dual System Estimate (DSE)*** of the true population. The *E sample* is an enumeration sample drawn from cases already enumerated in the census, but selected for independent re-interview for the purpose of estimating ***census erroneous inclusions*** when compared to original census records. In practice, the *E sample* can overlap completely with the P sample in order to reduce costs. It, therefore, consists of the same EAs selected for the PES. Omissions, however, are not the complete story in the art of census evaluation. Errors can and are made in a census itself that affect the overall under-and over-count measurement. As indicated earlier the following are some of the other errors:

(a) the census can contain duplicate or multiple enumerations;
(b) the census could have people or housing units designated to the wrong geographic area therefore cannot match the PES interview;

(c) People could be less than perfectly enumerated, thus, there could be insufficient information for matching to the PES interview;

(d) The census can erroneously enumerate someone who should have been enumerated elsewhere or the enumerator could make fictitious persons.

106. The resulting tally can be represented in a contingency table or two- by - two tables. It symbolically shows inputs into the Dual System Estimation of the True Population.

Table 5. Tally of a two-by-two observations

	In census	**Out of census**	
In PES	$\hat{N}_{11}$	$\hat{N}_{12}$	$\hat{N}_{1+}$
Out of PES	$\hat{N}_{21}$	$\hat{N}_{22}$	$\hat{N}_{2+}$
Total	$\hat{N}_{+1}$	$\hat{N}_{+2}$	$\hat{N}_{++}$

Where

$\hat{N}_{11}$ is an estimate of the number1 of people counted in both the census and the PES

$\hat{N}_{12}$ is an estimate of the number of people counted only in the PES

$\hat{N}_{21}$ is an estimate of the number of people counted only in the census

$\hat{N}_{22}$ is an estimate of the number of people missed by both the census and the PES

$\hat{N}_{1+}$ is an estimate of the total number of people counted in the PES

$\hat{N}_{+1}$ is the total number of people counted correctly in the census (thus erroneous inclusions are factored out

$\hat{N}_{++}$ is the estimate of the total number of people

Dual System Estimate of the size of the total population is given by:

$$\hat{N}_{++} = \left(\hat{N}_{+1}\right)\left(\hat{N}_{1+}\right) \Big/ \hat{N}_{11}$$

107. The Dual System Estimate raises the corrected census total (where erroneous enumerations are subtracted for the census population) by the total estimate of the number of people in the PES divided by the estimate of the number that matched to the census. We shall explore more of the Dual System Estimation methodology in Chapter 9.

6.2. Two-way matching

108. A two-way match is conducted between census records and the P including the E sample to identify omissions and erroneous inclusions. The matching exercise, in addition, produces an estimate of the matched population.

6.3. Alternative procedures for measuring missed and erroneous enumerations

109. There are three basic procedures that can be used in PES to evaluate coverage in censuses. The major difference in the procedures is the treatment of ***movers.*** This refers to the people whose location at the time of the PES differs from their locations at the time of the census.

6.3.1. Procedure A

110. This procedure reconstructs the households as they existed at the time of the census. A respondent is asked to identify, during a PES, all persons who were living in the sample household on census date. These persons are then matched against names address or location on the census questionnaire. From such information estimates of the number of and the per cent *matched for no-movers and out-movers* can be made.

6.3.2. Procedure B

111. Procedure B identifies all current residents living or staying in the sample household at the time of the PES. The respondent is asked to provide the address(es) or location(s) where these persons where living or staying on census date. They are then matched against names on corresponding census questionnaire(s). Estimates of the number and per cent matched for *non-movers* and *in-movers* are calculated.

6.3.3. Procedure C

112. Most PES questionnaires collect basic demographic characteristics, as in the census, such as age, sex, relationship to the designated person, and marital status. The questionnaire has to be unique to the procedure adopted. For procedure C, for example, the questionnaire is designed to obtain a listing of all persons currently living at the sample address or location and all possible locations (as in procedure B) of the members of household on a census day including a listing of persons who belonged to the sample address on census day, but were not resident at the time of the PES. The location where the latter might have been enumerated in the census (as in procedure A). Mention should be made that, one purpose of addresses is to enable in-movers to be matched to the census.

113. Note that estimates of non-movers and movers come from procedure B and match rate estimates come from procedure A. It is clear from the above that procedure C is a combination of

procedures A and B. It maximizes the features of the above two procedures thus, minimizes matching difficulties and improves the estimation of movers. The difference between procedures A and B relates to primarily to movers. It should be pointed out that while many countries use procedure C there are others countries such as Australia which used procedure B in the 2006 PES.

6.3.4. Strengths and limitations of procedures A, B, C

114. The three procedures differ mainly with respect to the handling of movers; this implies that their relative advantages and limitations also pertain to movers.

115. ***Procedure A:*** an advantage with this procedure is that the matching of non-movers and out-movers is relatively simple and inexpensive because the search is limited to the sample areas and their adjacent areas. The limitation of procedure A is that movers (out-movers) are no longer at the sample address or location. It is common that the mover goes with his/her whole family; therefore, no family member may be living at the sample address. In this case the mover may be reported by a proxy respondent. Attempting to get the new address of an out-mover can be difficult and expensive. In the case of a *de facto* census, problems associated with procedure A are multiplied. The enumerator has to ask about people who were staying at address or location during the census. This includes visitors etc. This procedure, therefore, is susceptible to the underestimation of the out-movers and because movers are more likely to be missed in a census this leads to under estimation of the census omissions.

116. ***Procedure B*** The field procedures are simpler than those of procedure A. It provides a better estimate of movers, than procedure A. However, it is associated with difficulties and higher cost of matching, as it involves searching in-movers in areas where they were during the census. Some of these areas may not be in the PES sample resulting in an extended matching exercise. This is even more complicated in most developing countries where addresses may not be existent, especially in rural areas and at best inadequate. The other problem is that one is not always sure whether failure to match indicates an actually omitted person or is incorrectly located.

117. ***Procedure C:*** procedure C is somewhat more expensive than procedures A and C since it requires enumeration of in-movers in addition to out-movers and non-movers. The method is a combination of procedures A and B, thereby taking advantages of the features of each to reduce matching difficulties, and in addition, improve the estimation of movers.

6. 4. Treatment of missing data

118. Missing data are compensated by imputation and in some cases weighting. Imputation is a statistical process for predicting data where there was non-response in the survey. Missing data can be as a result of limited resources, time constraints, refusals and errors due to processing.

119. At the outset it should be pointed out that depending on the volume and particular items that are missing, the problem may be handled either by making weighting adjustments or imputing the missing data. The U.S. Bureau used both techniques (U.S. Bureau of the Census, 1985). With respect to the P-sample and E-sample, in cases where entire households were missed the households were dropped from the sample and weighting adjustments was made. During non-interview adjustment the weights of non-interviewed P-sample housing units were spread proportionately to successfully interview housing units taking into account the similarity of characteristics.

120. However, for cases where particular data items were missing, the missing data value was imputed. In this case the imputations were performed using the ***nearest neighbour*** approach. The imputation values that were picked were relatively similar to the cases that were missing. The missing data process, therefore, consist of three basic steps. The first being an adjustment of weights of housing units to account for whole household non-response. At the second stage the missing data are imputed for individual P-sample person characteristics. At the third stage the missing data process imputes for unresolved cases, for example, unresolved P-sample residents and P-sample possible matches or E-sample correct enumerations. Similarity categories may include age, sex, race or tribe and household size.

121. In 2006 the Australian Bureau of Statistics conducted a post enumeration survey in which the non-response dwellings were imputed using the ***hot-deck methodology*** during census processing, based on information provided during the census about the dwellings and their residents. The PES was used to measure the accuracy of this imputation procedure.

122. It should, however, be mentioned that lower amounts of imputed data generally mean higher quality data. Thus, the quality of census and PES data is generally affected by the amount of information provided by respondents. Non-response in this case negatively affects the quality.

123. Through the PES we require estimates say on demographic characteristics to classify respondents by demographic characteristics, in cases where data was not collected imputations may take place.

Chapter 7. Matching Operations

124. After the process of data capture is finalized and all census data prepared for an EA, the next step is to match the two sets of results. The matching can be accomplished by computer assisted matching, followed by clerical matching of the remaining non-matches and possible matches.

125. In general, the basic process of matching involves comparing addresses, names and demographic characteristics between census and PES results. It is, therefore, an operation whereby households, housing units, and persons enumerated during a census and PES are compared for similarities. The operation is done in the office. This operation can be done by computer but for some countries manual matching is still the norm. Computer matching can first involve scanning of the questionnaires and installation of matching software. The advantages of the computer matching include speed and assurance of objectivity by using matching algorithms which are mathematically sound. Matching models have been developed which can be employed by countries (U.S. Bureau of the Census, 1985). A two-way matching is used to identify census erroneous enumerations and omissions. Under Procedure C the matching is attempted for non-movers and out-movers.

126. The process involves clerk's first gathering material for EAs assigned to them. The material may include:

(a) Address list from both the PES and census;
(b) Census questionnaires for the selected EAs;
(c) PES interview results for the EAs;
(d) Maps for the EAs used during the census and a different set from the PES.

127. It should be underscored that the gathering of materials for an EA can be cumbersome. The basic process of matching involves: (i) comparing person's names and demographic information between the PES and census forms. (ii) The matching clerk matches the P sample to the census throughout the EA or search area. The clerk also looks for duplicates within the E sample.

128. In practice, the matching is carried out in two phases. During the first phase strict matching rules are used resulting in obvious matches and possible matches. If a particular household is not found in a particular EA, for example, the search may be done in the neighbouring EAs as it is possible that some households may be enumerated as part of the neighbouring EAs both during the census and PES. This may occur in situations where EA boundaries are wrongly interpreted by field staff. After the initial or preliminary match, field reconciliation is done to obtain additional information to help resolve suspicious cases. During the final phase of matching, the possible matches are re-examined at times more following subjective or relaxed rules to determine additional matches. Reconciliation visits are mounted in order to minimize the net matching error (*i.e.* the difference between erroneous matches and erroneous non-matches).

7.1. Steps followed in carrying out matching

129. The following steps are necessary for carrying out the matching:

(a) There is need to identify and determine the EAs to be searched. This requires staff that have good knowledge of the geography of the country and the availability of relatively good maps to work with.

(b) The assigned clerks should search for households listed checking for addresses, in the PES questionnaire in a specified EA. At the same time the clerk can look for groups of names, allowance being made for misspellings and misrecording.

(c) Determine which persons listed on the PES form are listed on the census questionnaire. Possible matches should be carefully examined, paying special attention to stipulated matching rules.

(d) Cases which cannot be classified on the basis of available information should be referred to the field for reconciliation visits. This enables the collection of additional information which may facilitate proper classifications of match status.

(e) After analyzing the results of reconciliation visits, matching can be finalized.

7.2. Matching rules

130. It is imperative to prepare detailed matching rules. When developing the rules consideration must be given to the possibility of making *erroneous matches* and *erroneous non-matches.* The former is defined as cases which are classified as matches when in fact the PES case was not actually enumerated in the census. The latter are defined as cases classified as non-matches which in fact do not correspond to a case enumerated in the census. It is worth pointing out that if exact agreement of characteristics is needed to establish a match, there is bound to be an excessive number of erroneous non-matches. However, as rules are relaxed and made flexible to allow for more matches and fewer non-matches, an increasing number of erroneous matches will occur. It is, therefore, necessary to develop the matching rules designed to put in place a system which minimizes net error. The objective of the matching procedure is to determine the number of matches. The estimate can be accurate if net matching error is equal or close to zero.

131. Matching rules specify the characteristics, such as age, sex, name etc. by which persons and households enumerated in the census and PES are to be matched. Tolerance ranges, which particular records, must agree with should be defined and specified. Such tolerance ranges allow for limited degree of misreporting within the census or PES. The ranges can vary according to characteristics. A good example is that no tolerance may be made for gender differences, but relatively large tolerances may be allowed for age, especially in some developing countries where most of the population has no birth certificates. However, very careful development of matching rules should be considered, because

while net matching error might be minimized by flexibility of tolerance ranges, gross error, thus***, erroneous matches*** plus ***erroneous no-matches*** might be large.

7.3. Matching process

132. This is usually the most complex part of the PES that requires judicious implementation. Once the PES enumeration exercise is over , the re-interview households are matched to the PES households so that the case-by-case comparisons can be made. Matching can be computer based, manual or a combination of the two methods.

7.3.1. Computer-assisted matching

133. We give an example of the recent use of computer matching in PES. The Australian Bureau of Statistics used the computer matching system in matching its PES results. This system called the ***Match and Search*** **System (MSS*)*** *software* system was built explicitly for the PES processing. The system made it possible for PES processors to search, view, compare and recorded matches between the census and PES data. The processors used the system to record matches of dwellings and persons between the PES and Census and to search for persons included on the census forms at all alternative addresses. It is clear from the above description that there is still involvement from the PES processors who confirmed any decisions made.

134. With respect to matching of dwellings, an attempt was made to establish whether each dwelling included in the PES was enumerated in the Census. The activity involved searching through census information in order to locate the address of the PES dwelling. The census district number was the search unit for the MSS. Through this system strict procedures were put in place to ensure consistency and accuracy when carrying out the matching process. For dwelling non-matches all efforts were made to locate the addresses in the expected census districts or the neighbouring census districts. All the PES non-matches were confirmed by the PES Processing supervisor.

135. After the PES dwellings had been matched to the census information, person matching was done for all persons in dwellings in the PES. In most cases the same people were in the same dwelling during both in the census and census enumerations. In this case a dwelling match was found to be a strong indicator that the person match was successful within the dwelling (Australian Bureau of Statistics, 2007). When this was not the case a further search for the PES respondent was carried out at any alternative addresses provided during the PES enumeration. In situations where a corresponding census dwelling could not be found for a PES dwelling, the census district was searched in order to identify residents of the dwelling. In establishing if a PES person matched a census person, comparison was made of responses to some key variables common to both the PES and census forms.

136. Mention should be made that the MSS incorporated an algorithm that combined filed match codes to produce a final match code. In 2006, the ABS conducted experiments demonstrating the usefulness of fully automated matching in locating matches and potential matches between PES and Census. It is planned that this technology will be used in the next PES, in conjunction with clerical review of any uncertain matches.

7. 3.2 Manual matching

137. We hereby briefly describe the three tier matching followed in manual matching:

(a) The first tier of matching would involve clerks to classify definite matches only. Very tight tolerances are used at this stage, keeping erroneous matches to an absolute minimum. For example, in the South African 2001 PES the initial matching tier/phase involved searching through the census records for the selected EAs in order to identify cases corresponding to the PES enumeration records vice-versa (thus, a two-way case by case matching was conducted of the two sources);

(b) With respect to the second tier specially trained clerks are required who would use more complex rules to classify the remaining cases as definite matches and definite non-matches or status unknown. The latter, however, could be possible matches;

(c) The third tier involves supervisors and professional staff working together to resolve the status of unknown cases. Any auxiliary information found on the problem case could be used at this stage. Out of the third tier will come definite matches, definite non-matches and cases that need follow up in the field to obtain additional information?

