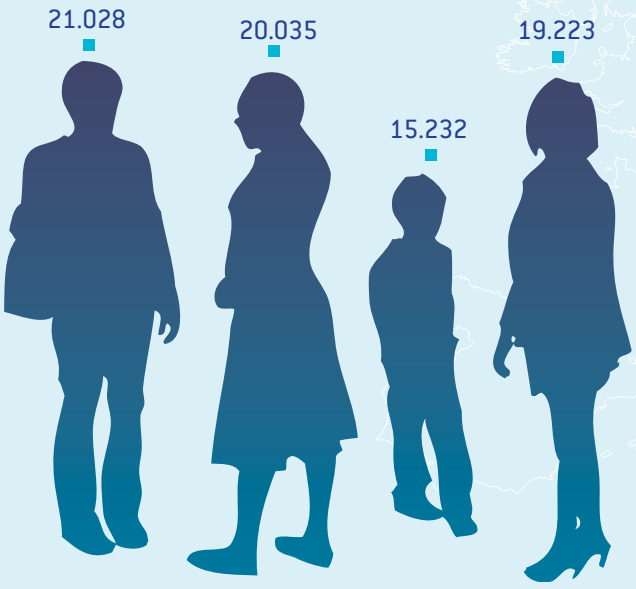
ask
AGJENCIA E STATISTIKAVE TË KOSOVËS
AGENCIJA ZA STATISTIKE KOSOVA
KOSOVO AGENCY OF STATISTICS

REKOS 2011

Kosovo Population and Housing Census 2011

FINAL RESULTS

EVALUATION OF CENSUS RESULTS WITH THE POST ENUMERATION SURVEY



Census Project Multi-Donor Trust Fund

UKaid from the Department for International Development
SWEDEN
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Agency for Development and Cooperation SDC
LE GOUVERNEMENT
DU GRAND-DUCHÉ DE LUXEMBOURG
Coopération luxembourgeoise
MINISTRY OF FOREIGN AFFAIRS OF DENMARK

NOTE TO THE READER

The population and household census was carried out by door to door enumeration in April 2011, with as reference date the 31 March 2011, at midnight. It was based on a specific Law on the Census (03/L–237) approved on 7 October 2010 by the Kosovo Assembly. One of the most important provisions of the law regards data protection and confidentiality of personal information. This law also defines the obligation to conduct a census post enumeration survey immediately following the census enumeration.

Kosovo census is compliant with international methodology recommendations for the 2010 censuses of population and housing prepared by United Nations Economic Commission for Europe in cooperation with the Statistical Office of European Communities (Eurostat). All topics covered and data collected refer to internationally agreed definitions.

All the data reported in this document refer to the 2011 Census. The Census results include data from 34 municipalities. Due to objective reasons the enumeration could not be carried out in the northern municipalities, which has been recognized by the Census Trust Fund Steering Committee. The Kosovo Agency of Statistics will publish estimates for these municipalities later on.

Kosovo Population and Housing Census 2011

FINAL RESULTS

EVALUATION OF CENSUS RESULTS WITH THE POST ENUMERATION SURVEY

Technical assistance to the Kosovo Agency of Statistics for the preparation of a Population Census

An EU-funded project managed by the European Union Office in Kosovo implemented by Istat, CIRPS-Sapienza University of Rome, ICON-INSTITUT and Rrota

EUCEP 2011 CENSUS SUPPORT PROJECT

Table of Content

List of Tables

List of abbreviations

AII	Aggregate Index of Inconsistency
ASK	Kosovo Agency of Statistics
CEO	Chief Executive Officer
CI	Confidence Interval
CSProX	Integrated system for statistics data capture and analysis
DSE	Dual System Estimation methodology
EA	Enumeration Area
EU	European Union
GDR	Gross Difference Rate
GIS	Geographic Information System
ID	Identification Code
IMO	International Monitoring Operation
IT	Information Technology
MySQL	Database Management System
NDR	Net Difference Rate
PES	Post Enumeration Survey
RA	Rate of Agreement
SQL	Structured Query Language

Foreword

In my capacity of Chief Executive Officer of the Kosovo Agency of Statistics, I am pleased to provide some introductory thoughts to this document, which arrives at an important turning point in the history of Kosovo official statistics.

When, in 2004, a strategic planning process of the population and housing census was set in motion, the process of establishing our institutional priorities and enabling us to make wise resource-allocation choices in the months and years to come was at its starting point. The work in front of us was enormous, and from the very beginning, our main objectives was to ensuring the highest possible quality to our census data and guarantee that census results would be widely recognized and used with confidence by the users.

This is a challenging goal in a context of a country lacking reliable census data for 30 years, in which the entire population called to participate in the operation by providing answers to a large set of questions does not fully understand the scope and crucial need of the operation.

It appeared quite soon that Kosovo would need support in all census planning and implementation aspects, including public awareness, and I take this opportunity to express my highest appreciation to all institutions, bodies and individuals having provided political, financial and technical support to us in this titanic undertaking. My specific gratefulness goes to the census International Monitoring Operation (IMO) members, to all Donors – especially the EU and some of its member States - having generously shared the census costs with Kosovo, and to all foreign institutions and bodies having provided us with experts' assistance in all technical issues.

In all this assistance received, our institution grew considerably: we understood that collecting and disseminating data was not enough, we realized that we have to assess the quality of our data, and moreover, that we have to share our conclusions on this quality with our users. I personally consider this shift towards full transparency about our work as one of the greatest successes of our institutional building process.

Statistics is a science, and it has throughout the years developed theories and models also to assess the data quality. One of these instruments is specific to census coverage and quality assessment, and is known as post enumeration survey (PES).

PES relies on totally specific methodology and requires sound statistical knowledge as well as experience. We were lacking both. Yet, with the help of our EU-funded technical assistance project, and with continuous guidelines received from our IMO Steering Committee partners, ASK implemented this survey and was able to estimate the census coverage errors as well as to assess the census content quality. The results of this whole process are presented in this report. We have taken also as much attention as possible to describe the operation in a transparent way, describing weaknesses and limitations PES has faced and trying to provide our users with clarifications on how to interpret PES results.

With this document, and with the institutional commitments that lie behind it, I and my staff hope having met your expectations for clear and visible information on the quality of the Kosovo census 2011. A census quality report is complementing this particular PES exercise to ensure a full evaluation of the data quality. We look forward to your comments and remain open to your suggestions in view of better servicing your needs.

Isa Krasniqi
Chief Executive Officer

Executive Summary

- The quality of population and housing census data is crucial, among others for building public trust and understanding in official statistics. 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 results.

- 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.

- Numerous methods are available to estimate the coverage and content error of censuses. Among these post enumeration survey is a very important and specific method for evaluating census data.

- For the first time in its census history, Kosovo implemented a PES. This report defines the PES implemented 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; and as main PES outcome, the evaluation of coverage and content error. The conclusion highlights the usefulness of the PES and care that must be taken in its results' interpretation.
- The sample size for PES was defined by law and limited to 0.5% of the total population. The unique reference sampling frame available to extract the sample was spatial based and consisted in the last version of the total census Enumeration Areas, which amounted 4,681 for the whole Kosovo territory. Among 23 retained EAs in the sample, only 20 could be used for the PES, the other three ones pertaining to the small Kosovo area not included in the enumeration process. The smallness of sample size is considered the larger limitation of the Kosovo PES and restricted its evaluation to the national level.
- The PES questionnaire includes questions and other information that serve four purposes:
    - Defining for each re-interviewed individual if she/he belongs to the census target population;
    - Defining census address of each individual;
    - Enabling later on a successful record linkage at the address level and at the individual level of PES records with the census records;
    - Allowing the content errors estimates for certain characteristics of the individuals.

  The information gathered by the PES questionnaire is used for estimation of local

(region) and overall (Kosovo-enumerated) coverage, as well as for quality checks of selected characteristics collected in the census.

- Kosovo PES fully complied with the principle of independence from the census process. Estimations have been based on the Dual System Estimation (DSE) methodology which is based on capture and recapture methodology. The methodology estimates the total population. The DSE model is conceptualized that each person has a probability of being either included in the census or not included in the census as well as either included or not included in PES. Chapter 9 presents all coverage and content results estimates, while a specific appendix describes the used methodology.

- Among all coverage error rate indicators, the census omissions has the highest level (4.3% at national level) while erroneous inclusions rate is the lowest (2%), especially in urban areas (0.99%). The net under-count rates were 2.30% at national level; 3.83% at urban level, and 1.43% at rural levels. Urban areas had thus a slightly larger under-count than rural ones. The erroneous inclusion rates were 2.0% at national level, 0.99% in urban areas and 2.5% in rural areas. These results are at the level usually encountered in good quality censuses.
- The content error analysis was performed for age, sex, marital status, and ethnicity using the matched persons in P-and E- samples. The content error in the census and PES was exceptionally low for all demographic characteristics except certain categories of marital status. Aggregate inconsistencies indices for the evaluated characteristics ranged between 3.3 and 5.4 except for marital status which was 13.5. The standard threshold to assess errors as being "low" is when this index is lower than 20. In Kosovo case, the content errors estimated can thus be declared as "very low".
- Considering that it was the first census in 30 years, the results of census from PES evaluation look very good. The data quality in terms of completed data is also exceptionally good for those who responded in PES or census.

- The match rates and correct enumeration rates are over 90 percent and in some cases 100 percent.

- PES also provided good outcomes from the reconciliation visits since all statuses needed from matching operations were confirmed except for 53 out of over 6,000 persons in P-sample (less than 0.8%).

- The whole PES undertaking and its results lead to practical warnings and recommendations, which are presented in the conclusions of this report.

Introduction

Kosovo conducted its first census in 30 years. Implementing a census is a massive undertaking and requires major planning and resources.

A census of population, households and dwellings is often the largest data gathering exercise in any country. In principle, it requires counting everybody in the country on census night.

Regrettably despite all efforts to include northern Kosovo in the census operation, enumeration in this area could not take place, as the necessary co-operation of the population in the North could not be guaranteed. Consequently, census in this report refers only to the area of Kosovo where the census could be implemented.

Despite this fact, Kosovo census yields a wealth of valuable information for analyzing changes in the socio-demographic profile of the population, and for monitoring, planning, and decision making at the national and local level, by government, business and the general community.

It is also integral to the derivation of reliable post-censual population estimates and for charting future demographic trends. Given the strategic significance of the census data and its diverse applications, Kosovo Agency of Statistics (ASK in the following), like other national statistical organizations, made concerted efforts to ensure universal coverage of its 2011 census.

Yet, censuses everywhere tend to miss some people. It is almost impossible to conduct a perfect census since errors can be introduced at many different points in the data collection and processing operations. Therefore, it is important to assess the census quality including census errors. Incomplete coverage may result from, for example, inadvertent omission of young children, difficulty in enumerating people on the move and those living in apartments, as well as people not willing to cooperate with census enumerators.

There are a myriad of statistical procedures that demographers and others use to check the accuracy of census coverage.