138. It should be noted that matching characteristics may be specific depending on the country. It is advisable for planners of the matching exercise to experiment with various mixes of characteristics and tolerances in pre-tests. Collection of sufficient matching information from respondents through a pre-test or a subsample of the PES is helpful in resolving questionable cases. If detailed information is collected from a subsample of the PES, it is advisable to confine the initial matching to the subsample so that matching rules for the rest of the PES may be established. It is necessary to explicitly state and document matching rules so that each matching staff follows the same criteria. It goes without saying that a matching operation that uses explicitly stated rules will be much more controllable and statistically defensible, as it will most likely yield better estimates with smaller net and gross matching errors.

7.4 . Reconciliation interviews

139. It is advisable to carry out field reconciliation visits in sample EAs and adjacent areas following the initial or preliminary matching phase. The follow-up is directed to non-matched persons and

households. Such visits give an opportunity to identify erroneous census enumerations and the resolution of doubtful cases in order to achieve a realistic and definitive match status for every P and E sample element.

140. For persons who may appear in the census records but not in the PES, reconciliation visits will permit:

(a) a determination of whether such persons were usual residents as of census date (i.e., correctly enumerated) or whether they were erroneously enumerated;
(b) collection of additional information to help determine the final match status.

141. In the case of persons appearing on the PES list but not on the census record, reconciliation visits help in:

(a) Confirming whether such persons were usual residents as of the census date, in this case non-movers or out-movers. On the other hand whether they arrived or were born after the census (in-movers);
(b) collecting additional information to determine the final match and to confirm whether such persons were residents of particular households as of the census date.

142. Field reconciliation visits should be treated as part and parcel of the PES Dual System Estimation methodology. Such visits facilitate the follow-up of the E sample and help to validate matching status cases which were doubtful during the initial matching exercise. In addition, the results help in making final decisions to eliminate cases with insufficient information for matching.

Chapter 8. Tabulations

143. It is advisable to have a tabulation plan and dummy tables for the PES. Dummy tables are draft tabulations that include everything except the actual data. At the minimum the tabulation outline should specify the table titles and column stubs including identifying substantive variables to be tabulated; background variables to be used for classification; and the population subgroups, to which the various tables apply.

144. Various tabulations can be derived from the PES results, namely *initial tables* that facilitate the calculation of coverage error; tables documenting *coverage and content analysis*; including tables showing *standard/sampling errors; coefficient of variations and confidence intervals* for estimates of coverage error. Such tables are tentative, because some elements may not necessarily be included at the final stage. For example, some countries conduct PESs specifically only to evaluate coverage error; in that case tabulations related to content error may not be necessary. With respect to standard errors and confidence intervals PES planners may select core items to be tabulated and reported.

145. Initial tables are only an input into table series on coverage analysis, while tables on coverage analysis, content analysis and standard errors and confidence intervals are usually considered for publication.

146. Tabulations resulting from the PES are in different categories, namely as earlier stated initial tables, and those related to the measurement of coverage and content errors. The following initial tabulations have been found to be useful in analyzing coverage error (Dauphin and Canamucio, 1993):

(a) estimated number of non-movers, by age group and sex;
(b) estimated number of out-movers, by age group and sex;
(c) estimated number of matched non-movers, by age group and sex;
(d) estimated number of matched out-movers, by age group and sex;
(e) estimated in-movers, by age group and sex;
(f) estimated number of erroneous inclusions in the census, by age group and sex.
(g) estimated number of correctly enumerated persons missed in the PES, by age group and sex.

147. A sample of initial input tables into coverage analysis is given below.

Table 6. Estimated number of non-movers, by age group and sex

Age	Sex			Total
	Male	Female	Undetermined	
Under 5 years 5-14 years 15-19 years 20-29 years 30-44 years 45- 64 years 65 or above Undetermined				
Total				

148. You could also have tabulations of:

(a) match rates for out-movers, by age group and sex;

(b) actual census count for household population, by age group and sex.

149. Tabulations for estimates of coverage measures would include:

(a) E-sample estimates of the population enumerated in the census, by age group and sex;

(b) P-sample estimates of the total population, by age group and sex;

(c) Census omissions and census omission rates, by age-group and sex;

(d) Census erroneous inclusions and census erroneous inclusion rates, by age group and sex;

(e) Net census coverage error and net coverage error rate, by age group and sex;

(f) Gross census coverage error and gross census coverage error rate, by age group and sex;

(g) Dual-System Estimate of the True Population, by age group and sex.

Table 7. Census omissions and omission rates, by age group and sex

Age group	Sex						Total	
	Female		Male		Undetermined			
	Total	Rate	Total	Rate	Total	Rate	Total	Rate
Under 5 5-14 15-19 20-29 30-44 65 or above Undetermined								
Total								

150. Tabulations which are relevant for content error estimate include:

(a) Sex as reported in the census and as reported in the PES for matched persons;
(b) Age group as reported in the census and as reported in the PES, for matched persons;
(c) Family relationship as reported in the PES for matched persons, etc.

Table 8. Gender as reported in the census and PES

Sex(census)	Sex (PES)			Total
	Female	Male	Undetermined	
Female Male Undetermined				
Total				

Table 9. Net difference rate and index of Inconsistency by gender

Gender	Total consistent cases	Total (Census)	Total (PES)	Net difference rate			Index of inconsistency		
				Rate	95% Conf.Int. Lower	Upper	Index	95% Conf. Int. Lower	Upper
Female Male Undetermined									
Aggregate Index									
Gross Difference Rate (off diagonal proportion) = --- per cent Rate of Agreement (Diagonal proportion) = --- per cent									

151. Tabulations for standard errors and confidence intervals for estimates of coverage may include:

(a) summary estimates of coverage measures at the stratum/domain levels;
(b) P-sample estimate of the PES population, by age group and sex;
(c) E-sample estimate of population enumerated in the census, by age group and sex;
(d) Census omissions, by age group and sex;
(e) Census omission rate, by age group and sex;
(f) Census erroneous inclusions , by age group and sex;
(g) Net difference rate and inconsistency for sex;
(h) Net difference rate and inconsistency for age group;
(i) Net difference rate and index for relationship to head of household or reference person;
(j) Net difference rate and index of inconsistency for marital status;
(k) Census omission rate, by age group and sex;
(l) Census erroneous inclusions, by age group and sex;
(m) Census erroneous inclusion rate by age group and sex;
(n) Net census coverage error , by age group and sex;
(o) Net census coverage rate, by age group and sex;
(p) Gross census coverage error, by age group and sex;
(q) Gross census coverage error rate, by age group and sex;
(r) Dual System Estimate of the True Population by age group and sex;

Table 10. Sample of standard errors, coefficient of variations and confidence intervals for summary coverage estimates at domain or stratum level

Estimate	Standard Error	CV (%)	95 % confidence interval Lower limit Upper limit	No. of observations
Total No. of non-movers				
Total No. of matched non-movers				
Matched rate of non-movers				
Total No. of out-movers				
Total No. of matched out-movers				
Matched rate of out-movers				
Total number of in-movers				
Percentage of in-movers				
Total number of matched out-movers				
Total PES population				
Total census omissions				
Census omission rate				
Total census erroneous inclusions				
Census erroneous inclusion rate				
True Population				
Net coverage				
Net coverage error rate				
Gross coverage error				
Gross coverage error per unit enumerated				
True census population				

Table 11. Standard errors, coefficient of variations and Confidence intervals for PES estimates of Census erroneous inclusions, by age group and sex

Estimate	Standard Error	CV (%)	95 % confidence interval Lower limit Upper limit	No. of observations
Both Sexes				
Under 5 years				
5-14 years				
15-19 years				
20-29 years				
30-44 years				
45-64 years				
65 or above				
Female				
Under 5 years				
5-14 years				
15-19 years				
20-29 years				
30-44 years				
45-64 years				
65 or above				
Males				
Under 5 years				
5-14 years				
15-19 years				
20-29 years				
30-44 years				
45-64 years				
65 or above				

Chapter 9. Dual System Estimation

152. The Dual System Estimation is implemented in PES to estimate the True Population of persons in households. The Chandrasekaran-Deming estimator, assuming independence, is expressed as follows:

$$\hat{N}_{++} = \hat{N}_{+1}\left(\hat{N}_{1+} \Big/ \hat{N}_{11}\right)$$

See the interpretation of the symbols in paragraph 106.

9.1. Correctly enumerated persons

153. In order to operationalise the dual system estimator there is need, as a first step, to define, in the census, the list of persons *correctly enumerated*. This is essential because since the general *net coverage model assumes that events like duplications, non-existent or out-of-scope cases have been* identified in both the census and PES and accounted for in the estimation. According to Mule, 2008, "correctly *enumerated*" has four aspects, namely, appropriateness, uniqueness, completeness and geographic correctness.

154. *Appropriateness*: this means that the person should be included in the census. For example, if the census date is 20th August 2010, people who die before or are born after this date are not part of the "*target census population*." Indeed, records of erroneous enumerations of say, fictitious people are not part of the "*target population*."

155. *Uniqueness:* refers to the need to measure the number of people included in the census and not necessarily in census records. If there are duplicate records, the count of records must be reduced for the purpose of the Dual System of Estimation.

156. *Completeness*: means that the census record must be sufficient to identify a person. If records lack sufficient identification information, it will be difficult to ascertain whether such a person was appropriately and uniquely included in the census. It may also not be possible to determine whether the person was included in the PES. Lack of sufficient Information on addresses pose a special challenge in many developing countries.

157. *Geographic correctness:* People must be included in the census in the EAs they are supposed to be included. Enumerations outside the EA that are enumerated in the census are not considered *correctly* included in the census for Dual System Estimation.

158. In order, therefore, to come up with a good Dual System Estimates the PES organizers should try to ascertain correctly the enumerated population.

9.2. Proportion of people captured in the census

159. Having defined the set of correctly enumerated persons, the next step in the Dual System Estimation is to estimate census coverage. See the formula below.

Census coverage rate (which is called the match rate) = N_{11}/N_{1+}. Thus, matched population relative to the PES population.

160. It should be pointed out that people who move between the census reference date and the time of the PES present a challenge for designing a Dual System Estimation for census application. People who move, for example, nomads, homeless persons, street kids, etc., are more likely to be missed by the census and PES.

161. It is advisable to apply the Dual System Estimation procedure within substrata formed, for instance, by the sex-age groups. Small substrata should, however, be avoided as they are susceptible to large sampling errors and bias.

9.3. Steps followed in Dual System Estimation

162. A number of census coverage estimates can be based on initial tabulations referred to in chapter 9. Population estimates are calculated for selected population parameters. The estimates are population estimates from the P sample and the E sample (Dauphin and Canamucio, 1993). It is helpful to identify all the elements that are essential in making Dual System Estimates. We hereby assign symbols to various estimates to facilitate the developments of compact standard formulas. In this case:

$a =$ total number of non-movers (estimated from P sample);
$b =$ estimated total number of out-movers (from P sample);
$c =$ estimated total number of in-movers (from P sample);
$d =$ estimated total number of matched non-movers (based on matched cases between census and P sample);
$e =$ the total number of matched out-movers (based on matched cases between census and P sample);
$f =$ total number of erroneous inclusions in the population (from the E sample)
$g =$ total number of census cases correctly enumerated in the census but missed in the PES (from E sample)

163. Table. 12. Below gives data that is subsequently used to illustrate the calculation of some coverage error and rates.

Table 12. Initial table with hypothetical PES and matched results

Symbol	Items	Values
a	Non-movers	1,300,200
b	out-movers	165,000
c	in-movers	175,000
d	Matched non-movers	1,179,450
e	Matched out-movers	98,780
m	Matched Population	1,278,230
f	Erroneous inclusions	41,777
g	Correctly enumerated but missed in the PES	338,500

Matched Population = Matched non-movers + Estimated matched out-movers

$\hat{N}_{11} = d + e$ *= 1,278,230.*

9.4. Census population estimate

164. The census estimate is obtained as follows:

Census Population = Matched non-movers + matched out-movers + population erroneously included in the census +population correctly enumerated in the census but missed in the PES

Census population $= d + e + f + g$ *= 1,658,507*

9.5. PES sample estimate of total population

PES population = Number of non-movers + in-movers

PES population $= a + c$ *= 1,475,200*

9.6. True Population

165. This is the population estimated from the PES multiplied by the population from the census after correcting for erroneous inclusions and divided by matched population between the census and the PES.

$$\text{True population} = \frac{PES\ Population \times (Census\ Population - Erroneous\ inclusions)}{Matched\ Population}$$

= 1,475,200(1,658,507-41,777)/1,278,230* $\approx$ *1,865,861*

9.7. Net coverage error

166. This is the difference between what should have been counted, thus, the True Population and what was counted in the census.

Net coverage error = True population – Census population

= 1,865,861 - 1,658,507 = 207,354

9.8. Net coverage error rate

167. The measure is the total net error relative to the Dual System Estimate of the True Population. It is an important indicator of the quality of census coverage.

$$\text{Net coverage error rate} = \frac{\text{True population – Census population}}{\text{True Population}} * 100$$

$$= \frac{207{,}354 * 100}{1{,}865{,}861} \approx 11\%$$

9.9. Estimating census omissions

168. *Net coverage error = True population – Census population= Omissions – Erroneous Inclusions*
Then,
Omissions = True population – Census population + Erroneous Inclusions
= 1,865,861 - 1,658,507+41,777 = 249,131

9.10. Census omissions rate

169. The census omission rate is the missed population relative to the PES population estimate.

$$\text{Omission rate} = \frac{Omissions}{TruePopulation} \times 100$$

$$= \frac{249131}{1865861} \times 100 \approx 13\%$$

9.11. Coverage rate (match rate)

170. Is the matched population between the census and PES relative to PES population.

$$\textit{Coverage rate} = \frac{Matchedpopulation}{PESpopulation} \times 100$$

$$= \frac{1{,}278{,}230}{1{,}475{,}200} * 100 \approx 87\%$$

9.12. Erroneous Inclusions and its rate

171. The erroneous inclusions as earlier stated include fabrications, out-of-scope, geographic misallocations, etc. It will be recalled that the objective of the E sample is to provide an estimate of erroneous inclusions. This facilitates the correction in the Dual System Estimate of the True Population.

$$\textit{Erroneous inclusion rate} = \frac{ErreneousInclusions}{CensusPopulation} * 100$$

$$= \frac{41{,}777}{1{,}658{,}507} * 100 \approx 2.5\%$$

9.13. Gross coverage error

172. Some countries use it as an indicator of the operational quality of the census enumerations. It is the sum of omissions and erroneous inclusions.

Gross coverage error = Omissions + Erroneous inclusions

= 249,131 + 41,777 = 290,908

9.13. Gross coverage error rate per unit enumeration

173. This is the absolute gross error relative to the census enumerated population.

$$\textit{Gross coverage error rate per unit of enumeration} = \frac{(\text{Omissions + Erroneous inclusions})}{\text{Census Population}} * 100$$

$$= (290{,}908/1{,}658{,}507) * 100 \approx 18\ \%$$

174. This ratio is to be used with caution, especially at sub-national levels, as erroneous inclusions may be omissions in another location.

Chapter 10. Evaluation of Content Error

175. Evaluation of census content error involves the estimation of variance and bias components of total error in a census statistic. Consider content error of characteristics, say age. This characteristic response may be designated as x_i .

A model describing total error is:

Recorded value for observation x_i = True value + Error

$$x_i = u_i + e_i$$

Where u_i the true value for the characteristic is, in this case, age for an observation i , and e_i is the error introduced for the observation by errors committed by say respondents, enumerators, data processors, etc. In our illustration we shall assume, for simplicity, each of the N observations in the population responded to the questions related to a characteristic, such as age.

176. The above model assumes that:

(a) For every observation i in the population $(i = 1,2,---,N)$ a true value u_i for the characteristic in question exists.

(b) The observed value, and the true value u_i differ by and additive error term e_i .

(c) The error term values e_i is a random variable.

177. Content error also known as response error is estimated only for matched persons and for selected variables, such as age, education level, marital status, relationship to reference person or head of household, etc. The only response error component commonly estimated in PES is variability and not bias. Variability can be measured by four indicators, namely, (i) net difference rate; (ii) index of inconsistency; (iii) the gross difference rate; and (iv) the rate of agreement.

178. The table below has hypothetical data used to illustrate the calculation of the net difference rate (NDR) and the Index of inconsistency $(\hat{I})$.

Table 13. Age group (hypothetical data) as reported in the census and as reported in the PES

	Age group (PES)							
Age group (census)	*0-4*	*5-14*	*15-19*	*20-29*	*30-44*	*45-64*	*65 +*	*Total Census* ($x_{\bullet i}$)
0-4	*2,107*	*170*	*16*	*25*	*22*	*12*	*6*	*2,359*
5-14	*161*	*5,372*	*192*	*61*	*25*	*11*	*8*	*5,830*
15-19	*13*	*179*	*2,493*	*132*	*16*	*9*	*3*	*2,846*
20-29	*26*	*61*	*124*	*3,832*	*156*	*24*	*8*	*4,231*
30-44	*21*	*21*	*18*	*149*	*4,435*	*200*	*17*	*4,859*
45-64	*15*	*14*	*9*	*25*	*198*	*3,088*	*84*	*3,432*
65 +	*7*	*11*	*4*	*11*	*16*	*84*	*1,132*	*1,267*
Total PES ($x_{i\bullet}$)	*2,351*	*5,827*	*2,856*	*4,234*	*4,868*	*3,428*	*1,258*	*24,822*

10.1. Net difference rate

179. The net difference rate (NDR) is the difference between the number of cases in the census and the number of cases in the PES that fall under each response category relative to the total number of reported persons in both the census and PES in all response categories. We illustrate below the calculation of the i^{th} category (15-19 age group)

$$NDR_i = \frac{x_{\bullet i} - x_{i\bullet}}{n} \times 100 = \frac{(2846 - 2856)}{24{,}822} * 100 = -0.04\%$$

For $i = 1, - - - -, s$

Where $x_{\bullet i}$ = unweighted census number of cases in the i^{th} category

$x_{i\bullet}$ = unweighted PES number of cases in the i^{th} category

n = unweighted total number of reported persons in both census and PES

s = total number of response categories for characteristic x

This is a measure of bias only when the reinterview is considered more accurate than the original response.

10.2. Index of inconsistency

180. The index of inconsistency is a relative number of cases for which the response varied between the census and PES. Like in the previous case of NDR this index is calculated for each response from the i^{th} category (in our example, 15-19 age groups).

$$\hat{I}_i = \frac{\left(x_{\bullet i} + x_{i\bullet} - 2x_{ii}\right)}{\frac{1}{n}\left\{x_{\bullet i}\left(n - x_{i\bullet}\right) + x_{i\bullet}\left(n - x_{\bullet i}\right)\right\}} \times 100$$

For $i = 1,2,---,s$

Where x_{ii} =*number of cases where category* i *was given as response in both the census and the PES*

$$\hat{I} = \frac{716}{5{,}047} * 100 = 14.2\%$$

10.3. Aggregate index of inconsistency (for categories of characteristics)

$$\hat{I} = \frac{\left\{n - \sum_i^c x_{ii}\right\}}{\left\{n - \frac{1}{n}\sum_i^c x_{\bullet i} x_{i\bullet}\right\}} \times 100$$

10.4. Gross difference rate

181. The gross difference rate (GDR) is calculated for the variable as a whole such as age. It is the number of discrepancies between the census responses and PES responses relative to the total number of matched persons.

$$GDR = \left\{\frac{n - \sum_i^s x_{ii}}{n - \frac{1}{n}\sum_i^{sc} x_{ii}}\right\}$$

10.5. Rate of agreement

$$\textit{Rate of agreement} = \frac{\sum_i^s x_{ii}}{n} = \frac{22{,}459}{24{,}822} = 0.90$$

182. The following table provides guidelines for interpretation of different content error measures.

Table 14. Interpretation of content errors

Measure	Level		
	Low	Medium	High
Index of inconsistency	π 20	20-50	ϕ 50
A Absolute value of NDR relative to the mean or proportion (NDR/P)	π 20	20-50	ϕ 50
Aggregate index of inconsistency	π 0.01	0.01-0.05	ϕ .0.05

Chapter 11. Adjusting census figures

183. Adjusting of census figures can be politically sensitive, in some countries, despite knowing coverage error rates from the PES. However, there are some countries that adjust some figures taking into account the PES results. Australia, South Africa, and the United Kingdom are among countries that adjusted some census figures using PES results

184. . In general, countries could adjust census results if some errors such as coverage error are substantial and the validity of census results is questionable.

185. As the decision to adjust census figures is sensitive, it is bound to be decided at highest levels of the government bureaucracy. In some countries the decision may be taken at cabinet level. There is also some critical statistical consideration which should be very carefully weighed in. For example, where are the adjustments going to take place? Consideration must be made to what geographic domains the adjustment would cover knowing that such adjustments have an effect on demographic distributions. The level of geography is important and what variables are to be adjusted, for example some demographic characteristics. Discussion of adjustment techniques are covered in appreciable detail in the U.S. Bureau of the Census classic publication on Evaluating Censuses of Population and Housing (1985).

Chapter 12. Challenges and Conclusions

12.1 Challenges of Post Enumeration Surveys

186. There are a number of problems which constrain some countries from carrying out post enumeration surveys. We briefly present below some of the key problems and suggested solutions:

(a) In some developing countries there is lack of technical personnel with requisite skills and experience in survey methodology and in designing and implementing the whole PES process. These challenges are in areas such as, sample design, implementation, matching and estimation. However, it can be argued that if a country can design and conduct good and efficient household surveys, the chances of conducting a good post-enumeration survey are high. There is therefore need for countries to develop and maintain capacity in sample survey methodology and implementation;

(b) Lack of financial resources after investing so much in a census. It is against this background that it is advisable that, with respect to financial resources for the PES, the planning should be an integral part of the overall census Programme;

(c) It is not possible to maintain the theoretical independence required between the census and the post enumeration survey. The commonly adopted strategy is to devise practical or operational approaches of maintaining independence. For example, planning and management of a post enumeration survey has to be undertaken by personnel that is separate from census personnel. In practice independence is maintained by putting into place seemingly independent field procedures that are implemented to try and improve the enumeration in the post enumeration survey compared to the census count. These include: (i) the use of enumerators and supervisors who are better qualified than those used in the census or the best staff used in the census; (ii) assigning the post enumeration survey staff to areas in which they did not work during the census; (iii) using a questionnaire which asks more detailed and probing questions on selected characteristics; (iv) census results for designated areas should not be known by staff who are assigned to those areas; (v) all census materials from the selected PES areas should be collected before PES enumerators go into the field;

(d) The design of the survey, matching and estimation procedures may be perceived to be complex. These problems can be solved or mitigated by having good sample survey methodologists and analysts including employing qualified and well trained enumerators and matching clerks and supervisors;

(e) The post enumeration interview can be demanding. It, usually, incorporates questions to determine if a respondent should really be counted at the residence in question. In addition, the post enumeration interview takes place after the census interview, at which point the respondent may feel overburdened and not be as forthcoming with accurate information. In some cases there can be outright refusals. In order to solve some of the above problems, it is advisable to develop a relatively short questionnaire administered by qualified, motivated and well trained enumerators.

(f) Past dismal failures of conducting PESs, in some countries, discourage such countries and others from conducting post enumeration surveys. However, some successful PESs have been conducted during the last 30 years, therefore, there are many lessons to learn from and in most countries the art of survey research has improved. This may apply to some countries which experienced dismal failures in PES in the 1970's and 1980's (ECA, 1999).

(g) Questions have been raised as to the rationale of conducting a post enumeration survey when its results may not be used for adjusting census population figures. The latter is not the only and exclusive reason for conducting a PES, there are many other and compelling objectives (refer to Section 2. 1)

(h) In some countries, census planners feel that it is enough to put in place good quality assurance procedures at various stages of a census. The truth is that a census is a mammoth operation such that despite the assurance procedures put in place, error is bound to creep in. It is against this background that an evaluation of census results is still necessary despite the quality assurance procedures having been implemented in a census;

(i) Some countries are ambivalent about conducting a post enumeration survey after the fatigue resulting from the census operation. A census being a gruelling and taxing operation, which saps the energy of those associated with it, discourages some national statistical/census offices to conduct a PES. Others feel that exposure of discrepancies between the census and post enumeration survey results, to users, would be detrimental to the reputation of the census or statistical organization. With respect to fatigue this is reason why we advocate that the PES should be as independent, to the extent possible, from the census activities. So that a different team would be responsible for the PES. In the case of the ambivalence of having evaluation figures for the census, it should be recognised that it is common in all credible statistical studies to have quantitative measures of error. PES evaluation results, therefore, would enhance confidence among informed ***users*** of census data, contrary to the negative view stated above. The evaluation would not diminish the importance of the census as long as users

understand the limitations of the data and errors do not affect the major uses of the data.

187. It should further be emphasized that an evaluation of a census is necessary for a number of reasons, among them is the fact that, the population census is, for many countries, the most extensive and expensive data collection exercise. In addition, censuses have in recent years become complex. With vast amounts of resources spent, there is usually considerable pressure on census takers to ensure that census results are accurate to facilitate informed decision making at national and other domain levels. In addition, because of the massive nature of the census operation, it is inevitable that some inaccuracies such as errors of coverage and content/responses are inevitable. The major difference among countries is the extent of such error. In view of the above it becomes necessary to evaluate the census results and the PES is one of the most plausible methods of evaluations, especially in developing countries.

12.2. Conclusions

188. For most developing countries, lack of comprehensive vital statistics registration data; detailed demographic data on fertility; mortality and migration makes it difficult or impossible to evaluate a census by explicitly applying the demographic analysis approach to census evaluation. The most plausible method of evaluation for many developing countries is to resort to conducting post enumeration surveys. As earlier stated, the PES is a probability sample survey which is supposed to be conducted immediately after the census. Its results are used, mainly, to measure coverage and content errors of the main census. However, it is also found to be a useful exercise enabling the study of operational aspects of the census that can be improved in future. It should be mentioned that some countries limit the evaluation exercise to measuring only coverage error, arguing that it is more complex to measure content error. Some countries use the PES results to adjust the census data.

189. As stated above the main objectives of the PES is to evaluate the census by measuring coverage and content error. This is imperative because the census is one of most expensive and comprehensive statistical operations a country can undertake, involving many enumerators, data entry staff etc. Therefore, a census operation cannot be devoid of error. It is also a fact that for many countries, their planning is evidence based. A census is a rich source of socio-economic data, provided at small area levels which are indispensable as benchmark input data into the planning process. It is, therefore, necessary that census results are assessed to highlight some limitations, in order for *informed users* to have confidence in the data.

190. However, it should be noted that conducting a PES is demanding in terms of planning; sampling design; data collection and supervision; matching of PES and census results. The prerequisites for a successful PES are having adequate resources; qualified enumerators and supervisors; good survey statisticians and analysts; and efficient and careful implementation of all the activities related to the survey. Notwithstanding these requirements, for most countries there exist household survey programmes, implying that the staff of many national statistical systems have experience of conducting

sample surveys. We can therefore posit that most developing countries can ably conduct a PES with respect to sample design and data collection and general analysis of sample data.

191. The only unique aspect of the PES is the matching exercise which is somewhat complex. However, this can be undertaken with special training of matching clerks and judicious guidance and supervision in their work. In short, if a country can successfully carry out large-scale sample surveys, the chances of conducting a good PES are high. The PES can be conceived as yet another household survey which just demands special attention in areas such as matching.

192. The dual system estimation methodology, which is key to the PES philosophy, assumes independence between the census and PES. At the outset we should point out that, in practice, it may not be possible to attain utopian independence; however, this does not preclude the adoption of practical steps in executing a PES programme by maintaining separate activities between the PES and the census.

References

Australian Bureau of Statistics (2007). Census of Population and Housing - Undercount, Canberra

Central Statistical Office (2000). Post enumeration survey, Lusaka, Zambia.

Dauphin Marjorie and Canamucio Anne (1993) Designing and Implementation of a Post-Enumeration Survey: Developing Country Example. International Statistical Programs Centre, U.S. Bureau of the Census, Washington, D.C.

Economic Commission for Africa (1999). Census Post Enumeration Activities, Addis Ababa, Ethiopia.

Farber, James (2001). Quality Indicators of Census 200 and the Accuracy and Coverage Evaluation, U.S. Bureau of Census, Washington, D.C.

Fosu Gabriel B. (2001). Evaluation of population census data through demographic analysis. Paper presented at the United Nations Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospect. New York 7- 10 August 2001.

Mule Thomas (2008). 2010 Census Coverage measurement Estimation methodology. U.S. Bureau of the Census, Washington, DC.