These include: (a) checks against demographically derived estimates, (b) comparison of census figures with administrative records and other sources, and (c) a post-enumeration survey (PES).

A PES is undertaken shortly after the census to evaluate the completeness of census coverage. It involves an independent re-enumeration of a statistically designed sample of all dwellings and the people within them covered by the national census. The basis of the methodology lies on the comparison of the re-interviewed

persons with the ones enumerated during the census.

Kosovo Agency of Statistics in the past (1981) has evaluated certain aspects of the general quality of census data. However, it has never attempted to measure the level of undercount or over count directly, using for example a post enumeration survey. In 1981 a pilot test as a preliminary to a PES was conducted (census control). This was an attempt to control the census quality but was not in line with PES methodology standards.

The Census law approved in Kosovo[1] includes the implementation of a PES immediately following the census enumeration, for a period of one week and on a sample of 0.5% of the population. The 2011 PES was the first to be undertaken in the history of Kosovo censuses.

The main objective of the PES was to measure the level of national coverage (undercount and over count) as well as the content errors in the Kosovo 2011 Census. This report describes and discusses the salient features of the 2011 PES, including its scope, methodology, the information gathered and the results of this operation.

CHAPTER 1

Miscount and its Sources

In such a large and complex exercise as a census, it is inevitable that some people are missed and some counted more than once.

Reasons for people and dwellings being missed are many and vary according to the type of persons and of situations; it includes:

- Dwellings entirely missed by enumerators
- People deliberately avoiding the census – refusing or unwilling to respond (ex. for fear that information given will be used against their interests when they mistrust the census confidentiality on personal data)
- People being reluctant to open their door to strangers
- People shifting from one house to another around the time of the census
- Multiple households living at the same address
- People being away temporarily (ex. work, school)
- People having no usual residence (ex. transients, street kids)
- The address is wrongly registered during the enumeration process
- Newborn babies being overlooked.

Conversely, there are situations in which people can be over counted:

- students living away at school or university (and also being counted at the home of their parents)
- children under joint custody
- people living away from home while working
- people shifting from one house to another around the time of the census
- people living in institutions
- erroneous enumeration of deceased persons, babies born after census night, residents
- temporarily overseas on census night, emigrants, etc

Coverage and content errors

This report discusses the coverage errors and the content errors in the census that were measured with the PES. Coverage errors include omissions and erroneous inclusions in census counts.

Omission occurs when a person should have been included in the census but was not included. This could happen due to misunderstanding of the question or concept, curb-stoning, intentionally not reporting a person living in a household for any particular reason, etc., or person is reported in a wrong place.

Erroneous inclusion occurs when a person is included but should not have been included. This could happen due to misunderstanding of the concept and includes duplicates, fictitious persons and reporting a person in a wrong place.

A content error occurs when person or households characteristics are reported or coded in error. This could happen due to respondent not having the correct information, could not recall the correct answer, or errors introduced during coding or processing operations.

It is worth to recall that the census had to count the usually resident population, as defined by international standard methodology. Census population corresponds to all persons who usually resided in Kosovo for at least 12 months at the census date, or had the intention to reside in Kosovo for at least 12 months at that date; persons with diplomatic status, foreign military personnel and persons who have their usual residence out of Kosovo are excluded.

Census coverage evaluation methods

A number of methods are used to evaluate censuses. These include demographic analysis, comparing against alternative data sources such administrative data or against results from a large demographic surveys, and evaluation of quality check results. ASK carried out qualitative analysis based on this methodology at aggregated data level and users can find its results in the census quality report. On its side, PES could help evaluate coverage errors for detailed demographics at micro-data level such as by urban/rural, age, ethnicity, sex, etc. Therefore, ASK also used a Post-evaluation Survey to measure coverage errors in addition to alternative evaluation approaches it has used.

CHAPTER 2

PES Objectives

The Post Enumeration Survey (PES) addresses three main issues:

- Were all the people who belong to the census target-population enumerated?
- Were only the people who belong to the census target-population enumerated?
- Is the reported address of each person who was enumerated, the right census address?

The PES in Kosovo has four specific objectives:

- Quantitatively evaluate the accuracy of the census collected data in terms of coverage and content error, at: national, regional and urban/rural areas, with specific emphasis in coverage of ethnic minorities. The latter was considered important in the Kosovo context, to measure whether all communities had a fair coverage in the census.
- Provide quantitative information required for determining the success of the 2011 Kosovo's population and housing census, and enhance its credibility.
- Furnish information on possible sources and causes of errors.
- Serve as a basis for improving the implementation and preparation of future censuses.

The 2011 PES is a sample survey of individuals in private dwellings. The 2011 PES has two sets of targets – evaluation of coverage, and content errors. These are defined below.

Coverage errors

The objectives of the Kosovo PES are to provide the following coverage and their analysis for the entire nation and for each region by urban and rural type of area:

- Net coverage errors and net coverage error rates in census;
- Number and rate of omissions in census;
- Number and rates of erroneous in census;
- Number and rates for duplicates in census;
- Variances and standard errors for each of the above estimates;
- In addition, the objectives of the Kosovo PES are to provide some measures of coverage errors for each region by urban/rural by age by ethnicity.

Content errors

The objectives for content errors are to evaluate the following measures for the four demographic groups namely age, sex, marital status, and the ethnicity at the national level:

- Net difference;
- Index of inconsistencies;
- Aggregate index of inconsistencies;
- Gross difference; and
- Rate of agreement.

CHAPTER 3

PES Questionnaire

The PES questionnaire (see annex 1) includes questions and other information that serve four purposes:

- Defining for each re-interviewed individual if she/he belongs to the census target population;
- Defining census address of each individual;
- Enabling later on a successful record linkage at the address level and at the individual level of PES records with the census records;
- Allowing the content errors estimates for certain characteristics of the individuals.

The information gathered by the PES questionnaire is used for estimation of local (region) and overall (Kosovo-enumerated) coverage, as well as for quality checks of selected characteristics collected in the census.

The questionnaire is divided into four parts:

- Identification of the dwelling's geographical location, and of all questionnaires belonging to the same dwelling unit;
- Classification of the use of the building and of the dwelling unit on Census Day and at PES time;
- Listing all the people who reside or resided in this dwelling unit;
- Individual questionnaire which aims to classify each person vis-à-vis her/his belonging to the Census Population and to identify her/his Census Address, and to facilitate the record linkage with the census data (after the PES is over).

Geographic location

POPULATION AND HOUSING CENSUS
POST ENUMERATION SURVEY

Census Law No. 03/L-237 " Kosovo Official Gazette", no.84/2010
All the data in this questionnaire are protected by law and will be used only for statistical purposes.

Municipality	Settlement	EA	-	Building Code	Entrance No.	Dwelling No.

This part of the questionnaire is essential to ensure that the record linkage between PES re-interviewed persons and Census enumerated persons correspond to the same location. Location codes used in PES questionnaire as well as enumeration area maps used for data collection have been the same than the ones used in the census, in order to optimize this aspect. Since in Kosovo there is no system of street, building and dwellings fixed addresses, likewise in the census, a PES "address" is described by:

- The code of the municipality (2 digits, pre-defined code);
- The code of the settlement (4 digits pre-defined code);
- The code of the enumeration area-EA (3 digits pre-defined by the census geography maps)
- The sub-code of the EA (one digit: for increasing the PES data collection quality, some EAs have been divided between two interviewers; the sub-code was indicated as "a" or "b");
- The building code (3 digits pre-defined on the EA maps; if a building was NOT on the EA map, a serial NEW code was entered by the interviewer as by instructions received);
- The entrance code (2 digits, serial number entered by the interviewer as by instructions received);
- The dwelling code (3 digits, serial number entered by the interviewer as by instructions received).

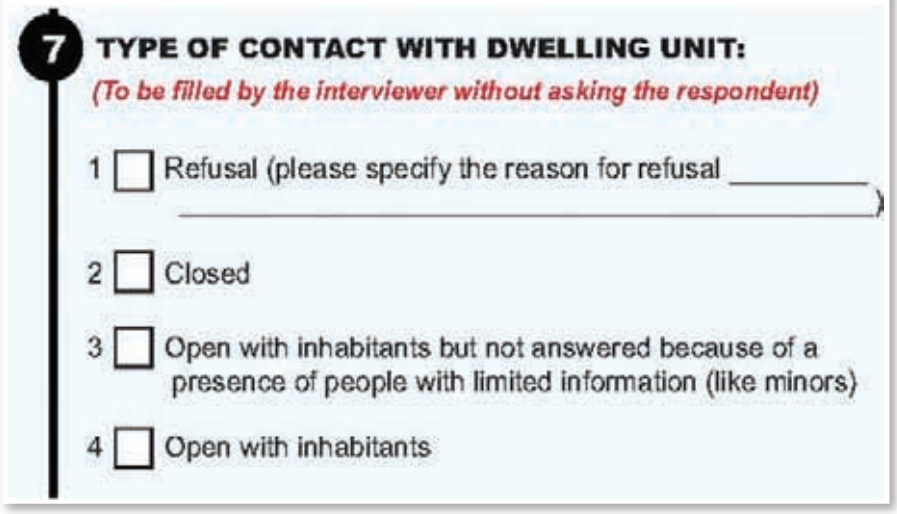
In order to facilitate the management of PES questionnaires, each of them received a pre-defined serial number. In addition, the possibility of a dwelling having more than 13 and even 26 persons was foreseen with the codes of "continuation questionnaires", linking questionnaires between them in case one form (limited to gather data for 13 individuals) would not be sufficient.

Classification of the use of the building and of the dwelling unit on Census Day and at PES time

Questions Q1 to Q6 aim at verifying the changes in the status and usage of the building and housing unit that could occur between the census day and the PES operation.

Both PES and the census are interested only in people usually resident in private residential housing units. The census also collected data for individuals residing permanently in collective living quarters (institutions) but these persons were not included in PES coverage estimates.

Description of the reason of not having conducted an interview



The PES questionnaire also contains a specific question on the reason for data not being collected, which contrary to the census questionnaire includes "refusal of the inhabitants" as an option. This question is very important since in some areas, a certain part of the people refused to take part in the PES operation.
A "closed" dwelling corresponds to a place designed for habitation where people could not be found despite repeated visits by the interviewer.

Only PES questionnaires with code 4 "Open with inhabitants" could be used for the PES estimates of coverage.

Listing all the people who reside or resided in this dwelling unit

The second and third pages of the questionnaire are aimed at defining the "status" of each individual living in the dwelling at the moment of the census (tables Q9; Q12 and Q15).

Table Q9:

LIST OF PEOPLE RESIDING IN THIS DWELLING UNIT TODAY:

Seq. no.	First name	Last name	Father's name	Did you have a different address on the midnight between March 31st and April 1st, 2011?		Did you have two addresses on the midnight between March 31st and April 1st, 2011?	
01				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
02				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No

The first table above (Q9) is used to define whether a person should have been counted in the census or not at this address. Each person who responded not having changed address between the census day and PES interview is a candidate for checking if she/he have been enumerated properly.