Kish, Leslie (1965). Survey Sampling, John Wiley & Sons, Inc., New York.

Statistics South Africa (2004). Post-enumeration survey: Results and methodology. Pretoria, South Africa.

U.S. Bureau of the Census (1985): Evaluating Censuses of Population and Housing Censuses, statistical Training Document ISP-TR-5 Washington, D.C.

Onsembe Jason (2003). Improving Data Quality in the 2000 Round of Population and Housing Censuses, UNPFA Country Technical Services Team, Addis Ababa, Ethiopia.

United Nations, (2005). Household Sample Surveys in Developing and Transitional Countries. Studies in Methods. Series F No. 96. Sales No. E.05/ XVII.6. New York.

Unite Nations, (2008). Designing Household Survey Samples: Practical Guidelines. Studies in Methods, Series No. 98. Sales no. E06XVII.13. New York.

United Nations (2008). Principles and Recommendations for Population and Housing Censuses Rev 2. Series M. No. 67/Rev.2. Sales No. E.o7.XVII.8, New York.

Whitford David and Banda Jeremiah (2001). Pose-enumeration surveys (PES's): are they worth it? Paper presented at the United Nations Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospect. New York 7- 10 August 2001.

Annex 1. Country examples of Post Enumeration Surveys

The country practices, of carrying out post enumeration surveys (PESs), are summarized and presented, in this manual, in order to share good practices and challenges from selected countries including lessons learnt. The countries selected are from different continents and diverse statistical background. The four post enumeration surveys reported here were conducted during the 2000 round of censuses (1995- 2004) while one was carried out during the 2010 round of censuses in 2006. The country post enumeration survey reports were the sources of material included in this manual, for details refer to each country's report given in the references.

1. Australia

Australia conducted a census in August 2006. A post enumeration survey (PES) was conducted, to measure net undercount, three weeks after the census. The PES determined whether each person in the sample was counted during the census and verified their characteristics such as age, sex, usual residence status and indigenous status. In general, all people present on census night, with the exception of diplomats and their families were included on the census form at the places where they stayed. It should, however, be pointed out that whenever a census is undertaken issues relating to completeness and accuracy, of the results, tend to arise. The census is a complex operation, therefore, it is inevitable that there were some errors, for example some people may have been missed and some appeared more than once. It was against this background that a census post enumeration survey was conducted by the Australian Bureau of Statistics (ABS) immediately after the census. The study provided an independent check on census coverage and identified some demographic characteristics of persons that were missed or duplicated in the census. The following were given as possible reasons why people got missed in the census:

(i) If they were travelling and were difficult to contact
(ii) There was insufficient space in the census form/questionnaire in the household and they did not get additional forms
(iii) Some would be respondents mistakenly thought to have been counted elsewhere
(iv) The person completing the form (proxy respondent) might have thought that, for example, young babies, the elderly or visitors should not be counted in a census
(v) Some respondents owing to confidentiality concerns did not want to be enumerated (refusals)
(vi) The dwelling in which prospective respondents stayed in was missed because it was difficult to find
(vii) The dwelling they stayed in was mistakenly classified as unoccupied

Some people were counted more than once in the census, among other reasons, because:

(i) They were included on the census form at the dwelling where they usually lived, even though they stayed and were counted elsewhere on census night
(ii) They were overseas on census night and should have not been counted at all but were included on the census form at the dwelling they usually live.

Like any other census, in the 2006 census some persons were missed (under-count) and others counted more than once (over-count). Net undercount of the census is the difference between the number of persons counted in the census and the number of people who should have been counted. It was necessary to measure coverage error for the following reasons:

(i) To augment the census count for the purpose of deriving an estimate of the resident population for 30 June of the census year
(ii) To provide census data users with an assessment of the completeness of the census count, thus helping them to make informed interpretation of census results
(iii) To assess the effectiveness of census implementation procedures with a view of improving the execution of future censuses.

The 2006 post enumeration survey introduced the following improvements compared to the last PES:

(i) The scope of the PES was expanded to include remote areas and discrete indigenous communities
(ii) The PES planners introduced the use of Computer Assisted Interviewing Instruments which replaced the paper questionnaires used in the previous PESs
(iii) ABS extensively utilized a computerized Match and Search System (MSS) in processing the PES data including the recording of more detailed information (metadata) about the reasons for matching decisions
(iv) An improved estimation methodology was adopted, including: specific adjustment for non-response, weighting “later returns” and imputed dwellings designed to make the dwellings representative of this combined group in the census
(v) A new estimator was developed allowing for: (a) An enhanced model for measuring coverage response error; and (b) Handling of differences in reporting characteristics between the census and *the PES such as age, indigenous status, etc.*

1.1 Scope and coverage of the 2006 PES

While the PES was supposed to cover every person present in Australia, on census night, except for foreign diplomats and their families, however, taking into account practical realities the following persons were not included:

(i) People who were in public dwellings such as hotels, motels, hospitals etc.
(ii) Homeless people were not counted because the selection process was based on dwellings
(iii) Babies born after 8 August 2006
(iv) People in Cocos (Keeling) Islands, Christmas Islands, Australian Antarctic Territory, and Jervis Bay Territory

The PES did not obtain information about people who died between the census and PES. However, it obtained information about Australian residents who were overseas during the PES enumeration period and who departed sometime in August, provided they lived with people who remained in private dwellings in Australia. It covered remote areas and discrete indigenous communities. Inclusion of these communities in 2006 ensured that the geographic coverage of the PES was more complete than was the case in the past.

1.2 Independence from the Census

The object of the PES is to provide an independent check on census coverage, namely *operational independence and population independence.* The former implies that the operations of the census do not influence the PES and vice versa. *Population independence* implies that there should be no population subgroups missed in the census; if that is the case the chances to miss such a person or dwelling would be high. This was a tougher requirement to achieve, but the PES estimation process adjusted for this by subdividing the population into subgroups where the assumption of population independence was more likely to be true.

Operational independence of the PES from the census was maintained, to the extent possible, at every stage of the exercise, which included enumeration, processing and administration.

(i) PES sample selection was done from an independent frame
(ii) Separate staff were used in the census and PES
(iii) The PES sample was maintained confidential so that census field staff and office staff were not aware which areas were included in the PES
(iv) PES enumerators were not employed as census enumerators in the same areas vice versa

Census forms which were received after the start of the PES field work were taken to be late and were treated differently in PES estimation.

1.3 Method of enumeration

The method of face-to-face interview was employed using well trained and experienced enumerators to collect information from respondents, unlike in the census were most forms well self-completed. The PES was operationally of a small scale compared to the census. It was, therefore, safe to assume that the PES delivered an accurate estimate of coverage error of persons and dwellings. On the assumption that the census and the PES were independent from each other, the estimate of the

percentage missed by the PES but found in the census and percentage missed by the census but found in the PES was used to calculate estimates of the percentage missed by both the PES and census.

However, despite the efforts to maintain independence between the census and PES, a person who was missed in the PES may have been missed in the census. This results in *correlation bias* in the PES estimates. To minimize this bias the PES estimation took into account stratification which ensured that different subgroups had different likelihood of being missed.

1.4 Sample Design

The PES was a stratified multi-stage area sample. Australia was first divided into around 100 geographic areas. These areas were then divided into strata according to population density, remoteness and growth. The following steps were followed in selecting the sample:

(i) At the first stage, census collection districts (CDs) were selected systematically, in each stratum, with probability proportional to size (PPS)

(ii) With respect to the second stage, each selected census collection district was divided into blocks one of which was selected systematically with probability proportional to size

(iii) At the third stage , dwellings were selected systematically with equal probability in each of the selected block

For areas with thinly populated areas an additional sampling stage was undertaken that preceded the selection of CDs to ensure that the sample was not very geographically dispersed. In general, the probabilities of selection at each stage were set so that the ultimate selection of each dwelling within a state or territory had equal probability of selection (EPSEM) in the survey.

1.4.1 Special sample design for discrete indigenous community

Collection districts with identified indigenous population of more than 75 per cent were classified differently from the rest of the population for both operational and sampling reasons. The discrete Indigenous communities were clustered into “sets” comprising main communities and their associated outstations and collection districts.

The selection of main communities was done with probability proportional to size of the set. The objective was to select a representative sample, taking into account cost constraints, enumerator workloads and expected sample size. When a community was selected in the sample, the selection of dwellings within the community followed the same procedures used in the selection of private dwellings with selected blocks in the non-Indigenous Community Frame. An enumerator listed all the dwellings within the selected community. Each dwelling was given a number and dwellings were systematically selected with a random start. A selection of outstations associated within each selected main community was also included in the sample. Each outstation had an equal chance of selection and once selected all dwellings were included in the sample. The above sampling strategy ensured that within each community or territory, all private dwellings and discrete indigenous community dwellings had the same chance of selection.

1.5 Field work

Enumerators collected information by determining whether each person in the sample should have been counted in the census under the subgroup such as age, sex, indigenous status and region of usual residence. For the private dwelling sample, the enumeration was done between 1 and 24 September 2006, with an initial sample of 40,000 dwellings. For the indigenous communities frame component, the enumeration took place between 11 September and 12 October 2006 with a sample of about 350 dwellings. The response rates for private dwellings and indigenous community frame samples were 94 and 91 per cent, respectively.

1.6 Matching

Matching, as earlier stated in the manual, is a comparison of records for dwellings /households or persons from the PES with records from the census.

1.6.1 Dwelling match

The matching involved the determination whether every dwelling selected in the PES had been counted in the census. It involved searching through census information, to locate the addresses of the PES dwellings. The search was performed at the census district level, thus the collection district number was the key search unit for the *Match and Search System*. ABS put in place strict procedures to ensure consistency and accuracy when carrying out the dwelling matching. In order to declare a dwelling to be a non-match, efforts were made to locate the dwelling in the census district or the neighbouring census district. All the dwelling non-matches were verified and confirmed by the PES processing supervisor.

1.6.2 Person matching

Once a dwelling was matched, person matching was conducted for all the people in the PES dwelling. In most cases they found out that the same people were in the dwelling during the PES and census enumerations. A dwelling match was, therefore, found to be a strong indicator that the person matching would be successful within the dwelling. When the latter was not the case a further search for the PES respondent was carried out at any alternate address provided during the PES enumeration.

The persons' responses to questions pertaining to the following variables were compared between the census and the PES to determine the match status: (i) age, (ii) sex, (iii) date of birth, (iv) marital status, (v) indigenous status, (vi) country of birth, and (vii) relationship in household. Finally, a field match code was recorded to indicate the strength of the match against each of the above fields. In carrying out the matching exercise, the *Match and Search System* (MSS) incorporated an algorithm which combined field codes to produce a final person match code. The MSS was complemented with manual confirmation of the person match coded calculation prior to progressing to the next record. Fields such as name, sex, and age were considered more important in determining the personal match code.

1.7 Quality assurance

A quality assurance process included a 100 per cent recoding procedure by a different processor. No identifier was included on the workloads such that it was not possible for the PES processors to know whether they were processing an original or quality assurance workload. Discrepancies at dwelling and personal levels were flagged by the Matching and Search System. Such records where then presented to the supervisors for adjudication. This was performed in order to ensure that consistent procedures were applied when processing the records. The Supervisor reviewed the discrepancies between the original and quality assurance coding and made a decision on which one was correct. If the supervisor came to the conclusion that the coding of the above was wrong the PES record was reprocessed by the PES supervisor. The quality assurance process was also found useful in identifying areas where the processors had difficulties that needed intervention of supervisors.

1.8 Net under-count

Net undercount was defined as the difference between the PES estimate of the number of persons who should have been counted in the census and the actual census count. It should be noted that imputed persons for non-responding dwellings were included in the calculations. The table below presents some estimates of rates of undercount by region.

Table. 1 Net under-count by State/territory of usual resident

State/territory	No. of persons	Rate (%)
New South Wales	157,578	2.4
Victoria	113,596	2.3
Queensland	148,409	2.7
South Australia	36,281	2.3
Western Australia	64,150	3.2
Tasmania	9,535	2.0
Northern Territory	15,909	7.6
Australian Capital Territory	4,027	1.2
Australia	**549,486**	**2.7**

1.9 Estimation

The PES estimation involved assigning a weight to each selected PES dwelling and therefore, to each person for whom a PES response was obtained. The PES estimate of the population who should have been counted in the census was a weighted sum of the number of persons in the PES sample who should have been counted in the census. The *net undercount* was obtained by taking the PES count of people in the category that should have been counted and subtracted the census count of people in that category. While the net under-count for a category of persons was the net result of the PES estimate of *gross under-count, gross over-count,* differences in classification between the PES and census and imputation error in the census.

1.9.1 Dwelling weighting

The weighting was done in two stages. At the first stage the weighting adjusted the PES selection weight such that the adjusted weights added up to the census private dwelling count. The adjustment were done within a set of census dwelling categories, based on region, dwelling structure and type of census response as it related to occupancy. A first-stage weight adjustment was also applied to private dwellings selected in the PES but missed in the census. At the second stage, the dwelling weighting involved the application of a non-response adjustment so that responding PES dwellings represented other dwellings from which no responses were obtained.

1.9.2 Person weighting

First stage personal weight adjustments were chosen such that the PES estimate of persons enumerated in private dwellings and discrete Indigenous Community dwellings, in a set of categories, matched the actual census categories. The categories were region, sex, age by 16 age groups, and country of birth, marital status, and indigenous status or whether sampled in the indigenous community frame. The weight adjustment applied to a person did not depend on whether they responded in the census but only on characteristics reported in the PES. At the final stage the initial person weights were adjusted so that the PES estimates represented persons in non-private dwellings as private and Indigenous Community Frame dwellings. At this final stage only region, age and sex was the information used from non-private dwellings as the other benchmark variables were not sufficiently reliable.

2. Republic of Macedonia

The State Statistical Office (SSO) of the Republic of Macedonia conducted an independent post enumeration survey (PES), shortly after the census, from 16 to 22 November 2002. The census was carried out from 1 to 15 November 2002.

2.1 Objectives of the PES

The objectives of the PES were to measure coverage error and the validity of responses to some questions in the census. In the latter case the PES evaluated the quality of some census data, with

regard to the consistency of the application of instructions, definitions and classifications used during the data collection exercise.

2.2 Methodological guidelines

The main methodological principles were to ensure that the:

(i) PES was an independent sample survey
(ii) PES was conducted immediately after the census
(iii) Same definitions and classifications were used in the PES as in the census
(iv) PES had some resources, though limited, under the overall census programme
(v) PES data was matched and compared to census data.

2.3 Coverage and content error measurement

The units of observation of the PES were households, dwellings and persons. The coverage evaluation covered a number of households; number of persons belonging to households and those whose residence was not at the place where the census took place; as well as the number of dwellings.

The following characteristics were subject to the evaluation of content/response error, with respect to geographic, demographic, education and socio-economic characteristics:

(i) Place/country of birth and immigration status
(ii) Date of birth, sex and marital status
(iii) Education attainment and school currently attended
(iv) Employment, occupation, status of employment, industry, working hours and unemployment.

They also evaluated data on tenure status of households and the ownership of land and livestock. With respect to dwellings, the following were evaluated: type of living quarters, occupancy status, living floor space, availability of dwelling facilities including subsidiary rooms.

2.4 Method of data collection

The face-to-face interview method was used in data collection. Enumerators went round collecting data from respondents in selected enumeration districts. The PES was designed as an independent survey which covered census units in selected enumeration districts.

2.5 Sample Design

The frame comprised of a total 7,712 census districts areas, covering the whole country. The PES was a two-stage stratified clustered sample design. The enumeration districts (EDs) were the first stage units of selection. At the second stage 10 per cent of the households were selected systematically in each of the selected district. The stratification was geographic and by type of settlements. The country was stratified into 41 regions. The stratification by settlement entailed the grouping of the census districts into urban and rural settlements. The sample allocation in strata was by probability proportional to size (PPS), where the measures of size were number of enumeration districts and

population size in each stratum. Eighty (80) enumeration districts were included in the sample. It is stated, in the country report, that the sample size was small because PES planners took into consideration the anticipated scope of work during matching of PES data with census data; the time needed to conduct the PES, and the limited funds earmarked for conducting the survey.

2.6. Questionnaires and control forms

The survey instruments included household, dwelling and persons' questionnaires and control forms. In general, the PES instruments were much shorter compared to those of the census.

2.7 Implementation

The activities related to the implementation of the PES can be classified into three categories, namely: (i) activities undertaken prior to the data collection exercise; (ii) activities related to field work; and (iii) activities carried out after data collection.

2.7.1 Prior data collection

The activities in this category related to the preparation of lists of enumeration districts with their identification particulars and assembling of maps with descriptions of the enumeration boundaries, for the selected enumeration districts. They also paid special attention to the selection of supervisors/ instructors (28) and enumerators/ controllers (80) and their subsequent training. The supervisors and enumerators were selected from the best that participated in the census. The training took only one day and particular attention was paid to training them on the art of appropriately canvassing households and the matching process.

2.7.2 Field work

The task of an enumerator was to first visit the assigned area and prepare a strategy for visiting all households and dwellings in the selected enumeration district including referring to the map and the description of the enumeration district. After this initial exploration of the enumeration district the enumerator visited and enumerated all households and dwellings including the number of household members present on the census reference date, which was 31 October 2002. Every tenth of the household in the selected enumeration districts were re-interviewed. The supervisor helped the enumerator in the ultimate sample selection by giving the enumerator a random start, in addition to other supervisory duties.

2.7.3 Activities after data collection

2.7.3.1 Matching

Supervisors, enumerators and persons selected by local authorities matched the selected persons in the households in the PES and census. Through the matching exercise similarities and differences between the PES and census data were established which became useful in determining the matching status of various characteristics. This information was critical to the evaluation of the census quality. Finally, each supervisor prepared a detailed report on the matching process and field work. Supervisors were also responsible for delivering PES materials and reports to the State Statistical Office.

2.7.3.2 Data entry and tabulations

Data entry and processing of the PES results was done after the data entry of census results. The calculation of totals, rates, indices and confidence intervals was done using SAS software. The focus of the analysis was on the evaluation of the discrepancies between PES and census results, from which they determined coverage error.

2.7.3.3 Quality assurance

Quality assurance was maintained during all phases of the PES, namely, field work, data entry, data tabulation and analysis. To enhance, quality assurance, monitoring groups were invited from the European Commission with whom the PES organizers discussed PES results.

2.7.3.4 Results

The PES results were analyzed and reports written, solely, for the State Statistical Office purposes. However, the following were some of the findings according to the report given in the reference to this manual:

(i) The census underestimated the population by 1.4 per cent
(ii) Data on ethnicity was of very high quality
(iii) Variables on economic status were not of good quality. Discrepancies were, particularly common for categories such as self employment and unpaid family workers in rural areas
(iv) There were glaring discrepancies between the census and PES with respect to agricultural items.

The supervisors' reports brought to light the following:

(i) There were refusals in the PES although relatively marginal
(ii) Some boundaries of primary sampling units were not very well defined this contributed to some enumerators to crisscross between enumeration districts
(iii) Because in some enumeration districts they had large numbers of households, two enumerators were assigned to collect information, in cases where there discrepancies between the enumerators, this led to discrepancies in the matching process
(iv) There were some discrepancies in the number of households in some dwellings, this could be attributed to the fact that the respondents who answered questions in the census and PES questionnaires were different
(v) As earlier stated, it was reported that enumerators and supervisors were selected from the best census staff.

The general conclusion, drawn from the PES results in comparison of census results, what the 2002 census had a good coverage and the majority of census topics were of good quality. On the on the other

hand, the knowledge of knowing that agricultural items were of poor quality in the 2002 census contributed to the proper planning and implementation of the June 2007 agricultural census.

3. South Africa

South Africa carried out a post enumeration survey in November 2001 which designated 6-7 November as a PES reference night. This took place soon after the census which was conducted in October 2001 with a reference night of 9-10 October 2001. The PES provided a statistical basis for estimating census coverage error, and for adjusting some census results. In addition, the survey provided the basis for evaluation content error with respect to the following characteristics/variables: sex, age, relationship to head of household; marital status; population group; home language; and highest level of education.

3.1 Target population

The PES sought to estimate the total number of persons and households in housing units and workers hostels on the census night. The units of observation were persons who spent the census night and/or the PES night in the above mentioned living quarters.

The PES sampling frame was a database of enumeration areas (EAs) demarcated during the mapping exercise prior the census. Four major strata were apportioned in the frame, namely, formal urban, informal urban, tribal areas and formal rural. EA types included in the sample were: farms, hostels, informal settlements, small holdings, tribal settlements, and urban settlements. While industrial, institutional, recreational and vacant EA types were considered as out of scope. The domains of estimation included national, urban, non-urban (rural) and province.

With respect to living quarters the PES target population included: (i) persons living in non-seasonal housing units, and (ii) persons living in worker's hostels. The population consisting those living in other types of collective quarters or the homeless was excluded from the PES. Living quarters that were excluded included: (i) residential hotels; nursing homes (homes for the aged); student hostels; tourist hotels/motels; inns; institutions and the homeless on the street.

3.2 Sampling frames and sample design

PES methodology called for a two-way match with census records. In general, it was expected that the boundaries of PES areas were to correspond exactly to those for the census areas. Another criterion was that the area clusters, to the extent possible, must be relatively small, about 100 households and uniform. For South Africa, the primary sampling units were the census enumeration areas (EAs). The base for the area sampling frame was the geographic information system (maps and database of EAs) developed during the census mapping exercise. For perfect independence between the census and PES, the PES frame should ideally have been a separate frame of EAs. For practical reasons this was not feasible taking into account the high cost, time and resources necessary of developing an independent frame.

The PES was based on a single-stage stratified cluster sample design, where, as earlier stated, EAs were primary sampling units. Once an EA was selected the ultimate sampling units were households in housing units or hostels. All households and hostels in a sample EA were enumerated to permit comprehensive matching against census records.

3.3 Procedure for coverage data collection and analysis

Procedure C, which is a combination of procedure A and B (refer to chapter 6 of the manual for the description of the said procedures) was chosen for the South African 2001 PES. The procedure took the advantages of features of both procedures A and B by reducing matching difficulties and also improving the estimation of movers. Below is a list of characteristics of procedure C which attracted its adoption:

(i) Sought to identify all persons in a household in selected EAs on the reference night of the PES and any other persons in the household on the reference night of the census
(ii) Classified each person as either non-mover, out-mover, or in-mover with regard to the person's presence status on census night
(iii) It facilitated the matching to the census records only the persons present on the census night, thus, the non-movers and out-movers
(iv) Estimated numbers of non-movers and movers were based on in-movers (as in procedure B)
(v) Matching rates for movers were estimated based on out-movers (as in procedure A).

3.4 The P and E Samples

The PES involved two samples, namely the P and E samples. The ***P sample*** (population sample) consisted of the PES sample EAs drawn from the same target population, but independent from the census. Its purpose was to facilitate the estimation of ***census omissions*** when compared to census records. On the other hand, the *E sample* was the enumeration sample drawn from cases already enumerated in the census, but selected for independent checks for the purpose of estimating ***erroneous inclusions*** when compared to census records. Not all census enumerated cases belonged to the E sample. Cases that were out of scope for the PES were not included in the E sample. Estimates of erroneous inclusions provided a correction factor needed in the Dual System Estimation (DSE) of the ***True Population***.

Although in theory the E sample was supposed to be separate from the P sample, in practice, South Africa allowed for a complete overlap with the P sample in order to reduce costs and improve the precision of estimates. In this case the E sample comprised of the same EAs selected for the PES. A two-way match was subsequently carried out between the P sample and the E sample to identify both omissions and erroneous inclusions in the census count. The matching also produced the estimate of ***matched population***, a component required in the Dual System Estimation.

3.5 Stratification

In order to improve the efficiency of the PES sample design, the sampling frame was categorized into strata. For this purpose, variables correlated to coverage error were selected such as geographic

area because density and accessibility affect the quality of the census coverage. Stratification, in addition, facilitated the calculation of separate estimates for designated domains. The first level of stratification, therefore, corresponded to the geographic domains of estimation, namely, province, and urban/no-urban zones of residence. With regard to secondary stratification advantage was taken of other variables correlated with the extent of coverage, such as subdivisions that were delineated and each possessed a high degree of internal homogeneity with regard to socio-demographic characteristics. The sampling frame of EAs was, therefore, sub-stratified by geography type, namely, urban formal, urban informal, tribal area, and rural formal. Further, in order to benefit from implicit stratification the EAs were ordered geographically, in serpentine format, within each EA type and a systematic sampling approach was used.

3.6 Sample size and standard errors

The total sample size was limited to 600 EAs taking into account operational and budgetary constraints. The average standard error was expected to be 1 percentage point at the provincial level.

Since the reliability of the PES estimates for individual provinces depended on the sample size allocated to each province, they had to ensure a minimum number of sample EAs for the smallest provinces. Therefore, the sample size of Northern Cape Province was increased to a minimum of 40 EAs. Standard errors for selected variables for the PES results were calculated using CENVAR module of the US Bureau's CSPro/IMPS software, for the purpose of reporting on the reliability of estimates and for determining future sample sizes in similar household sample surveys. At the national level, standard errors of 0.5 percentage point were obtained. At provincial level, the standard error was within 1 percentage point as expected for all provinces except KwaZulu –Natal and Gauteng which had standard errors of 2.8 and 1.7 percentage points, respectively, mainly because of the loss of samples EAs.

Proportional allocation was adopted within domains. Within each explicit substratum, EAs were selected systematically with equal probabilities after geographic ordering them. The selection was with equal probability which permitted a self-weighting sample design in each of the explicit strata. Systematic selection offered the convenience and efficiency because it provided implicit stratification. Sampling with probability proportional to size was not possible, because EA measures of size were either nonexistent or not reliable.

Table. 2 Sample allocation and expected standard errors (SEs)

Province	**Modified proportional allocation of 600 EAs**	**Expected SE for census undercount rate**
Eastern Cape	90	0.0083
Free State	40	0.0108
Gauteng	100	0.0061
Kwazulu-Natal	100	0.0075
Limpopo	70	0.0101
Mpumalanga	50	0.0112
North West	50	0.0066
Northern Cape	40	0.0215
Western Cape	60	0.0076
South Africa	**600**	**0.0031**

We list below a sequence of activities undertaken during the 2001 PES.

(i) A pilot PES was conducted in March 2001 in 60 EAs in conjunction with a pilot census.

(ii) This exercise allowed for the evaluation and improvement of census questionnaires and provided a feed back for the census procedures

(iii) The listing and enumeration was conducted, in all, the nine provinces from 7 to 30 November 2001. All dwellings and structures within the selected EAs were listed and households in housing units and hostels were enumerated

(iv) The initial matching phase involved the searching of census records for the selected EAs in order to identify cases that corresponded to the PES enumeration records and vice versa

(v) The field reconciliation followed the initial matching

(vi) The final matching exercise used the results of the reconciliation visits to assign definitive match status to each of the pending cases

(vii) Questionnaires were manually keyed into data files. The data entry work was verified

(viii) A data validation phase took place to detect and correct missing or otherwise erroneous inclusions and to ensure data file integrity.

3.7 Data collection, matching and data processing

3.7.1 Questionnaire design

The PES questionnaire consisted of some demographic questions from the census questionnaire and questions aimed at identifying the residence and ***mover status*** of households and persons in the PES. Although the questionnaire was designed for scanning and semi-automated matching, this was not possible owing to problems with both the scanning system and the matching software. In view of the above a manual processing system was implemented.

In accordance with ***Procedure C*** the data collection identified all persons in the household at the time of the census as well as those at the time of the PES. Adopting a *de facto* enumeration this meant the persons who spent the reference nights in the household. They, therefore, enumerated every person, young or old, including babies, the elderly, visitors and non-citizens, who were present in the household on either or both reference PES night of 6-7 November 2001 and 9-10 October 2001 which was the census night.

Although the enumerators did not have to classify each person, they in reality listed:

(i) The ***non-movers***, thus those persons who were present in the household on both the PES and census nights

(ii) The in-movers, those who were present in the household on PES night but were elsewhere on census night

(iii) Those born after the census

(iv) Out-movers, those who were elsewhere on PES night but present on census night;

(v) Those who died after census (they were also treated as out-movers).

3.7.2 Field work

The main goal of the PESs enumeration exercise was to conduct an exhaustive count, of households contained in selected EAs without omission or duplication. It also aimed at exhaustive enumeration within each household and hostel of all persons present on PES night or census night or on both nights. In addition, it meant accurately measuring the requisite characteristics of the enumerated population. During the field operation efforts were made to maintain operational independence between the PES and the census.

The reference night for the PES data collection was the night of 6-7 November 2001. Originally it was planned that this would take place between 10 to 31 October 2001. It should be noted that owing

to the extension of field operations in November, the PES was delayed in some areas into December 2001. Like the census the PES was a *de facto* enumeration. Enumerators and supervisors were drawn from Statistics South Africa's household survey programme, taking advantage of their better qualifications and being experienced survey field workers while ensuring operational independence from the census. Detailed training guidelines were developed; provincial managers and assistant managers were trained at the head office who then trained field staff in the provinces.