For people having moved or people having two addresses, further investigations are necessary to check the enumeration status.

Table Q12:

LIST OF PEOPLE WHO LIVED IN THIS DWELLING UNIT, BUT TRAVELLED TO LIVE ABROAD AFTER APRIL 1ST 2010 AND DID NOT RETURN UNTIL MIDNIGHT BETWEEN MARCH 31ST AND APRIL 1ST 2011 (EXCLUDING VISITS AND VACATIONS):

First name	Last name	Father's name	Gender 1. M 2. F	Date of birth (dd.mm.yyyy)	Ethnicity *(Use the codes from the list)*	Marital Status *(Use the codes from the list)*	Does he/she think to stay abroad over 12 months 1. Yes 2. No

Kosovo was exposed to important international emigration in the recent decades. In order to ensure that persons living abroad on a permanent basis were not included in the census, their identification was done through the above table.

Table Q15:

LIST OF PEOPLE WHO LIVED IN THIS DWELLING UNIT AT MIDNIGHT BETWEEN MARCH 31ST AND APRIL 1ST, 2011 AND ARE NOT LIVING HERE TODAY (INCLUDING PERSONS THAT HAVE PASSED AWAY SINCE 1ST OF APRIL 2011):

First name	Last name	Father's name	Gender 1. M 2. F	Date of birth (dd.mm.yyyy)	Ethnicity *(Use the codes from the list)*	Marital Status *(Use the codes from the list)*	Reason for not living here today 1-Moved out – in Kosovo 2-Moved out abroad 3-Deceased

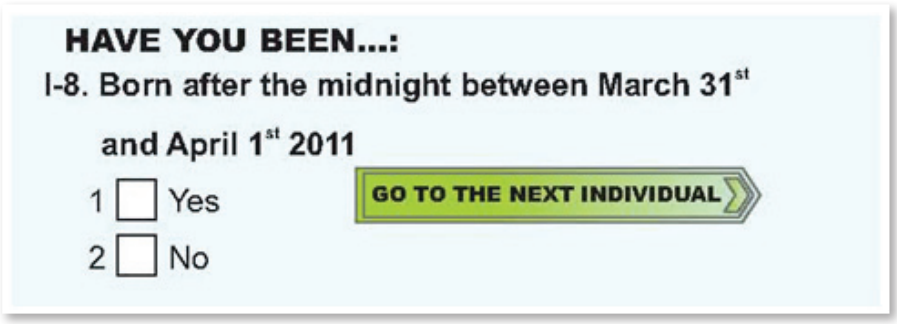
From the above table, people who should have been enumerated in the census but cannot be interviewed in the PES are identified. They include persons passed away between census day and PES time; persons that left this dwelling because they marry for instance, etc.

The individual questionnaire

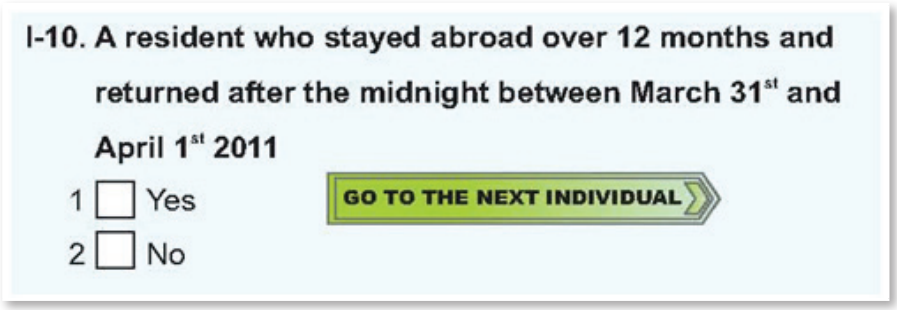
Once the general picture of persons living in the dwelling was sketched in the three above tables, each resident was asked a series of additional questions aimed at checking his/her eligibility to be enumerated during the census (I8 to I13), and if yes, where (I14 to I17). The information collected on the gender, age, ethnicity and marital status of individuals (I4 to I7), are data used in the census content errors estimates. The data collected on names and surnames (I1 to I3) are used to optimize the quality of linkage between PES records and Census ones.

Each person that should NOT have been enumerated in the census was identified; these cases correspond to each individual falling under a category "GO TO NEXT INDIVIDUAL" in the questions I8 to I13).

For example:

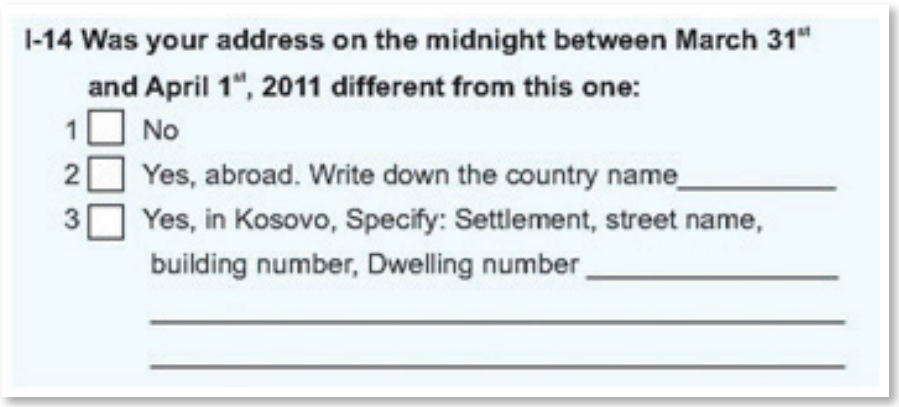


Persons born after the census day but before the PES time should not have been enumerated, or:



Persons residing in Kosovo at PES time but were not Kosovo residents at census day should not have been enumerated.

Once an individual is found as candidate for census enumeration (which is the case for all persons responding to the question I14) the remaining questions serve to identify at what place the person should have been enumerated. The question I14 makes an additional control of the address of the respondent at the time of the census, in order to exclude in the coverage measures people who should not have been enumerated.



I-14 Was your address on the midnight between March 31st and April 1st, 2011 different from this one:

1 ☐ No

2 ☐ Yes, abroad. Write down the country name_____

3 ☐ Yes, in Kosovo, Specify: Settlement, street name, building number, Dwelling number _____

_____

_____

Together with PES questionnaires additional field documents were prepared in order to allow for proper administration of the PES process as well as for guiding the field force in the data collection phase. Interviewers and supervisors instructions manuals were produced, accompanied with interviewers and supervisors report books, where field staff daily reported data on the survey progress, problems encountered, households interviewed etc.

CHAPTER 4

Sample Design

As stated in the introduction, the sample size for PES was defined by law and limited to 0.5% of the total population. The unique reference sampling frame available to extract the sample was spatial based and consisted in the last version of the total census Enumeration Areas, which amounted 4,681 for the whole Kosovo territory.

The sample was stratified with two factors: "ethnicity" and "urban/rural" type of area. PES sample design did not include institutional populations. The status of the sampling frame is described in table 1.

Table 1 – Initial sampling frame distribution of EAs by stratification factors

	Type of area			
Ethnicity	**1=Urban**	**2=Rural**	**3=Unknown or not inhabited**	**Total**
0=Not applicable	13	51		64
1= Albanian	1153	2690	3	3846
2=Serb	51	363		414
3= Ethnically mixed	85	183		268
4=Turkish		7		7
5=Bosniak		38		38
6=Goran		44		44
Total	**1302**	**3376**	**3**	**4681**

The following changes have been made in the sampling frame before the extraction of the sample:

- The 64 EA with "not applicable" code of Ethnicity are not inhabited, so they have been eliminated;
- The EAs with Ethnicity = 4,5,6 are inhabited by people speaking also Albanian, so they have been set equal to 1 (Albanians);
- The 3 EAs without indication of type of area code urban/rural have been eliminated.

So the initial sampling frame was transformed into a new one before the sampling procedures (table 2).

Table 2 – Final sampling frame distribution of EAs by stratification factors

	Type of area		
Ethnicity	**1=Urban**	**2=Rural**	**Total**
1= Albanian	1153	2779	3932
2=Serb	51	363	414
3= Ethnically mixed	85	183	268
Total	**1302**	**3376**	**4614**

As the census law limited at 0.5% of the total sampling frame the size of the PES sample, the size of the latter was to be composed of 23 enumeration areas. The sample (see Table 3) is the result of two independent selections:

- 21 EAs on a frame excluding the enumeration areas located in the North of the municipality of Mitrovicë/Mitrovica: 4391 EAs in total; and
- 2 EAs on a frame of composed of enumeration areas present only in north of Mitrovica: 223 EAs.

According to the sample design, 23 enumeration areas (EAs) were selected using balanced stratified design. Region, urban/rural and ethnicity variable were the stratifiers. The 16 sampling strata were formed by cross-classification of region (8) and ethnicity (where 1=Albanian, 2=Non-Albanian or mixed). The twenty one EAs were selected from these 16 strata. The 2 sampling strata in Mitrovicë/Mitrovica North were formed by Urban/Rural and one EA was selected from each strata. Table 3 presents the allocation of EAs in 18 different strata. All households and persons in the selected EAs were in the sample.

Likewise in the census there was lack of respondent cooperation in Mitrovicë/Mitrovica and Mitrovicë/Mitrovica-North regions sampling strata. Therefore, the coverage error estimates in the report do not include these areas. These strata are identified by "*" in table 3.

Table 3 - Allocation of sample EAs to sampling strata

	ETHNICITY		
Region	**Urban**	**Rural**	**Total**
PRISTINA CAPITAL	2	1	3
REGION FERIZAJ	1	2	3
REGION GJAKOVË	1	1	2
REGON GJILAN	1	2	3
REGON MITROVICA	1*	2	3
REGON PEJË	0	2	2
REGION PRISTINA	1	2	3
REGION PRIZREN	0	2	2
REGION MITROVICA NORTH	1*	1*	2
Total	11	12	23

NOTE:
Coloured cells marked with an asterisk correspond to enumeration areas excluded from coverage and content errors evaluations due to lack of P- and E- sample data in these areas;
PRISTINA CAPITAL corresponds to both the capital city and its surrounding settlements. The two urban EAs of Pristina Capital are located in the capital city and the rural EA of Prishtinë/Priština Capital is located in the region outside the capital city.

The sample was extracted evidently BEFORE the census took place. Since the latter could not take place in the northern part of the Ibër/Ibar river (including in Mitrovicë/Mitrovica-North, whose only the part located in south Ibër/Iibar edge was included in the census), a total of 3 enumeration areas included in the sample were not covered by the PES, because located in a zone excluded from the census.

At the end, the geographical area for PES data collection was composed of only 20 enumeration areas.

In order to be able to cover the 20 enumeration areas (which were envisaged to be covered in one week by census law) as soon as possible, some enumeration areas were divided in two parts, based on the number of dwellings recorded in ASK GIS database and as a result a total of 37 PES interviewing areas were used for the PES.