In order to further ensure independence from the census, PES personnel did not have prior knowledge of preliminary census results for the EAs for which they were responsible and did not know in advance the EAs that were included for the PES.

Field work consisted of two phases, namely, listing and enumeration. For every selected EA field workers listed housing units and other structures (including vacant buildings, businesses, schools etc.), independent of any previous census listing. The work of an enumerator included was to:

(i) Identify all housing units and hostels within the assigned EA and all eligible persons in each household without relying on the census results
(ii) Interview all individuals in each household in every housing unit and workers hostels
(iii) Ask for the stickers that were left by the census enumerators. Enumerators were asked to get them or write down the number on the PES questionnaire.

Enumerators were expected to complete the PES questionnaire correctly following the instruction manual.

3.7.3 Data capture and data validation

It should be recalled that the PES data processing plan was for automated capture by scanning. The questionnaire and reconciliation visit forms were, therefore, designed for automated processes. Due to problems with the scanning system, a manual processing system was implemented instead. This required manual matching, key-from- paper data entry and manual processes for tracking and validation. This approach was, however, not efficient because the questionnaires were not designed for manual processing.

3.7.4 Initial matching phase

Coverage status was determined through a case-by-case comparison of the PES independently enumerated cases with the census original records. A two–way matching was conducted of the PES and **census records. This exercise helped to account for persons included in both the census and PES and** persons included in one source but excluded in the other. It also enabled the discovery and removal of ***erroneous inclusions*** in the PES and census records. Thus the initial matching phase involved searching through the census records in order to identify corresponding cases from the PES enumeration records and vice versa. Thus cases enumerated in the census but missed in the PES and those enumerated in the PES but missed in the census were identified. A matching manual was published for use by matching

clerks. Cases of possible match in the PES but not in the census with insufficient or unclear information in the PES and in the census were identified as suitable for reconciliation visits.

3.7.5 Field Reconciliation

Subsequent to the initial matching a reconciliation operation was mounted. This was an integral part of the PES Dual System Estimation methodology, since they provided follow-up for the E Sample and helped definitively identify cases with insufficient matching information.

The purpose of reconciliation visits was to determine the final status of the unresolved cases thereby:

(i) resolving the final match status for possible matches
(ii) determining whether households and persons enumerated in the census but not in the PES were correctly or erroneously enumerated in the census
(iii) to clarify doubtful cases or cases with insufficient or unclear information in order to assign a final match status
(iv) Investigating EAs where boundary or enumeration quality problems were suspected.

It should be underscored that the purpose of reconciliation visits was not to add to the PES enumeration but only to verify the already collected information. This preserved partly the independence between the census and the PES. Statistics South Africa produced a detailed manual on reconciliation visits. During the reconciliation visits enumerators also recorded whether the dwelling was found inside or outside assigned the EA boundaries and/or whether the dwelling was seasonal. In all cases an interview completion status was recorded.

3.7.6 Final matching

During this phase results of the reconciliation visits were used to assign a definite match status to each pending case.

3.8 Estimation of the True Population

As earlier stated, two operationally independent sources were used to arrive at the estimate of the ***True Population*** under the Dual System of Estimation, namely the census of October 2001 and the PES conducted in November 2001. The census was done on a complete enumeration basis while the PES was a probability sample survey. The Dual System Estimate provided an estimate of cases included in the PES and excluded from the census and vice versa. In the final analysis the ***True Population*** was compared with the ***corrected census enumerated population*** and the difference between the two was the ***net under-count***.

The census enumeration produced 35. 11 million persons and the estimate for the PES were 35.58 million. Applying the DSE methodology the estimated ***True Population*** of South Africa was 42.63 Million. The ***net under-count*** in the true population was estimated at 17.6 per cent.

3.8.1 Adjustment of households and population

South Africa is one of the countries that used its PES results to adjust some selected census results. For example, the adjusted census household total corresponded to the Dual System Estimate of the ***True Population*** of households. The adjustment procedure for households was similar to that used for persons. It involved creating homogeneous adjustment domains /classes with similar coverage rates. Such classifications were based on geography type, province, household size and population group of head of household; and then calculating a common adjusted population, under-count rate and adjustment factor for each domain. The national adjusted household total was obtained by summing across the adjusted domains. Only households in the in-scope sub-universe were adjusted. Similarly, the adjusted population was obtained by multiplying the appropriate adjustment factor to the actual census count in the adjustment domain and then summing across the domains (Adjusted population = adjusted population within in-scope sub-universe +unadjusted balance population).

In applying the procedure this involved the partitioning of the population into within in-scope sub-universe and balance of the population. Each person or household was first established to be in or out of the target population. Thus only the cases in the in-scope sub-universe were eligible for the designated adjustment factors. The non-eligible cases received the adjustment factor of 1. The eligible person was then assigned, on an individual level, the adjustment factor corresponding to the adjustment domain. Similarly, each household was assigned on an individual basis the adjustment factor corresponding to the adjustment domain it belonged to, thus the class of the head of household.

3.8.2 Sampling weights

The sampling frame consisted of 72, 487 EAs. As earlier stated the PES sample was a one-stage stratified cluster sample design. The EAs were drawn with equal probability within each explicit stratum. A systematic selection was adopted in this case. They used the SAS procedure survey select. The base sampling weight of a sample EA was equal to the reciprocal of its probability of selection.

Within each EA, the weight for each household and each person was supposed to be equal to the EA sampling weight, since their probability of selection, given the selection of the EA, was supposed to be equal to 1.

3.9 Coverage evaluation

Coverage estimates were calculated for the domain of interest. Initial tabulations in the dual system were:

(i) Estimated number of ***non-movers*** and percentage of total population represented by non-movers
(ii) Estimated number of out-movers and percentage of total population represented by out-movers
(iii) Estimated number of in-movers and percentage of total population represented by in-movers
(iv) Estimated number and rate of matched non-movers
(v) Estimated number and rate of out-movers

(vi) Estimated number of matched in-movers
(vii) Estimated number of census erroneous inclusions
(viii) Estimated number of census correctly enumerated persons missed in PES
(ix) Estimated number of and percentage of census persons with insufficient information
(x) Estimated number and percentage of PES erroneous inclusions and PES insufficient information cases

The list below gives the characteristics calculated under the Dual System of Estimation.

(i) Census population (uncorrected for erroneous inclusions and insufficient information)
(ii) Census population (corrected for erroneous inclusions and insufficient information)
(iii) PES population
(iv) Matched population
(v) PES persons missed in the census (total)
(vi) PES persons missed in census (rate)
(vii) Coverage rate
(viii) Census correctly enumerated persons missed in PES
(ix) PES persons missed in census (rate)
(x) Census erroneous inclusions (total)
(xi) Preliminary Dual System Estimated True Population
(xii) Net error (net under-count) –Rate
(xiii) Gross error (total)
(xiv) Gross error (rate relative to true population)
(xv) Adjustment factor for census
(xvi) Final Dual System Estimate of True Population

3.10 Content evaluation

Content error is the deviation of observed census value from a true value, obtained from the PES, for a given characteristic. The following variables in the census questionnaire were repeated in the PES questionnaire for matching and content error analysis:

(i) Age
(ii) Relationship to head of household
(iii) Marital status
(iv) Population group
(v) Home language
(vi) Highest level of education

It should be noted that the same concepts and definitions; wording; response categories; and pre-codes were maintained in the PES as were used in the census. The evaluation was:

(i) Limited to matched cases
(ii) Limited to the in-scope sub-universe , consisting of housing units and hostels within selected EAs in scope EA types
(iii) A comparison of unedited PES and socio-demographic responses

The table below gives selected measures of content error, namely net difference rate, index of inconsistency and gross difference rate.

Table.3 Net difference rate, index of inconsistency, and gross difference rate by age group

				Net difference rate			Index of inconsistency		
Response category	**Total consistent cases**	**Total in census**	**Total in PES**	**Rate**	**95% Confidence Interval limits**		**Index**	**95 % Confidence Interval**	
					Lower	Upper		Lower	Upper
0-4	17,038	19,068	18,974	0.05	-0.02	0.11	11.51	11.15	11.88
5-14	42,589	46,239	46,176	0.03	-0.05	0.12	10.16	9.92	10.40
15-19	19,955	22,741	22,853	-0.06	-0.13	0.02	14.05	13.69	14.43
20-29	31,518	34,733	34,763	-0.01	-0.09	0.06	11.23	10.95	11.51
30-44	37,079	40,565	40,629	-0.03	-0.11	0.05	10.85	10.59	11.11
45-64	25,690	28,455	28,467	-0.01	-0.08	0.07	11.33	11.03	11.64
65+	8,960	10,008	9,947	0.03	-0.01	0.07	10.73	10.26	11.21
Total	**182,829**	**201,809**	**201,809**	**-**	**-**	**-**			
Aggregated index of inconsistency							**11.28**	**11.12**	**11.45**
Gross difference rate = 9.4 % (off-diagonal proportion)									
Rate of agreement =	**90.6%**								

Source: Post enumeration survey: Results and methodology, Statistics South Africa 2004

4. Suriname

Suriname is one the countries which used both the Post enumeration survey estimates and demographic analytical methods to evaluate the census results of its 2004 Population and Housing census. The Demographic analysis included the following methods: Demographic analysis based on a single census; Whipple's index; Myer's Blended Index; the United Nations age-sex accuracy index; Sex ratio at birth and mean parity pattern for children ever born; Simple extrapolation; Population

projections; Inter-censal survival ratios; and the Demographic balancing equation. We briefly describe, below, each of these methods.

4.1 Demographic analysis based on a single census

The focus of the demographic analysis using one source of census results focused on measuring internal consistency and revealing demographic regularities/irregularities. They computed age-sex accuracy indices for the population of 2004. A high index value outside the acceptable interval could mean that age and/or sex error, but also selective under-coverage or age and sex selective migration. The country report did not provide the acceptable range of the index.

4.2. Whipple's Index

The Whipple's Index used the single aged distribution on the age interval 23-62 of the census population and measured the digital preferences of terminal digits 0 and 5. Deviation from 100 indicated the digital preferences. The Whipple's index for both males and females was 101.2 and, the index was 100.4 for males and 102.0 for females, respectively. In interpreting the index the following guide was adopted: < 105 very good, 105- < 110 good, and 110 - <125 satisfactory and 125 and over is considered poor.

4.3 Myers' Blended Index

This index used a single age distribution of the population, of age ranges 10-89 and 20-99 to calculate a measure of tendency for ages ending in certain digits preferred compared to others during enumeration. The index is 0 when there are no discrepancies and it is 90 when everybody has chosen to report the same terminal digit. The Myers' Blended Index for both sexes was 1.48. The index was 1.81 for males and 1.35 for females using computed age range 10-89. In ranking the indices < 10 good, 10-20 satisfactory and above 20 was considered poor.

4.4 The United Nations age-sex accuracy index

This index utilizes group data up to the age of 70 and is based on the assumption that differences between sex ratios for successive quinquennial age groups and deviations of quinquennial age rations from 100 are likely to be slight for all ages up to age 70 if the statistics are accurate and if the population has not been severely affected by birth cycles, birth deficits, military casualties and discontinuous migrations (UN, 1952). The United Nations Age-Sex Accuracy Index was 20.9 without a correction for population size, while after the correction for population size, the index was 18.5.

4.5 Sex ratio at birth and mean parity pattern for children ever born

In assessing quality or internal consistency the sex ratio at birth for questions pertaining to children ever born were used as well as the pattern of mean parity. The ideal situation is that the sex ratios at birth should not vary systematically with age of mothers. Their values should range between 102 and 107. As regards to mean parity per woman, under normal circumstances it should exhibit a pattern of increasing with age. The results from the analysis, in general, exhibited the expected pattern except the sex ratio at birth for the age group 15-19 which was outside the expected range. Refer to the table below.

Table. 4 **Boys and Girls ever born by age groups of women including Sex Ratio and Mean Parity**

Age group	No. Of women	Boys ever born	Girls ever born	Sex ratio at birth	Mean parity per woman
15-19	23,035	2,455	2,420	101.4	0.2
20-24	21,399	8,207	7,790	105.4	0.7
25-29	18,674	13,770	13,205	104.3	1.4
30-34	19,156	21,571	20,861	103.4	2.2
35-39	18,083	25,065	24,035	104.3	2.7
40-44	16,327	25,189	24,280	103.7	3.0
45-49	12,992	22,907	22,333	102.6	3.5
50-54	10,485	20,135	19,583	102.8	3.8
55-59	8,027	17,094	16,444	104.0	4.2
60-64	7,059	15,675	15,257	102.7	4.4
Total	**155,437**	**172,068**	**166,208**	**103.5**	

4.6 Demographic analysis based on two censuses and comparison of birth statistics

We briefly describe the two approaches, namely, simple extrapolations and population project.

4.6.1 Simple extrapolations

The extrapolation was based on the comparison of the enumerated population in July 1980 with the enumerated population in March 2003. Assumptions were made about fertility, mortality and migration levels. The growth rate was estimated at 1.34 per cent per year which over a period of 16 months would have resulted in a population of 490,000. Further, a simple extrapolation of the growth rate resulting from a comparison of the enumerated population in 1980 with the estimated Population in 2003 (thus enumerated population + omissions). This produced an annual growth rate of 1.46 per cent per year, which over a period of 16 months would have yielded a population of 504,0000. The census enumerated population was 492,829. Which fell in the range 490,000 and 504,000?

4.6.2 Inter-censal Survival Ratios

The enumerated population of x years at a census in year $t(Px, t)$ was compared with the enumerated population of $x+n$ years of age at census in year $t+n(Px+n, t+n)$ where $n>0$. this method however, approximately works well in closed population. In small populations the influence of migration is relatively significant. The table below gives the results for Suriname.

Table . 5 Inter-censal survival ratios 1980-2004 (24 years) by age and sex

Age group		Males			Females		
1980	**2004**	**1980**	**2004**	**Survival ration**	**1980**	**2004**	**Survival ratio**
0-4	24-28	23,487	19,553	0.8325	22,944	19,279	0.8403
5-9	29-33	24,926	20,090	0.8060	24,506	19,020	0.7761
10-14	34-38	25,043	19,170	0.7655	24,724	18,291	0.7398
15-19	39-43	23,712	18,722	0.7896	23,499	17,280	0.7354
20-24	44-48	15,990	13,616	0.8515	16,615	13,670	0.8228
25-29	49-53	11,160	10,694	0.9582	11,887	11,187	0.9411
30-34	54-58	8,556	7,353	0.8594	9,270	8,175	0.8819
35-39	59-63	7,449	6,430	0.8632	8,410	7,328	0.8713
40-44	54-68	7,265	5,347	0.7360	8,007	5,799	0.7242
45-49	69-73	7,006	4,502	0.6426	7,263	4,829	0.6649
50-54	74-78	5,857	2,730	0.4661	5,895	3,111	0.5277
55-59	79-83	4,640	1,512	0.3259	4,670	1,867	0.3998
60-64	84-88	3,128	520	0.1662	3,231	851	0.2634
65-69	89-93	2,704	216	0.0799	2,538	344	0.1355
70 +	94+	4,896	56	0.0118	5,962	124	0.0208
		175,819	**130,513**	**179,421**		**131,155**	

The 2006 PES report, posits that although Suriname population is far from closed, the survival ratios were plausible.