Table 4 indicates the inclusion probability and weight of each enumeration area included in the original PES sample selection.

Table 4 - Weight and Inclusion probability by Ethnicity and belonging stratum

Region	EA ID code	ETHNICITY	Stratum	weight	Probability of inclusion
Prizren	2	1	1	141	0.007092
Prizren	1	1	1	141	0.007092
Prishtinë/Priština	3	2	2	2	0.5
Prishtinë/Priština	1	1	3	222	0.004505
Prishtinë/Priština	1	1	3	222	0.004505
Prishtinë/Priština City (Capital)	176	1	4	36	0.027778
Prishtinë/Priština City (Capital)	154	1	5	450	0.002222
Prishtinë/Priština City (Capital)	262	1	6	8	0.125
Pejë/Peć	2	1	7	411	0.002433
Pejë/Peć	13	2	8	64.5	0.015504
Mitrovicë/Mitrovica	94	2	8	64.5	0.015504
Mitrovicë/Mitrovica	8	1	9	469	0.002132
Mitrovicë/Mitrovica	2	2	10	26.5	0.037736
Gjilan/Gnjilane	28	2	10	26.5	0.037736
Gjilan/Gnjilane	2	1	11	410	0.002439
Gjilan/Gnjilane	1	2	12	44	0.022727
Gjakovë/Đakovica	2	1	13	762	0.001312
Gjakovë/Đakovica	90	1	14	79	0.012658
Ferizaj/Uroševac	76	1	14	79	0.012658
Ferizaj/Uroševac	3	1	15	605	0.001653
Ferizaj/Uroševac	1	2	16	128	0.007813

NOTE:

ETNICITY 1=Albanian

ETHNICITY 2=Non-Albanian areas (mainly Serb populated)

CHAPTER 5

Data Collection

For statistical purposes Kosovo is divided in seven "statistically-administrated" regions, each of them having a branch of ASK dealing essentially with data collection relying on a pool of regular Interviewers for the various surveys ASK is conducting. The PES organization relied fully on this existing structure to optimize the quality of the interviews in the field.

The PES data-collection staff was composed of managers located in ASK headquarters in Prishtinë/Priština, regular and reserve Supervisors to cover each of the seven regions, and interviewers distributed in the seven regions in Kosovo (each EA was allocated two interviewers representing different ethnic community, as required for ethnically mixed EAs).

There have been several levels of fieldwork staff involved in the PES operation. The following two levels are based in the field:

Interviewer – Most of the interviewers were selected among usual interviewers used for statistical surveys by ASK, provided they were not also participating in census enumeration. Each Interviewer was responsible for a specific PES enumeration area. The enumerator is the person who interacts most closely with the public and collects PES data for all buildings, dwelling units and individuals in the allocated enumeration area. Interviewers were chosen on the basis of their experience as survey data-collection, their knowledge of the area to be covered, and their ethnic belonging. During the whole PES operation, interviewers were object of very close follow-up and control of their quality of work.

Supervisor – Was the person responsible for managing approximately 2-3 PES interviewers. The Supervisor worked very closely with them. The Supervisor role was to help the interviewers to do their work efficiently, to assist them in case of difficulty, to undertake certain checks designed to ensure that their work was accurate, to help with administration and to support the logistics aspects. He/she was the person who reported to the management PES team. The role of the supervisors was to facilitate the correction of inefficiencies and to maintain satisfactory progress during the enumeration period. The supervision process also helped to ensure the coordination between the Statistical Agency and the field operations.

Manager of field work Operations – was responsible for managing the whole PES field work operation, and the person to whom the Supervisors were reporting. Since in Kosovo PES the ethnic composition of the two main ethnic communities was rather high, the PES team had one manager for each main ethnic group. The interviewers participated in the training organized in two main languages based on their preference.

Before the start of PES field work, ASK PES-team organized specific training sessions for field supervisors and interviewers. The training sessions took four days. Communication with the field staff was maintained on a daily basis. In cases troubleshooting or additional advices was required from the field managers were intervening.

The survey was carried out during the period 16 April–22 April 2011, following the planned completion of census field work. The PES was operationally of a small scale compared to the census. The survey period was chosen to avoid overlap of census enumerators and PES interviewers in the field, while being close enough to census date (31 March).

Data was collected by 52 specially trained interviewers using the PES questionnaire. Field work was monitored by 17 Supervisors and by the PES team (6 people).

The face to face interview method was used in data collection. Enumerators canvassed all selected enumeration areas using GIS maps that have been also used for the census. Maps were showing all existing buildings, which were numbered. Interviewers and Supervisors had instruction manuals and report books, in which they were daily reporting on the progress of the survey.

Problems/obstacles encountered in the data collection of PES

The main problems that happened during the field work can be classified in three main categories:

Unclear borders of some EAs

During the data collection, two enumeration area maps appeared to have unclear borders:

EA Gjakovë/Đakovica (EA 090; sub-divided in 090-a and 090-b). This was a result of the vast urbanization changes in that EA in the recent period.

EA Polac/Poljance (EA 008). Due to the geographic particularity of the EA, namely the road access to the EA, its supervisor reported prior to the start of the fieldwork that there were difficulties in canvassing and ensuring the proper borders of the EA.

EA Gojbulë/Gojbulja (EA 002) had similar problem to Polac/Polance.

The above problems were solved by involving the GIS department who sent the same person that were sent in the Census as well in order to ensuring that all interviewers could identify the correct borders of their EA.

Refusals to respond to the interview

EA Sushicë/Sušica 001 (subdivided in 001-a and 001-b) was supposed to be been inhabited by Kosovo-Serb population only according to GIS data, while during the PES interviewers came across a dozen households with Albanians living there, who refused to participate in a survey conducted by Serb interviewers in Serbian. The problem was solved by sending an interviewer from Ferizaj/Uroševac (adjacent municipality) in order to interview the Albanian-speaking residents.

Apart from having the occasional refusal in other EA's, in the Serb inhabited Enumeration Areas this refusal rate was significant. The Serb inhabited EA's: Graçanicë/Gračanica, Gushtericë e Ulët/Donja Gušterica, Gojbulë/Gojbulja (to a lesser extent), Pasjan/Pasjane and Sushicë/Sušica all have refusal rates that are higher. Table 5 below provides statistics on the refusals during PES data-collection.

Other problems

Other problems reported concerned isolated specific cases. In the EA Trudë/Trudna (001b) for example, an interviewer informing the household about the PES was thrown out of the house because the inhabitants were not in agreement with the census undertaking. Yet, the interviewer still managed to collect some data from some neighbours.

Unreliable respondents (EA Vitomiricë/Vitomirica - EA 013b). It also happened that information for the PES was obtained from next-door relatives, for instance in cases where people present were unable to respond due to some handicap.

The following table shows the household sample size and response rate by sample EA for Kosovo PES sample. The response rates for two rural areas in Gjilan/Gnjilane, two rural areas in Prishtinë/Priština, and one rural area in Ferizaj/Uroševac were very low. The refusals were handled in household non-response adjustment procedure during errors estimation.

Table 5 - Number of Sampled, Responded and Refused Households (HHs) in P- sample

Region	Settlement	No of HHs who Participated in E-Sample	Total No of HHs in PES	No of HHs who REFUSED to participate in PES	No of HHs who responded in PES	P-sample response rate
Gjakovë/Đakovica	**Urban**	47	44	0	44	100
Gjakovë/Đakovica	**Rural**	81	62	0	62	100
Gjilan/Gnjilane	**Urban**	109	104	0	104	100
Gjilan/Gnjilane	**Rural**	13	54	42	12	22.22
Gjilan/Gnjilane	**Rural**	24	80	55	25	31.25
Mitrovicë/Mitrovica	**Rural**	25	72	26	46	63.89
Mitrovicë/Mitro	**Rural**	32	36	0	36	100
Pejë/Peć	**Rural**	80	68	0	68	100
Pejë/Peć	**Rural**	62	75	3	72	96
Prizren	**Rural**	37	36	0	36	100
Prizren	**Rural**	41	40	0	40	100
Prishtinë/Priština	**Rural**	25	61	45	16	26.23
Prishtinë/Priština	**Rural**	36	97	62	35	36.08
Prishtinë/Priština	**Urban**	151	142	0	142	100
Prishtinë/Priština	**Urban**	101	104	5	99	95.19
Prishtinë/Priština	**Urban**	69	44	2	42	95.45
Prishtinë/Priština	**Rural**	54	54	0	54	100
Ferizaj/Uroševac	**Urban**	131	128	0	128	100
Ferizaj/Uroševac	**Rural**	37	33	0	33	100
Ferizaj/Uroševac	**Rural**	15	24	15	9	37.5
TOTAL		**1170**	**1358**	**255**	**1103**	**81.2**

CHAPTER 6

Data-entry

Data entry and processing of the PES results was done after the data entry of census. PES data were entered manually at the Kosovo Agency of Statistics. The data entry program was prepared by the IT experts. Program (form) for the data entry was prepared according to the specifications elaborated by the PES staff. The whole data capture system was programmed with CSProX and the database managed in MySQL format. The logical and arithmetical coherency within and between the tables was incorporated in the data entry program. Besides entering the data, the application could produce different check lists: number of entered questionnaires per day, number of questionnaires entered with an error, list of errors, and statistics about the staff keying the data. These lists helped to monitor the whole process of data entry carried out by the PES staff of the data entry.

The data entry was build-up following a set of specifications, as described below.

6.1 Data Capture Specifications

As for the census data capture site, the following relevant aspects were also adopted in the PES data capture system:

1. The **working unit** is an enumeration area. However, since EAs were divided to 2, and were put in different boxes marked as A and B, the working unit can also be one box (half EA). The identification of the separate halves will be kept. The IT people may assign a numeric value instead of the A-B characters.
2. Questionnaires are in two **languages**; Albanian and Serbian. Keying operators are assigned accordingly (6 out of the 20 EAs are Serbian-speaking).
3. The **database** includes 5 tables:
    - 3.1. Geographic codes up to the dwelling unit serial number of the continuation questionnaires, plus a table with answers regarding the use of the buildings and the dwelling units (Questions 1-8), answers regarding the people who supplied the information for the tables (Questions 11, 14, 17), answers regarding the total numbers of people in the different tables (Questions 10, 13, 16) and the comments (page 3).
    - 3.2. First table of the people residing in the dwelling unit (Q9)
    - 3.3. Second table of people who left abroad before census day (Q12)
    - 3.4. Third table of people moved out after census day (Q15)
    - 3.5. The individual questionnaire linked to table 1.
    - 3.6. Statuses may be added letter to the (SQL) database.

4. **Over-riding skip-over** – if answers were given even if they were supposed to be ignored (stipulated by a previous answer).
5. **Verification** of data capture is done for questions that have more than one answer, missing values in critical questions, failed edit-checks within (out of range values) and between fields.
6. **Out of range** values for all multiple choice questions is the number of options provided.
7. **Missing values** are marked as an error and when exceeding a certain number (like over 5 in one questionnaire) they will pop up for controller's check.
8. The PES adopts the **Controller** duties and interfaces as in the census (same methodology).
9. **Closing an EA** is done only after check vis-à-vis the totals provided in the control and monitoring page (report book). The report book will not be keyed in but the paper version may be used as needed.
10. **Training** is done on real questionnaires and their data are erased a re-captured during data capture process.
11. **Quality Assurance (QA)** during data capture
    - 11.1. Is done by double blind entry of all questionnaires
    - 11.2. Reports will be generated for the controller and will be used for guidance during the data capture process.

6.2 Specific Instructions for data capture

Opening an EA record

1. The localities and EAs numbers in the sample are provided along with the totals for each working unit (EA or half EA).
2. An EA record is open from the list.

Page 1

1. The Serial number of the questionnaire is the leading identification field.
2. Geo codes are captured. As in the census the text address is not captured. It may be checked if needed in the census evaluation process.
3. If a continuation box is ticked then the option of capturing the three serial numbers of possible questionnaires is opened.

4. Questions 1-7 – Only one answer is allowed. If more than one was ticked then the key operator refers to the expert on site.
5. Questions 2 and 5 – Ignore the Go To instructions and allow keying in of additional (even if irrelevant) answers.
6. Questions 3, 4, 6 – if collective living quarter was ticked, then the name of institution is entered (bring missing value to the operator again to verify missing value).
7. Question 7 can be answered in categories 1, 2, 3 there may be additional information on the next pages of the questionnaire (Don't stop here).

   If category 4 was ticked, then the list in question 9 must have at least one person (bring it to be checked again).
8. Question 8 – Telephone number is entered as text.

Pages 2-3

9. Question 9 – The list of people residing in the dwelling unit is captured from all questionnaires belonging to the same dwelling unit consecutively according to the total number of people provided in Question 10 of the master questionnaire.

   If the list is empty, there is still a possibility that the list in questions 12 and 15 will carry information.
10. Question 10 – The PES team went through all continuation questionnaires and calculated and added the total number of people residing in the dwelling unit near the answer to question 10. This is the value entered.
11. Questions 11, 14, 17 - Only one answer is allowed. If more than one was ticked then the first one is captured (it is the closer person provided the answer).
12. Questions 12 and 15 – The data are captured as are. However,

    12.1. In year of birth allow value to be from 1911 to 2011 (edit check).

    12.2. If day and month of birth are missing but year of birth is filled in – do not adopt the census rule (imputing January 1st to all).
13. Question 13 – The value cannot exceed 12.
14. Question 16 – the value cannot exceed 11.
15. Comments – To allocate as much space as possible. Comments should be checked and entered in ALL CASES.

Individual Questionnaire

16. The number of records that is automatically opened is identical to the total provided in Question 10 of the Master Questionnaire.
17. However, a special venue is opened for additional Individual questionnaires (This may happen when data was provided for people in Questions 12 and 15 but not only). The keying operator is asked to check if there are additional individual questionnaires filled in. The individual record is still linked to the dwelling unit and carries a flag of not being included in the list.
18. Sequential Number – is supposed to follow the order of the list in question 9. It is different if it was written on a continuation questionnaire (15 and over).
19. Questions I1 to I-7 are defined as critical questions. Missing values are marked as errors to be checked again.
20. I-1 to I-3 – Names are checked against the list in question 9 and if not identical the fields are emptied and the operator has to re-key them. If the order of the individual questionnaires does not follow the list in table 9, it is fixed after data capture.
21. I-8 – If Yes (option 1), check against I-5, if Okay follow the GoTo, otherwise continue and ignore the GoTo.
22. I-9 – Bring up a message to check weather there is additional information on this individual (on the same page0. If yea, allow to continue, if No - GoTo the next individual.
23. I-10 to I-13 – Follow the GoTo instructions.
24. I-14 and I-15 – Country name to be picked from a list as in the census. If not in the list, assign a 9999 code and enter the text.
25. I-14 and I-15 – Address in Kosovo to be captures as string.
26. I-14 to be checked against the list in question 9: If answered No in I-14 but in question 9 answered that did have a different address on census day then bring I-14 answer to be checked again.
27. I-15 to be checked against the list in question 9: If answered No in I-15 but in question 9 answered that did have a second address on census day then bring I-15 answer to be checked again.
28. I-16 and I-17 – to be defined as 5 and 7 questions respectively to allow answers to all options. Any reduction of irrelevant information is done during processing.

CHAPTER 7

Methodology

7.1 Principle of independence

As in any evaluation process, the principle of independence between the evaluated domain and the evaluation tools is a basic requirement. In order for the PES to achieve its objectives, its processes need to be independent from the census.

To ensure this independence, the PES:

- Was defined and managed by a group of persons who were not involved in the census and had no responsibility regarding the census;
- Used different field staff than the census one;
- Was conducted after the census field work was completed to avoid contact between census enumerators and PES interviewers. The 2011 PES used more tightly controlled collection procedures, and more experienced and better trained field staff than the census;
- The PES sample was maintained confidential so that census field staff and office staff were not aware which areas were included in the PES;
- PES was conducted immediately after the census;
- Same definitions and classifications were used in the PES as in the census.

7.2 Dual System Estimation methodology

PES used the Dual System Estimation methodology which is based on capture and recapture methodology (Chandrasekaran-Deming estimator (1949)). The methodology estimates the total population. The DSE model is conceptualized that each person has a probability of being either included in the census or not included in the census as well as either included or not included in PES. This can be described as in the following table.

Table 6 – DSE model

	In Census	Not in Census	Total
In PES	N_{11}	N_{12}	N_{1+}
Not in PES	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	N_{++}

In the above table, all the cells are observable except N_{22}; as a consequence all the marginal that included N_{22} are also not observable. The model assumes independence between the census and the PES. Hence the probability of being in ij^{th} cell P_{ij} is the product of two marginal probabilities that contains ij^{th} cell.

Thus, under the independence assumption, the estimate of total population is

$$DSE = N_{++} = \frac{(N_{+1})\,(N_{1+})}{N_{11}}$$

So the total population can be written as a function of number included in census, in PES and those included in both. This model is applied within each estimation domain usually called post-stratum.

In practice, the components of DSE are estimated from samples. In the table 6, N_{+1} is not the census counts. Thus, the census counts should be corrected for erroneous enumeration. Also, the persons with insufficient information for matching could not be matched with PES enumeration. Hence, census counts must be corrected for enumeration with insufficient information to match with PES enumeration.

Thus, DSE uses the following formula;

$$DSE = DD * \frac{CE}{N_e} * \frac{N_p}{M}$$

Where;

DD = the number of census data-defined persons eligible and available for PES matching,
= (census count) minus (enumeration with insufficient data for matching)
CE = the estimated number of correct enumerations from the E-sample,
N_e = the estimated number of people from the E sample
N_p = the estimated total population from the P sample
M = the estimated number of persons from the P sample population who matched census.

To implement the methodology, CE and M are obtained by matching P- and E- sample persons. The PES team performed the matching operations with an expert support.

7.3 Post-strata (Estimation Domains) for DSE

As stated above, DSE for total population is computed by estimation domain also called post-stratum. This is done to increase the efficiency of estimates by reducing the mean square of estimates. The definition of post-strata should be such that the persons (or households) within each group have the similar inclusion probability in census but it differs between different post-strata. In addition, each post-stratum should include sufficient number of sample cases.

7.3.1 Post-strata for person Estimation

Initially tallies for each of the 8 regions by urban/rural by ethnicity (2 groups Albanian and others) by sex and by age (4 groups) were obtained, totalizing 256 strata. However, many of these strata were either empty or contained very small numbers. Therefore, 4 age classes were reduced to 3. Also, many other post-strata for other than Albanian ethnic group were combined to have sufficient sample cases. As a consequence, only 38 post-strata were retained for person estimation. The following table shows the post-strata used for person estimation.

Table 7 - Post-strata definition for person estimation

		Albanian			Other
		Age group			
Region		0-19	20-45	46+	
Gjakovë/Đakovica	Urban				#
	Rural				#
Gjilan/Gnjilane	Urban				
	Rural	#	#	#	
Mitrovicë/Mitrovica	Rural				
Pejë/Peć	Rural				
Prizren	Rural				
Prishtinë/Priština	Urban				
	Rural				
Ferizaj/Uroševac	Urban				#
	Rural				

NOTE: '#' means an empty stratum. Prishtinë/Priština corresponds to both the capital city and the Prishtinë/Priština region. The urban corresponds to the capital city and the rural corresponds to the region outside the capital city.

7.3.2 Post-strata for Dwelling (Housing Unit) Estimation

The matching operations were carried out only for individuals and not for the dwellings. Therefore, the DSE and coverage errors of dwelling were not estimated.

7.4 Coverage Errors

Coverage error refers to either an under-count or over-count of units (persons or dwellings) owing to omissions of persons/dwelling units or duplication/erroneous inclusion, respectively. The following are the different types of coverage errors the PES estimated and evaluated.

Net coverage error

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

$$Net\ coverage\ error = True\ Population - Census\ count$$

Net coverage error rate

This 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.

$$Net\ coverage\ error\ rate = \frac{DSE - Census\ count}{DSE} * 100$$

Census omission

This is the difference between the true population and the census count with the exclusion of the erroneous inclusions.

$$Omission\ rate = True\ Population - Correct\ Enumeration$$

Census omission rate

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

$$Omission\ rate = \frac{Census\ Omission}{True\ Population\ estimate} * 100$$

Census erroneous inclusion

The erroneous inclusions include fabrications, duplications and geographic misallocations. It is computed as:

$$Erroneous\ inclusions = (Census\ Counts + Omission) - True\ Population$$

Census erroneous inclusion rate

Census erroneous inclusion rate is the census erroneous inclusions relative to the true population estimate.

$$Erroneous\ inclusion\ rate = \frac{Erroneous\ inclusions}{True\ Population\ estimate} * 100$$

The coverage error methodology is described in (Census 2004) and was adopted for Kosovo PES with modifications as necessary.

7.5 Content Errors

Content error measures the discrepancy between the census and the PES data. Content error is only estimated for the matched persons and for the selected variables in order to measure the inconsistency between the answers captured from census and PES to the same questions. It is important to know the inconsistency between data from the two sources (census and PES) for those items that are used for PES matching operations and forming post-strata. The inconsistencies in these items lead to bias in DSE results. Ideally, the data from the two sources should be the same. However, this is not the case in practice.

This inconsistency is measured by means of four indicators: the net difference rate; index of inconsistency (simple and aggregated), the gross difference rate; and the rate of agreement [UNSD (2010)]. It is also desirable to review bi-variate table to determine item groupings for matching, defining estimation domains (post- strata) and estimating coverage errors.

The content error analysis was done at the national level for sex, age, ethnicity and marital status. The content error indicators are described below.

Net Difference Rate

The net difference rate (NDR) is the difference between the number of cases in census and the number of cases in PES that fall under each response category, relative to the total number of matched persons in all response categories.