4.6.3 The Demographic balancing equation

The theory behind this approach is that the population at a certain time equals the base population plus natural growth plus net migration. Thus, $P_t = P_0 + (B_{0t} - D_{0t}) + (I_{0t} - E_{0t})$.

Where P_t is the population at the later date, p_0 is the population at an earlier date, B_{0t} is births and D_{0t} is deaths between the two dates and I_{0t} is immigration and $E_{0t} = emigration$

The above equation was applied on a relatively long period of 24 years between 1980 and 2004.

Table. 6 Estimated population using the Balancing equation

Census population(1980)	355,240
Natural growth (1July1980 and 2 Aug. 2004)	173,186
Net Migration (1July 1980 and 2 Aug.2004)	-62,859
Population growth	110,327
Population 2 Aug. 2004	492,829

Other methods used to evaluate the census included a comparison with birth statistics in vital registration; comparison of census infant with infant mortality from the Bureau for Public Health; and comparisons with other independent external data.

4.7 The 2004 post enumeration survey

It was recognized that despite every attempt to count every one belonging to the target population in a census, for various reasons it was not possible to enumerate every one. In Suriname in measuring the under-count they calculated the difference between the final census de *jure* count and True Population estimated using the Dual System of Estimation. The 2003 census displayed an undercount of 2.6 per cent.

4.7.1 Design and implementation of the PES

The PES was carried out in all the 10 districts of Suriname. About 116 enumeration districts (ED) were selected from a total of 844. Districts were treated as strata, with enumeration blocks. The blocks

were classified into as urban, rural, and interior. A proportionate probability sample (PPS) was selected. All households in the selected enumeration districts were enumerated.

Attempts were made to make the PES operationally independent from the 2003 census. For example, (i) the PES sample was only known to selected a number of the census team members, (ii) enumerators were assigned to different enumeration blocks from those they were worked in during the census; and (iii) the enumerators did not know the census results of the enumeration blocks they were assigned to.

The field work took place between 27 September to 11 October in 2004. It was undertaken seven weeks after the census date. And five weeks after the conclusion of the base population count. The target population for the PES was the non-institutional population. Another population subgroup excluded from the PES was the Garimpeiros in the gold fields. This group is very mobile. Enumerators in the PES worked in pairs while in the census, one enumerator covered an enumeration block. The PES questionnaire was simplified with fewer questions. While the census questionnaire had 47 questions the PES had only 16 questions. This may have enhanced cooperation from respondents which resulted into about 93 per cent overall response rate. For the interior areas the response rates were relatively low ranging from 84 to 86 per cent. According the PES results about 15,900 persons were missed during the 2003 census enumeration.

4. 8 The PES results

The following table gives estimates of undercounts by sex and age groups.

Table. 7 Under-count Numbers and rates by sex and age group

Subgroup	**Numbers**	**Under-count rate (%)**
Male	8,100	3.21
Female	7,800	3.03
Total	15,900	3.16
Age group		
0-14	-	3.72
15-59	-	2.54
60+	-	2.77

5. Uganda

Uganda carried out a post enumeration survey in January 2003 after the completion of the census in September 2002. The main objective of the PES was to measure coverage and content/response error of the census. Specific objectives included to:

(i) Evaluate the accuracy of the census by measuring coverage and content errors, at national, regional and urban and rural domains
(ii) Evaluate the quality of delineation of enumeration areas as viable sampling units as a basis for selecting samples for inter-censal household surveys
(iii) Use it as a basis for documenting lessons learnt to improve implementation of future censuses
(iv) Act as a source of information on sources and causes of errors
(v) Enhance skills in census evaluation in Uganda
(vi) Provide, if necessary, concrete statistical basis for adjusting census data.

5.1 Planning of the Post Enumeration Surveys

The post-enumeration survey was planned as part of the census programme. However, planning for the PES started late, in June 2002, with the initiation of a PES framework. The framework outlined the objectives of the exercise; survey strategies including methodologies to be adopted and it also outlined planned activities. In addition, the framework had a work plan, budget and a draft questionnaire. The survey strategy included details of a sampling design; data collection methods; procedures for matching census and PES records; reconciliation; data processing; and estimation of coverage and content errors.

The office of the Census Technical Officer, UBOS, implemented the PES, but drew expertise from the Institute of Statistics and Applied Economics as well as the UNFPA Country Support Team. The PES was implemented by the Technical Working Group chaired by the Deputy National Census Coordinator.

5.2 Data collection and analysis

The data collection and analysis activities included:

(i) *Planning*: Sample design and sample selection, data analysis and report preparations
(ii) *Field activities*: Collecting data through questionnaires from households and persons including during reconciliation visits
(iii) *Matching*: Office matching of census and PES household and persons' records
(iv) *Data processing, analysis and reporting*: Development of computer programmes, manual editing, data entry, tabulation, analysis and report writing.

5.3 PES sample design

The PES targeted all persons, in the country, who lived in private households. Persons in institutions, floating and those who were homeless were not included in the sample. A one stage

stratified cluster sample design was implemented in the selection process. The 2002 census enumeration maps and census household counts comprised the sampling frame. In order to maximize the precision of estimates, the administrative units within each sub-stratum were listed in a serpentine manner to facilitate the use of systematic sampling. The listing was done in three stages in:

(i) Districts within the strata
(ii) Sub-counties within districts
(iii) EAs within Parishes

5.4 Analytical Domains

The country was stratified into two main domains, namely, urban and rural. The urban stratum comprised 75 per cent gazetted of urban centres in the country. The rural major domain was split into four substrata, namely:

(i) Rural Central
(ii) Rural Eastern
(iii) Rural North and
(iv) Rural Western.

Estimates were calculated for each domain.

5.5 Sample size

The census EAs were the primary sampling units and the aim was to get reliable estimates for the strata. The following formula was applied to get a stratum sample size.

$n = \left(t_\alpha{}^2 pq\right) \big/ d^2$ (see section 5.3.2)

Where n = sample size

p = Universe proportion

$q = 1 - p$

d = Margin of error

t_α = T-statistics value for the 95% confidence interval (1.96)

The margin of error and confidence level was fixed at 0.03 percent and 95 per cent respectively and p was approximated from variables closely related to coverage. Other requirements included the need: (i) to maintain at the minimum of 30 EAs per stratum; (ii) to have at least two selections of PSUs per substratum; (iii) to reduce clustering. In total 350 EAs were selected for the PES. The Probability proportional to size (PPS) was used in distributing EAs among strata. The measures of size used during the PPS selection were the provisional census household count.

Table. 8 Percentage distribution of households and

Number of selected EAs

Stratum	% distr. of household	Number of EAs
Urban	14.1	52
Rural areas		
Central	21.9	69
Eastern	22.9	80
Northern	17.6	67
Western	23.4	82
All Areas	**100.0**	**350**

5.6 Questionnaire design and field instructional manual

The PES had two main instruments, namely, the questionnaire and field instruction manual. Control forms were also produced.

5.6.1 Questionnaire design

The initial draft of the questionnaire was designed by the Census Technical Office and it was revised by the PES technical working group and finally approved by the census technical and advisory committee. The questionnaire was designed to collect information relevant to the measurement of coverage and content error. The following items from the census questionnaire were included in the PES questionnaire:

(i) Full name
(ii) Relationship
(iii) Sex
(iv) Age
(v) Religion
(vi) Ethnicity
(vii) Marital status and
(viii) Existence of Agricultural holding

The questionnaire had five sections, namely:

(i) Identification particulars
(ii) Household matching particulars
(iii) Characteristics of household members
(iv) Characteristics of the out-movers
(v) Agricultural section

Information on agricultural holdings and ethnicity were, however, not used for matching

5.6.2 Enumerator's manual and maps

An enumerators' manual was developed alongside the questionnaire which contained instructions of how to enlist information from respondents. In addition, it contained concepts and definition of key terms. EA maps were reproduced and given to each enumerator. The maps were used for boundary identification and in guiding enumerators in covering selected EAs.

5.7 Fieldwork and logistics

Field work started after the preparations, which included preparation of EA maps, questionnaire design, and sample selection.

5.8 Staffing, recruitment and training

The Ugandan PES report indicated that high quality staff that had undergone adequate training was assigned to the PES. The organization of field work comprised the following cascading structure:

(i) Zonal supervisors were seconded from the Ugandan Bureau of Statistics (UBOS). These were responsible for training District Post Enumeration Officers (DPSO);
(ii) Assistant District Post Enumeration Officers (ADPSO) assisted DPSOs in cases where districts had nine or more EAs selected for the PES.

5.9 Logistics

The PES materials were distributed to enumerators during training and this enabled them to fill in the identification particulars. The PES field work, on average, took five days. However, in some large EAs the enumeration took longer. The reference date was the census night, thus all persons who spent a night, in a particular household, before the PES date was enumerated.

The enumerators also collected information about persons who were enumerated in a particular household during the 2002 census but were not members of the household during the PES reference night. These were the ***out-movers***. The supervision of PES field work was done at three levels, namely, national, zonal and district levels. At district level the DPESOs and ADPESOs supervised the enumerators. Each zonal supervisor oversaw at, least, three districts, while senior officers from UBOS were responsible for the national level supervision.

5.10 Recruitment and deployment of enumerators

The recruitment of enumerators was done by DPESOs. About 350 enumerators were recruited to cover 350 EAs in 55 districts. One enumerator was expected to cover an EA. However, EAs having more than

250 households were covered by more than one enumerator. In all, about 429 enumerators were recruited from the crop of better qualified 2002 census enumerators. While the enumerators were recruited from the parishes they belonged to, PES organizers made sure that they did not work in the same EAs as they worked in during the census.

5.11 Training

The training was aimed at familiarizing the participants on PES data collection procedures and reading of EA maps. The training e exercise started by training Trainers who were staff of the Ugandan Bureau of Statistics (UBOS). The training of DPESOs and ADPESOSs and enumerators was done at zonal level at six different training centres. The training was short which lasted only for two days. The training started with the debriefing of the DPESOs and ADPESOs on operational and administrative procedures, they were later trained jointly with enumerators.

5.12 Publicity

A standard message was put on radio informing people about the PES including dates of enumeration. The messages were in local languages, which informed people of the areas to be covered by the PES and what was expected of them. In order to entice maximum cooperation, communities in selected areas were mobilized prior to the PES enumeration.

5.13 Matching, data processing and reconciliation

A one week training session on matching procedures, for PES technical office staff was conducted by the CST/UNFPA expert in April 2003. These staff later trained matching and reconciliation clerks. A total of 30 matching clerks, three matching supervisors and four data entry clerks were recruited in September 2003. An additional six matching clerks and 10 data entry clerks were recruited in order to speed up the matching and data entry processes. Reconciliation enumerators were selected from the matching clerks taking into account the local language. They were briefed for one day about field reconciliation.

5.14 Matching

The purpose of matching was to determine whether a particular PES household/individual was enumerated in the census by comparing the individual's characteristics recorded in the PES and census. The matching clerks were divided into teams of three. Each team had a leader who was responsible for distributing work to individuals within a team. The team members first matched individuals within the matched households.

Specifically, the matching of households involved comparing names of administrative units, census household members and the names of household members. In matching households, team leaders were responsible for PES questionnaire booklets. Starting from the first listed PES household, the team leader read loudly the census number and names of household members. Households were declared as matching if the name(s) in the census questionnaire were the same to the name(s) in the PES questionnaire while tolerating minor spelling differences. The name of the head of household and/ or that of a spouse was adequate for deciding whether the household matched or not. Matching clerks

referred possible matches to supervisors who decided whether they were matches or non-matches. In matching households the following distinct categories were established:

(i) Matched households
(ii) Non-matches
(iii) Households created after the census (in-movers)

After matching households in a particular EA, each team member was assigned PES and census books to match individuals. The person's name and four characteristics were used to determine the matching status. The characteristics were relationship to head of household, age, sex and marital status. People over 10 years old and having at least three of the above characteristics the similar where considered matched. For people below the age of 11, relationship, age and sex were the variables considered in matching. If two of them were the same the individual was considered to be a match. Some varying age tolerances were taken into account in the matching process. The table below gives the tolerance limits.

Table. 9 Age tolerance limits used in matching persons

Age	Tolerance (in years)
< 10 years	± 1
10-19 years	± 2
20-40 years	± 3
> 40 years	± 4

Matching clerks then transcribed information from the census questionnaires to the PES questionnaires and assigned the matching and moving status codes for individuals who appeared in both questionnaires. The PES matching supervisors verified all the matched cases. The following were distinct categories assigned to individuals during the matching process:

(i) Match
(ii) Non-match
(iii) Born after census (out of scope)

5.15 Field Reconciliation visits

It was necessary to conduct reconciliation visits because during the process of office matching it was observed that a number of households and persons enumerated in the census did not match with cases enumerated in the PES. Similarly, a number of households and persons enumerated in the PES could not be matched with households and persons captured in the census. The purpose of the reconciliation, therefore, was to carry out field visits to identify through, re-interviews, ***erroneous enumerations*** which took place in the census. These were cases enumerated in the particular EA, during the census but reported by the PES as not staying in the EA and vice versa.

Specifically the reconciliation visits were meant to establish the status of:

(i) Households and/ or persons enumerated in the PES but not in the census
(ii) Households and/or persons enumerated in the census but not in the PES
(iii) Individuals who could not be matched even after applying flexible established matching rules

It should be noted that due to time and other resource constraints not all PES sample EAs were covered in the reconciliation visits. While all the four regions were covered, in selecting the PES sample some areas within the regions were deliberately excluded. The exclusions included:

(i) Districts which were experiencing insecurity and had mobile populations
(ii) Areas with relatively high match rates
(iii) EAs where it was not possible to retrieve all corresponding census questionnaires.

A total of 105 EAs were selected for reconciliation.

5.16 Data processing

Data capturing was carried out by using the Census and Survey Processing (Cspro) software. Data verification was done at 100 per cent rate throughout the exercise. Cspro was also used to generate initial tabulations. Such tabulations were then exported from Cspro to Stata and Microsoft Excel to facilitate the production of final tables.