Index of Inconsistency

It can either be simple or aggregated and is calculated for each response category. The simple index of inconsistency (I) is the relative number of cases for which the response varied between the census and the PES.

Aggregate Index of Inconsistency

The Aggregate Index of Inconsistency (AII) is a summary measure of the index of Inconsistency (that is for all the response categories of the characteristic as a whole).

Gross Difference Rate

The Gross Difference Rate (GDR) also referred to as "Off-Diagonal proportion" is calculated for the characteristic as a whole. It is the number of discrepancies between the census responses and the PES responses relative to the total number of persons matched. It is equivalent to the sum of all cells off the diagonal, for all categories, or the complement of the sum of the diagonal cells.

Rate of agreement

Rate of agreement (RA) is the complement of the gross difference rate (GDR).The rate of agreement indicates the level at which the information given in the Census matches that given during the PES. A low rate of agreement indicates a high degree of variability and vice-versa. The rate of agreement is therefore a good measure of the gross error for an item.

7.6 Variance Estimation

The variances for coverage errors were computed only for national level and urban and rural. Due to lack of sufficient data, variances for lower level of geography or demography were not computed. Jackknife estimator was used to estimate variances. A detailed description of the methodology is presented in appendix 2. For more details, see Wolter (1985, and 1986).

CHAPTER 8

Data Processing

8.1 Attribution of moving statuses

The first processing of PES data consisted in attributing moving statuses based on the information collected in PES questionnaires. Statuses to be attributed could be:

1 **Non-movers:** persons who were in a particular household as of the census and PES date.
2 **Out-movers:** persons who were in a particular household at the census date but moved out or were not part of the household at the time of the PES
3 **In-movers:** persons who were enumerated in a particular household during the PES but where not there during the census date
4 **Out of scope:** persons who do not belong to the target population as of the Census date, for example a child born after the census date

This stage was essential to understand, after the record-linkage, whether the census did or not include the appropriate people. The rules followed for attribution of status were as follows, based on PES questionnaire:
1 = Non-movers = If present in Table 9 (or has an individual questionnaire) and I-8 = 2 and I-9 = 2 and I-10 = 2 and I-11 = 2 and (I-11 = 1 and I-11.1 = 1) and I-12 = 2 and I-13 = 2 and I-14 = 1; **OR** Table 12 and intent to stay for more than 12 months =2;

2 = Out-movers = all those in table 15

3 = In-movers = If present in Table 9 (or has an individual questionnaire) and I-14 > 1

4 = Out of scope = If present in Table 9 (or has an individual questionnaire) and I-8 = 1 or I-9 = 1 or I-10 = 1 or (I-11 = 1 and I-11.1 = 2) or I-12 = 1 or I-13 = 1; **OR** Table 12 and intent to stay for more than 12 months =1.

8.2 Matching procedures

After PES data were captured, P-sample and E-sample were ready to be compared (by record-linkage methodology). To this aim, some programmes of fully automated record-linkage were first developed, and they could solve a large majority of the cases. For unsatisfactory matches a specific application has been developed with MS-ACCESS tool, allowing a computer-assisted comparison between all cases of E- and of P-sample. Paper questionnaires have been also examined when it was found impossible or uncertain to match cases, in order to ensure the quality of the processes.
The first identifiers for linkage were the geographical codes. This process had to be done with a specially-designed application since it was soon discovered that the geo-codes corresponding to

a same dwelling were not always identical in the PES and in the census. Indeed, the enumerators (census) and interviewers (PES) had the freedom to inserting the codes starting from the entrance one (see chapter on questionnaire). If they did not follow the same order, geo-codes were not matching for a same location. In order to overcome this critical problem, ASK IT staff developed an application allowing a computer-assisted matching procedure for the dwellings.

Once dwellings were identified, automatic processes were trying to link census and PES records, taking into account variables such as name identifiers, date of birth, sex, marital status, and ethnicity.

This operation led to a series of "perfect matched"; "partially" and "not" matched cases.
Perfect matches correspond to all cases where all variables used from P- and E- sample records were found perfectly identical. Partial (also called relaxed) matches correspond to cases where not all the variables were found exactly identical. In most cases, this was concerning the text strings (names, surnames) where data reporting or data capture errors were producing parasites. Semi-automatic checks were done on most of these cases. If not enough data were matching, the cases remained unresolved and were candidate for reconciliation visit.

8.3 Reconciliation visits

When cases were not solved or not giving enough information, re-controlling of the cases were done in the field. Interviewers were sent with a series of precise questions to ask about specific persons and households, in order to confirm, infirm or complete a status of a person (see next section). This process was done for each person of each EA separately.

8.4 The final enumeration statuses

Rules to decide whether a person was correctly enumerated or not were applied for both the P-sample and the E-sample.
They are described below.

8.4.1 Enumeration status in the P-sample

Possible statuses in the P-sample are:
1: Correct
2: Erroneous inclusion
3: Duplicate
4: Omission
5: Undefined

These statuses are obtained from a combination of the original status in the P-sample with the result of matching (record linkage) with the E-sample.

Possible combinations and the rules applied are described in table 8.

Table 8 – Rules for the attribution of enumeration statuses in the P-sample

Status in P-Sample			**Enumeration**
Non-mover			**status**
In E-sample	Perfectly Matched	1	
	Relaxed match	1	
	Not matched	**Go to reconciliation.**	
	After reconciliation:	Found, and resided in EA at census	4
		Found but resided out EA at census	1
		Not found	5
Status in P-Sample			
Out-mover			
In E-sample	Perfectly Matched	1	
	Relaxed match	1	
	Not matched	**Go to reconciliation.**	
	After reconciliation:	Found, and resided in EA at census	4
		Found but resided out EA at census	1
		Not found	5
Status in P-Sample			
In-mover			
In E-sample	Perfectly Matched	Go to reconciliation	
	Relaxed match	Go to reconciliation	
	After reconciliation:	Found, and resided in EA at census	4
		Found but resided out EA at census	2
		Not found	5
	Not matched	**1**	
Status in P-Sample			
Out of Scope			
In E-sample	Perfectly Matched	Go to reconciliation	
	Relaxed match	Go to reconciliation	
	After reconciliation:	Found, and resided in EA at census	1
		Found but resided out EA at census	2
		Not found	5
	Not matched	**1**	

Status in P-Sample			
Unidentified			
In E-sample	Perfectly Matched	Go to reconciliation	
	Relaxed match	Go to reconciliation	
	After reconciliation:	Found, and resided in EA at census	1
		Found but resided out EA at census	2
		Not found	5
	Not matched	**Go to reconciliation.**	
	After reconciliation:	Found, and resided in EA at census	4
		Found but resided out EA at census	1
		Not found	5

8.4.2 Enumeration status in the E-sample

Possible statuses in the E-sample are:
1: Correct
2: Erroneous inclusion
3: Duplicate
4: Omission
5: Undefined

These statuses are obtained from a combination of the original status in the E-sample with the result of matching (record linkage) with the P-sample.

Possible combinations and the rules applied are described in table 9.

Table 9 - Rules for the attribution of enumeration statuses in the E-sample

Person in E-Sample		
Is also in P-sample	Case already solved when matching P-sample with E-sample	
Person in E-Sample		
But is not found in P-Sample	All cases for reconciliation	
	AFTER RECONCILIATION:	
	Was resident in the EA	1
	Was not resident in the EA	2
	Person is not found in field	5

Estimates could start according to DSE methodology once all the cases have received an enumeration status in both P and E samples.

CHAPTER 9

Analysis of Coverage and Content errors

The results on coverage and content errors are presented in this Section. The PES sample was very small due to legal requirements. Therefore, coverage errors presented here are at the national and at urban/ rural levels for the entire territory. The content errors are presented only at the national level for four demographic characteristics. The Jackknife variance estimator (Wolter 1985, 1986) was used for the national estimates.

Note that in all tables presented at region level, Prishtinë/Priština corresponds to both the capital city and the Prishtinë/Priština region. The urban corresponds to the capital city and the rural corresponds to the region outside the capital city.

9.1 Coverage errors

The following table (Table 10) presents, total population estimate and coverage errors for Kosovo and urban rural areas in Kosovo.

The estimates in the table hereafter are computed independently for each geographical area and were not obtained by adding rural and urban estimates to obtain national level estimates. Therefore, summing them will not add to national level estimates.

The table shows that the census had under-counted Kosovo population by 40,587 persons. Both urban and rural populations were under-counted in the census. Census under-counted the urban population by 26,403 and the rural population by 15,513.

The net under-count rates were 2.30 (2.2; 2.4) percent at national level; 3.83 (2.48; 5.2) percent at urban level, and 1.43 (-0.49; 3.3) percent at rural levels. Urban areas had thus a slightly larger under-count than rural ones.

The Census omissions at the national level were estimated to be 76,109 and at the urban/rural levels were estimated to be 33,205 and 42,585, respectively. The omission rates were 4.3 (4.2; 4.4) percent at national level; 4.75 (2.9; 6.6) percent, and 3.6 (-2.1; 9.3) percent, for urban and rural areas, respectively.

Erroneous inclusions in Kosovo Census were estimated to be 35,523 and at the urban/rural levels were estimated to be 6,803 and 27073, respectively. The erroneous inclusion rates were 2.0 (-0.2; 4.2) percent at national level, 0.99 (-3.7; 5.6) percent in urban areas and 2.5 (-4.4; 9.4) percent in rural areas.

Among all coverage error rate indicators, the census omissions has the highest level (4.3% at national level) while erroneous inclusions rate is the lowest (2%), especially in urban areas (0.99%).

Table 10 - Coverage errors and their rates at national, urban and rural levels

		Urban	Rural	National
DSE	Persons	689621	1086492	1774784
	CI 95%	678526;700715	1022434;1150550	1658209;1891358
Census counts		663218	1070979	1734197
Net coverage Error	Persons	26403	15513	40587
	CI 95%	(17091; 35714)	(-5009;36035)	(38427;42447)
	Rate	3.83	1.43	2.30
	CI 95%	(2.48; 5.2)	(-0.49;3.3)	(2.2; 2.4)
Census omission	Persons	33205	42585	76109
	CI 95%	(19968; 46442)	(-6084; 91254)	(74288;77931)
	Rate	4.75	3.6	4.3
	CI 95%	(2.9; 6.6)	(-2.1; 9.3)	(4.2; 4.4)
Census erroneous inclusion	Persons	6803	27073	35523
	CI 95%	(2417; 11188)	(6531; 47613)	(19367; 51678)
	Rate	0.99	2.5	2.0
	CI 95%	(-3.7; 5.6)	(-4.4; 9.4)	(-0.2; 4.2)

Note: "CI 95 %" in the table 10 means 95% confidence interval.

As shown in table 11 hereafter, two strata in Prizren have their match and the correct enumeration rates equal to one. Also census enumeration had all persons data defined in these strata. Therefore, DSE and census enumeration counts were the same which resulted in a net coverage error of zero for these strata. This situation probably also was observed due to small PES sample size.

Table 11 - Net coverage errors for persons by post-strata

Net coverage errors By estimation post-strata (Persons)					
Region	Urban/ Rural	Albanian			Other
		Age group			
		0-19	20-45	46+	
Gjakovë/Đakovica	Urban	2,780	1,487	350	
	Rural	2,340	4,182	80	
Gjilan/Gnjilane	Urban	1,338	1,337	661	-71
	Rural				3,206
Mitrovicë/Mitrovica	Rural	1,074	-1,013	-506	5,076
Pejë/Peć	Rural	436	920	-1,076	1,155
Prizren	Rural	0	714	0	-1,276
Prishtinë/Priština	Urban	1,280	6,119	1,634	-241
	Rural	-1,302	-272	4,426	2,620
Ferizaj/Uroševac	Urban	343	434	692	
	Rural	1,784	4,772	791	6,110

NOTE: Positive number means under-count and negative means over-count.

Table 12: Net coverage error rates in percent for persons by post-strata

Net Coverage Rate (%)					
Region	Urban/ Rural	Albanian			Other
		Age group			
		0-19	20-45	46+	
Gjakovë/Đakovica	Urban	12.50	6.35	2.31	
	Rural	4.33	-8.61	0.29	
Gjilan/Gnjilane	Urban	5.41	4.93	4.10	-4.05
	Rural				25.81
Mitrovicë/Mitrovica	Rural	2.37	-2.36	-2.19	87.29
Pejë/Peć	Rural	1.06	2.22	-4.82	10.85
Prizren	Rural	0.00	0.96	0.00	-4.14
Prishtinë/Priština	Urban	1.66	6.46	3.07	-2.52
	Rural	-1.43	-0.29	8.45	14.02
Ferizaj/Uroševac	Urban	1.59	1.86	5.10	
	Rural	3.53	9.20	3.14	60.99

NOTE: Positive number means under-count and negative means over-count.

Tables on omissions and erroneous rates are discussed below. The tables with omissions and erroneous inclusions for each post-stratum are provided in Appendix 4. They show similar results as the net coverage errors.

The tables 13 and 14 hereafter present the omission rates and erroneous inclusion rates by post-strata. The observed rates for Albanians are much lower than the rates for "Other" ethnic groups. Some of the rates (95.62 and 63.63) are extremely high. Mitrovicë/Mitrovica rural strata has very low response rate and Ferizaj/Uroševac rural has very small size either due to EA sample or due to high refusal in the EA.

Table 13 - Omission Percent Rates by Post-strata

Omission Rate By Post-strata (%)					
Region	Urban/ Rural	Albanian			Other
		Age group			
		0-19	20-45	46+	
Gjakovë/Đakovica	Urban	12.50	6.35	4.08	
	Rural	6.35	2.03	2.08	
Gjilan/Gnjilane	Urban	5.95	5.98	4.76	5.96
	Rural				25.81
Mitrovicë/Mitrovica	Rural	2.37	0.00	0.00	95.62
Pejë/Peć	Rural	2.75	6.97	1.75	10.85
Prizren	Rural	0.00	1.87	0.00	3.08
Prishtinë/Priština	Urban	1.95	7.10	3.07	0.31
	Rural	0.00	1.12	8.45	15.83
Ferizaj/Uroševac	Urban	2.65	2.86	6.15	
	Rural	3.53	9.55	3.14	63.63

Table 14 - Erroneous inclusion rates (percent) by post-strata

Erroneous Inclusion Rates (%)					
Region	Urban/ Rural	Albanian			Other
		Age group			
		0-19	20-45	46+	
Gjakovë/Đakovica	Urban	0.00	0.00	1.78	
	Rural	2.02	10.64	1.80	
Gjilan/Gnjilane	Urban	0.53	1.05	0.66	10.01
	Rural				0.00
Mitrovicë/Mitrovica	Rural	0.00	2.36	2.19	8.32
Pejë/Peć	Rural	1.69	4.75	6.57	0.00
Prizren	Rural	0.00	0.91	0.00	7.22
Prishtinë/Priština	Urban	0.29	0.64	0.01	2.83
	Rural	1.43	1.41	0.00	1.81
Ferizaj/Uroševac	Urban	1.06	1.00	1.04	
	Rural	0.00	0.34	0.00	2.65

NOTE: Since most of the net coverage errors were accounted by omissions, the erroneous inclusion rates were comparatively smaller than other coverage error measures.

Most of the sample stratum had one EA in sample. An evaluation of how the sample strata estimates compared to enumeration counts was performed. The following table 15 presents the estimates based on the E-sample and the corresponding census counts. The table shows that E-sample estimates of sample stratum and the census enumeration counts have very large differences in almost all of the sample strata. These differences have significant effect on small area estimates of coverage errors.

The sample persons in P- and E-sample are not relatively that far apart as the census counts and their corresponding estimates. The estimates based on P-sample would show the similar results. The table also shows the difference between numbers of P- and E-sample responding households

Table 15 - E-Sample Estimates v/s Census Counts by estimation Post-strata

Region	Settlement	Age group	Ethnicity	E- Sample Estimate	Census Counts	P-sample Size	E-sample Size	Sample size difference
Gjakovë/ Đakovica	Urban	0-20	Albanian	432	19462	48	54	-6
		21-45	Albanian	576	21937	65	72	-7
		46+	Albanian	440	14820	51	55	-4
	Rural	0-20	Albanian	63906	51726	128	142	-14
		21-45	Albanian	87309	52756	158	194	-36
		46+	Albanian	49955	27834	99	111	-12
Gjilan/ Gjilane	Urban	0-20	Albanian	73161	23376	202	178	24
		21-45	Albanian	74394	25789	214	181	33
		46+	Albanian	59597	15469	156	145	11
Mitrovicë/ Mitrovica	Rural	0-20	Albanian	39378	44314	156	150	6
		21-45	Albanian	41822	43906	143	142	1
		46+	Albanian	21902	23616	79	75	4
Pejë/Peć	Rural	0-20	Albanian	72063	40626	234	232	2
		21-45	Albanian	75915	40474	248	253	-5
		46+	Albanian	52323	23395	156	166	-10
Prizren	Rural	0-20	Albanian	48397	79570	81	80	1
		21-45	Albanian	65941	73908	112	109	3
		46+	Albanian	49002	36609	79	81	-2
Prishtinë/ Priština	Urban	0-20	Albanian	49578	75818	499	533	-34
		21-45	Albanian	66523	88663	607	665	-58
		46+	Albanian	31962	51628	314	343	-29
	Rural	0-20	Albanian	108232	92458	143	142	1
		21-45	Albanian	108232	94560	145	142	3
		46+	Albanian	53354	47934	81	70	11
Ferizaj/ Uroševac	Urban	0-20	Albanian	61931	21210	278	279	-1
		21-45	Albanian	65705	22912	291	296	-5
		46+	Albanian	40400	12878	183	184	-1
	Rural	0-20	Albanian	14896	48732	60	78	-18
		21-45	Albanian	19108	47080	90	102	-12
		46+	Albanian	14902	24437	69	73	-4

Region	Settlement	Age group	Ethnicity	E- Sample Estimate	Census Counts	P-sample Size	E-sample Size	Sample size difference
Gjilan/ Gnjilane	Urban		Others	1829	15618.58	26	38	-12
Gjilan/ Gnjilane	Rural		Others	9217	11222.91	179	174	5
Mitrovicë/ Mitrovica	Rural		Others	739	264.9989	103	10	93
Pejë/Peć	Rural		Others	9491	6424.077	197	146	51
Prizren	Rural		Others	32093	24318.44	184	190	-6
Prishtinë/ Priština	Urban		Others	9820	17590.35	69	243	-174
Prishtinë/ Priština	Rural		Others	16072	7268.131	232	92	140
Ferizaj/ Uroševac	Rural		Others	3908	2147.892	37	39	-2

9.2 Content Errors

The content error analysis was performed for age, sex, marital status, and ethnicity using the matched persons in P-and E- samples. To analyze content errors, a bivariate *nxn* table was formed for each of the demographic under study. The groups of each demographic were formed to be consistent with the DSE estimation post-strata. The marital status was not used to estimate DSEs but there was an interest to learn about content error for marital status. The content error in the census and PES was exceptionally low for all demographic characteristics except certain categories of marital status.

The following table shows the consistency of captured data for age groups. It is important to review and understand the discrepancy in the data to improve matching operations, and estimation of coverage errors in census. One can see that diagonal cell in the table have the concentration of persons which means that the data on a person's age was mostly captured in the same age group in both P- and E- samples. In other words, there was consistency in reporting and capturing the data in census and PES. Numbers in off diagonal cells shows that person's data was recorded in different age categories for a person in both census and PES. If discrepancies are noticed, it is important to learn how these discrepancies are distributed.

TABLE 16 - P-and E-sample Persons' Response Captured Data by Age Groups

		P sample			
	Age group	**0-19**	**20-45**	**45 +**	**Total**
E sample	**0-19**	1905	43	27	**1975**
	20-45	39	2151	34	**2224**
	45+	29	28	1391	**1448**
	Total	**1973**	**2222**	**1452**	**5647**

Interpretation of content errors

The content errors are usually given as either percentage or a ratio. To aid in the interpretation, UNSD (2010) provided the following table that gives ranges for each measure.

Table 17 - Interpretation of Indices for Content Errors

	Level		
Measure	Low	Medium	High
Index of inconsistency	<20	20 - 50	>50
Aggregate index of inconsistency	<20	20 - 50	>50
Absolute Value of NDR	<0.01	0.01 -0.05	>0.05

Source: UNSD (2010), Post Enumeration Survey, Operational Guidelines

Over all the content error for the demographics studied here are extremely low. The indices for agreement are over 92% and as high as 99.3%. The content error results are interpreted using the above table. Content error for age is discussed below. The following table presents content error indices for age groups.

TABLE 18 - Content Error Indices for Three Age Groups

Age groups	NDR	Index of inconsistency
0-19	0.00035	5.37401
20-45	0.00035	5.34167
45+	-0.00071	5.47473
Aggregate Inconsistency Index	5.39152	
Gross difference rate	0.03542	
Rate of agreement	96.45830	

Five indices were computed for each characteristic under study. These indices for age are discussed below. The tables for other characteristics are included in appendix 4.

The net difference rate for age groups are between 0.00035 and 0.00071 for all age groups. Indices of inconsistency are between 5.34 and 5.47 which are low. The aggregate inconsistency index is 5.39% which is also in low range. The gross difference 0.035 is also small. Using criteria from the table 17, it is clear from the table 18 that content error indices for age groups are low, that is, a high level consistency in the data from two sources.