5.17 Estimates of coverage error

The table below shows some selected national coverage measures

Table.10 other estimates on coverage error

Domain	Coverage rate	Erroneous inclusion rate	Gross coverage error rate
National	94.3	3.7	9.3

Age group			
0-4	94.4	4.0	9.5
5-9	96.1	3.4	7.2
10-19	94.8	3.9	9.0
20-39	92.2	4.2	11.9
40 and above	95.4	2.6	7.2
Gender			
Female	94.5	3.7	9.1
Male	94.2	3.7	9.5
Rural & Urban			
Urban	88.2	7.7	19.1
Rural	95.0	3.3	8.2
Rural substrata			
Central	93.9	3.7	9.7
Eastern	95.6	2.5	6.9
Northern	93.7	5.2	11.3
Western	96.1	2.4	6.2

It will be observed that the above table does not suggest any significant difference of coverage rates between males and females. The urban coverage rate, however, was lower than that of the rural domain 88.2 and 95 per cent, respectively.

5.18 Measurement of content error

Content error was interpreted as errors in recording characteristics of persons who were enumerated in both the census and PES. Five characteristics were evaluated, namely:

(i) Sex
(ii) Age
(iii) Relationship
(iv) Marital status and
(v) Religion

Five measures were used in the analysis of content error. These were: (i) Rate of agreement; (ii) Net difference rate (ii) Index of inconsistency; (IV) Aggregate index of inconsistency and (v) Gross difference rate. The table below gives net difference rates and indices of inconsistency by characteristics.

Table. 11 Net difference rates and indices of inconsistency by characteristics

	No. of cases in PES	**No. of cases in census**	**Net difference rate**	**Index of Inconsistency**
Gender				
Male	105,867	106, 050	-0.09	3.97
Female	107,223	107,040	0.09	3.97
Relationship				
Head of household	45, 052	44,978	0.03	7.9
Spouse	32,455	32,407	0.02	8.9
Child	108,951	109,563	-0.29	11.1
Step child	3,922	4,011	-0.07	82.1
Parent of head or spouse	1, 037	1, 080	-0.02	48.8
Brother/sister of head or spouse	3, 683	4,011	-0.15	54.2
Other relative	16,213	15,109	0.52	44.1
Non-relatives	1,411	1,507	-0.05	53.8
Marital status				
Never married	5, 382	5,324	-1.88	22.2

Currently married/cohabiting (monogamous)	5,135	5,363	0.73	17.5
Currently married/cohabiting (polygamous)	610	418	-0.49	45.5
Widowed	529	534	-0.29	30.4
Divorced/separated	708	552	-0.39	52.7
Not applicable	5,063	5,236	2.32	13.8
Religion				
Catholic	90,288	90,620	-0.11	14.9
Anglican	78,840	77,501	0.63	18.0
SDA	2,289	2,564	-0.13	32.4
Pentecostal	9,634	10,252	-0.29	39.8
Moslem	24,013	24,168	-0.07	5.0
Others	6,841	6,900	-0.03	102.1
Age groups				
0-4	35,514	38,964	1.63	11.93
5-9	35,821	35,017	-0.38	24.10
10-19	53,987	51,892	-0.99	15.53
20-39	54,402	53,962	-0.21	11.63
40 & above	32,500	32,389	-0.05	8.37

5.19 Sampling errors

Sampling errors are random errors in survey estimates resulting from a fact that a sample rather than the entire enumeration of a population. It should be noted that a sampling error is the square root of a sampling variance. Sampling errors were computed for the following estimates:

(i) PES population
(ii) The Match rate for out-movers
(iii) Coverage rate and
(iv) Omission rate

Table. 12 Reliability of estimates based on selected indicators at national level

Variable	Estimate	Standard error	Confidence interval		Design effect
			Lower	Upper	
PES population	24,731,997	597,879	23,600,000	25,900,000	0.627
Match rate for out-movers	0.922	0.010	0.903	0.941	1.193
Coverage rate	0.978	0.004	0.971	0.986	1.432
Omission rate	0.022	0.004	0.014	0.029	1.432

5.20 Lessons learnt from implementing the PES

(i) The planning of the PES was deferred until after June 2001, although it was advisable to plan for the PES as part of the census programme for strategic reasons mainly pertaining to resource allotment. The end results was that the implementation of the PES was delayed resulting into field work being conducted after four months after the census enumeration instead of the envisaged three months. Because of the delay in planning for the PES field work a pilot PES was not conducted.

(ii) Owing to financial constraints the same unit thus, Census Technical Office, and field personnel planned and conducted both the census and the PES. This

arrangement, somewhat, compromised the fundamental of maintaining of independence between the census and the PES. This important requirement for a Dual System of Estimation. For the future, there is need to adhere to the independence philosophy as it pertains to its practical and operational ramifications.

(iii) At different stages of the PES, there was shortage of funds to pay for undertaking some activities. This led to delays in the implementation of some activities such as filed work, matching, and field reconciliation. Owing to the lack of sufficient financial resources, not all EAs with unmatched cases were covered. A sub sample was, therefore, selected. It is advisable to plan for resources well in advance of the actual implementation of the PES.

(iv) Due to the delay in executing the PES there was considerable migration in selected areas. It was difficult to capture some of the migratory population. This shortcoming impacted on the coverage rates and omission errors.

Notwithstanding the above mentioned shortcomings, the PES had its advantage in that staff of UBOS who were involved in the PES gained experience in planning and implementing a PES. It is argued, in the UBOS report, that the experience was invaluable for the country to organize the next PES.

References

Australian Bureau of Statistics (2006). Census of Population and Housing –Undercount, *www.abs.gov.au*.

Economic Commission for Europe (2008). Preparation, organization and conducting of post-enumeration survey in the State Statistical Office of the Republic of Macedonia, Joint UNCE/Eurostat Meeting on Population and Housing Censuses, Eleventh Meeting, Geneva.

Statistics South Africa (2004). Census : Post-enumeration survey: Results and methodology, Published by Statistics South Africa, Private Bag X44, Pretoria 0001.

General Bureau of Statistics Census Office (2006). Report on Census 2004 Coverage Evaluation, Suriname in cijfers n.230-2006/12.

Uganda Bureau of Statistics (2005). Post Enumeration Survey: 2002 Uganda Population and Housing Census, Entebbe, Uganda.

United Nations (October 1952, Population Bulletin no.2): Accuracy Tests for Census Age distributions tabulated in five-year at ten-year groups.

Annex 2. Sample PES questionnaire

FORM NO. P E S 2 0 0 0 1

Strictly Confidential

CENTRAL STATISTICAL OFFICE
P. O. BOX 31908, LUSAKA

Republic of Zambia
POST-ENUMERATION SURVEY
2000 CENSUS OF POPULATION AND HOUSING

QUESTIONNAIRE IDENTIFICATION	CENSUS STICKER INFORMATION (If available)	ASSIGMENT RECORD
1. PROVINCE NAME ☐	14. CENSUS BUILDING NUMBER ☐☐☐	NAME / DATE (MONTH, DAY)
2. DISTRICT NAME ☐☐	15. HOUSING UNIT NUMBER (HUN) ☐☐	Enumerator
3. CONSTITUENCY ☐☐	16. HOUSEHOLD NUMBER (HHN) ☐	Supervisor
4. WARD ☐	17. CENSUS DATE D D M M / / / /	Office Editor
5. CSA NUMBER ☐☐	**OCCUPANCY STATUS**	Matching Clerk
6. REGION 1. RURAL 2. URBAN ☐	18. **(OBSERVE)** Is this housing unit occupied? 1. Yes –**GO TO 21** 2. No ☐	Field Revisit (Enumerator)
7. SEA NUMBER ☐	19. **(OBSERVE)** What is the condition of this vacant unit? 1. Habitable for the year round 2. Habitable for seasonal use 3. Abandoned ☐	Data Entry Operator
AREA IDENTIFICATION	20. **(ASK NEIGHBOUR)** Was this housing unit Occupied at the time of 2000 census? 1. Yes – **GO TO OUT-MOVERS SECTION** 2. No – **END INTERVIEW** ☐	Data Entry Verifier
8. VILLAGE/LOCALITY	**HOUSEHOLD**	**SUMMARY COUNT**
9. CHIEF'S AREA	21. Was this household enumerated during the 2000 census? 1. Yes 2. No. **WRITE COMMENTS BELOW** ☐	MALE
10. RESIDENTIAL ADDRESS/VILLAGE NAME		FEMALE
PES STICKER INFORMATION		TOTAL
11. PES BUILDING NUMBER ☐☐☐		**SUPPLEMENTARY? INDICATE WHETHER OR NOT CONTINUATION SHEET WAS USED**
12. HOUSING UNIT NUMBER (HUN) ☐☐	**TYPE OF HOUSING UNIT**	
13. HOUSEHOLD NUMBER (HHN) ☐	22. (OBSERVE) What type of housing unit is this? 01. Traditional 02. Mixed 03. Conventional Flat 04. Conventional House 05. Mobile 06. Part of Commercial Building 07. Improved/Makeshift 08. Unintended 09. Institutional/Collective 10. Other ☐☐ ☐☐	

STATUS OF QUESTIONNAIRE

23a. Interview Status	1. Interview Completed (Occupied) 2. Non-contact (Occupied) 3. Not Interviewed (Vacant) 4. Non-residential ☐
23b. Household Match Status. **(FOR OFFICE USE ONLY)**	1. Match 1 to 1 2. Match 1 to 2 3. Match 2 or more to 1 4. Non-match 8. Not Applicable ☐

1

FOR ALL PERSONS - GENERAL CHARACTERISTICS								
SERIAL NUMBER	P- 1: HOUSEHOLD MEMBERS What are the names of all usual household members?	P - 2: RELATIONSHIP What is (name)'s relationship to the head of household? 01. Head 02. Spouse 03. Own Son/Daughter 04. Step Son/Daughter 05. Parent 06. Brother/Sister 07. Nephew/Niece 08. Son/Daughter In-Law 09. Grandchild 10. Other relative 11. Unrelated	P - 3: SEX Is (name) Male or Female? 1. Male 2. Female	P - 4: AGE How old is (name)? ENTER AGE IN COMPLETE YEARS ("00" if less than 1 year and "90" if 90 years or older)	P - 5: MARITAL STATUS (ASK PERSONS 12 YEARS AND OLDER ONLY) Is (name) 1. Married 2. Separated 3. Divorced 4. Widowed 5. Never Married 6. Living Together/ Cohabiting	P - 6: RESIDENCE ON CENSUS DAY Was (name) a usual resident of this housing unit at census time? 1. Yes 2. No- Because was a usual resident after census day 3. No- Born after the census day	P - 7: MOVING STATUS (FOR OFFICE USE) 1. Non Movers 2. In-Movers 3. Not Applicable	P - 8: MATCH STATUS (FOR OFFICE USE) 1. Match 2. Possible Match 3. Non-match 8. Not Applicable
01								
02								
03								
04								
05								
06								
07								
08								
09								
10								
11								
12								
13								
14								

2

FOR PERSONS WHO HAVE MOVED OR DIED

O - 1: OUT-MOVERS

Are there any persons who were usually members of the household at census time who are no longer usual members of this household (including those who died after census day)?

1. Yes – LIST THE NAME(S) OF ALL WHO HAVE MOVED SINCE CENSUS DAY AS OUT-MOVERS ☐
2. No – END INTERVIEW

SERIAL NUMBER	LIST OF NAMES OF OUT-MOVERS	O - 2: RELATIONSHIP What is (name)'s relationship to the head of household? 01. Head 02. Spouse 03. Own Son/Daughter 04. Step Son/Daughter 05. Parent 06. Brother/Sister 07. Nephew/Niece 08. Son/Daughter In-Law 09. Grandchild 10. Other relative 11. Unrelated	O - 3: SEX Is (name) Male or Female? 1. Male 2. Female	O - 4: AGE How old is (name)? ENTER AGE IN COMPLETE YEARS ("00" if less than 1 year and "90" if 90 years or older)	O – 5: MARITAL STATUS (ASK PERSONS 12 YEARS AND OLDER ONLY) Is (name) 1. Married 2. Separated 3. Divorced 4. Widowed 5. Never Married 6. Living Together/ Cohabiting	O - 6: MATCH STATUS 1. Match 2. Possible Match 3. Non-match 8. Not Applicable
01		☐☐	☐	☐☐	☐	☐
02		☐☐	☐	☐☐	☐	☐
03		☐☐	☐	☐☐	☐	☐
04		☐☐	☐	☐☐	☐	☐
05		☐☐	☐	☐☐	☐	☐
06		☐☐	☐	☐☐	☐	☐
07		☐☐	☐	☐☐	☐	☐
08		☐☐	☐	☐☐	☐	☐
09		☐☐	☐	☐☐	☐	☐
10		☐☐	☐	☐☐	☐	☐
11		☐☐	☐	☐☐	☐	☐
12		☐☐	☐	☐☐	☐	☐

3

FOR OFFICE USE AND REVISIT ONLY

PERSONS ENUMERATED IN THE CENSUS BUT NOT IN THE POST ENUMERATION SURVEY (NON-MATCHES)

SERIAL NUMBER	R- 1: NAMES INDICATE NAMES OF USUAL MEMBERS OF THIS HOUSEHOLD ENUMERATED IN THE CENSUS BUT NOT IN THE POST ENUMERATION SURVEY	R - 2: RELATIONSHIP What is (name)'s relationship to the head of household? 01. Head 02. Spouse 03. Own Son/ Daughter 04. Step Son/ Daughter 05. Parent 06. Brother/Sister 07. Nephew/Niece 08. Son/Daughter In-Law 09. Grandchild 10. Other relative 11. Unrelated	R – 3: SEX Is (name) Male or Female? 1. Male 2. Female	R - 4: AGE How old is (name)? ENTER AGE IN COMPLETE YEARS ("00" if less than 1 year and "90" if 90 years or older)	R - 5: MARITAL STATUS (ASK PERSONS 12 YEARS AND OLDER ONLY) Is (name) 1. Married 2. Separated 3. Divorced 4. Widowed 5. Never Married 6. Living Together/ Cohabiting	R - 6: MEMBER SHIP STATUS (PAST) Was (name) a usual resident of this housing unit at census time? 1. Yes 2. No	R - 7: MEMBER-SHIP STATUS (PRESENT) Is (name) still usual member of this household? 1. Yes 2. No	R - 8: MATCH STATUS 1. Match 2. Possible Match 3. Non-match 8. Not Applicable	R – 9: CENSUS ENUMERATION STATUS 1. Correct Enumeration 2. Erroneous Enumeration / Fabrication 3. Erroneous Enumeration / Duplication 8. Not Applicable
01									
02									
03									
04									
05									
06									
07									
08									
09									
10									