Tables in Appendix 4 show that the rate of agreement for age, sex, marital status and ethnicity ranged between 92.2% and 99.3% which was exceptionally high. Aggregate inconsistencies indices for these characteristics ranged between 3.3 and 5.4 except for marital status which was 13.5. Indices of inconsistencies ranged between 3.3% and 5.5% for all characteristics except for marital status. The indices of consistency for marital status were between 6.2% - 38.2%. The lowest index of inconsistencies was for never married and the highest was for divorced status. All marital statuses had medium level index of inconsistencies except for never married which has low level inconsistencies. The net difference rates were about zero and the gross difference rate was between 0.007 and 0.08 for all characteristics.

CONCLUSION

Over all PES worked in a very satisfactory way. PES collected complete data and had no missing data for persons' characteristics that were used for matching and estimation. This was an exceptional achievement for PES. The data on the similar characteristics was also complete in the census.

Considering that it was the first census in 30 years, the results of census from PES evaluation look very good. The data quality in terms of completed data is also exceptionally good for those who responded in PES or census.

The match rates and correct enumeration rates are over 90 percent and in some cases 100 percent.

PES also provided good outcomes from the reconciliation visits since all statuses needed from matching operations were confirmed except for 53 out of over 6,000 persons in P-sample (less than 0.8%).

The main – yet important - limitation of the Kosovo PES was the sample size. The sample size due to legal requirement was too small to estimate coverage errors below the national level. This had a significant impact on reliability of PES estimates. The sample size for the national level estimates could have been larger to allowing estimate coverage errors by geographic or demographic characteristics. Given the sample size constraints and no prior knowledge how coverage may vary by demography or geography in Kosovo, the sample design was good.

Another limitation of estimating coverage errors for all regions by urban rural or by sex or age groups was the lack of sample data in cross-classification of these variables. Part of the reason besides the sample size was that the most members of ethnic groups other than Albanian did not participate in the census and refused to answer both census and PES.

Because of the significant variation between EAs total population, an estimate for a stratum, based on sample EA was extremely different from the census counts for the stratum (see table 4.3). These large differences may have given unreasonable high coverage errors for a number of post-strata. These data are thus poorly reliable and should be used with caution.

The prior or historical demographic data was not available for sample design since this was the first census in 30 years. Also, information was lacking on what factors may affect the coverage in Kosovo census. In spite of lack of information on how coverage varies within urban or rural areas, urban/rural and ethnicity were considered good predictors of coverage error. Such predictors have been used in other countries.

The refusal rates for certain sample strata were high because non- Albanian ethnic groups (in particular Serbs) did not co-operate during census data collection process. Three sampling strata were could not be included in evaluating coverage and content errors since they were pertaining to the area where the census could not take place.

PES did not match the dwellings during the matching operations. Therefore, estimates of housing unit coverage errors were not produced.

Over all content errors were exceptionally low for all characteristics studied except for married without certificate, widowed and divorce marital status. The lowest index of inconsistencies (6.24%) was for never married and the highest (38.24) was for divorced status. The indices on consistency for never married and married with certificate were low level and married without certificate, widowed and divorce were in the medium range.

Based on the above mentioned limitations having distorted the PES process, the following cautions are recommended for current PES results interpretation and for future implementation of such survey:

a) Various estimates of coverage errors are produced with small number of persons and households in sample. Even though these estimates provide significant information about certain errors and for designing future census and PES, one should be careful in using these estimates.

b) Census unduplicated households that were part of the E-sample. Thus, E-sample did not represent the target population in census. Therefore, it has potential for bias in coverage error estimates. However, duplicated households in P- and E- samples were about 1%. Hence, the bias in estimates was expected to be small.

c) The PES was exposed to a piece meal approach in using experts and professionals in designing and implementing PES. Experience in implementing PES in other countries showed that it is on the contrary suitable that PES leaders have good understanding of the entire plan and the data requirements to complete PES efficiently. Therefore, it is recommended that future PES uses the same expert for both sample design and estimation activities. This expert should also be involved to the maximum extent possible in developing PES questionnaire, data collection instructions, matching operations including reconciliation aspect of the PES.

d) Due to small sample size requirement enforced by law, only one or two EAs were selected in a post-stratum. Such a small sample could not provide estimates with lower variance for estimates at sub-national level or by demographic groups at national level. Therefore, sample size should be determined by the objectives of PES and not by legal provisions.

e) An accurate matching operation is critical for the success of the PES. Also, it is one of the most complex components of PES operations. This should be well planned and more time should be devoted in developing and implementing the entire matching operation including the reconciliation questionnaire and data collection instructions.

f) A pilot test should be conducted to test various steps of the PES. If entire operation could not be tested due to lack of resources, at least the PES questionnaire, reconciliation questionnaire and the matching operation should be tested prior to finalizing operations for PES.

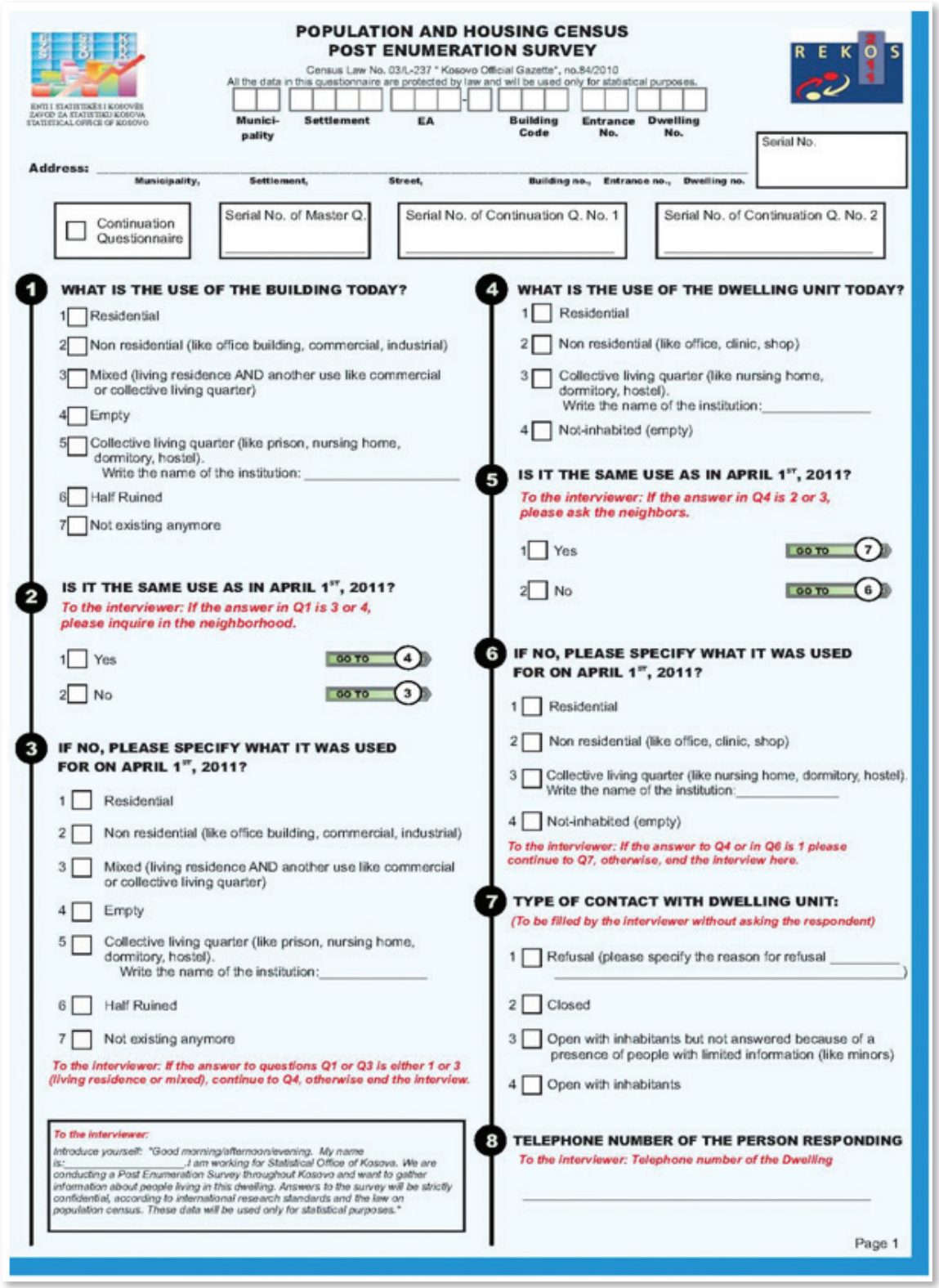
g) PES planning should be an integral part of overall census planning and operations. This should include budget, schedule, data products and other needed resources.

h) PES results should not be treated as the gold standard to evaluate census. PES operations should also be evaluated using various operational statistics such as quality checks during various operations, non-interviews or missing demographic characteristics of households and persons. In addition, PES results should also be examined and evaluated against other aggregate level data sources, if available.

ANNEX 1

The PES Questionnaire

Note:
The PES questionnaires contain a total of 16 pages each, bound together. The part “individual questionnaire” corresponding to the page 4 of the following is repeated 13 times in each PES questionnaire.



POPULATION AND HOUSING CENSUS
POST ENUMERATION SURVEY

Census Law No. 03/L-237 " Kosovo Official Gazette", no.84/2010
All the data in this questionnaire are protected by law and will be used only for statistical purposes.

Municipality | Settlement | EA | Building Code | Entrance No. | Dwelling No.

Address: Municipality, Settlement, Street, Building no., Entrance no., Dwelling no.

Serial No.

☐ Continuation Questionnaire

Serial No. of Master Q.

Serial No. of Continuation Q. No. 1

Serial No. of Continuation Q. No. 2

1 WHAT IS THE USE OF THE BUILDING TODAY?

1 ☐ Residential
2 ☐ Non residential (like office building, commercial, industrial)
3 ☐ Mixed (living residence AND another use like commercial or collective living quarter)
4 ☐ Empty
5 ☐ Collective living quarter (like prison, nursing home, dormitory, hostel). Write the name of the institution: _____
6 ☐ Half Ruined
7 ☐ Not existing anymore

2 IS IT THE SAME USE AS IN APRIL 1ST, 2011?

To the interviewer: If the answer in Q1 is 3 or 4, please inquire in the neighborhood.

1 ☐ Yes — GO TO 4
2 ☐ No — GO TO 3

3 IF NO, PLEASE SPECIFY WHAT IT WAS USED FOR ON APRIL 1ST, 2011?

1 ☐ Residential
2 ☐ Non residential (like office building, commercial, industrial)
3 ☐ Mixed (living residence AND another use like commercial or collective living quarter)
4 ☐ Empty
5 ☐ Collective living quarter (like prison, nursing home, dormitory, hostel). Write the name of the institution: _____
6 ☐ Half Ruined
7 ☐ Not existing anymore

To the interviewer: If the answer to questions Q1 or Q3 is either 1 or 3 (living residence or mixed), continue to Q4, otherwise end the interview.

To the interviewer:
Introduce yourself: "Good morning/afternoon/evening. My name is: _____. I am working for Statistical Office of Kosova. We are conducting a Post Enumeration Survey throughout Kosovo and want to gather information about people living in this dwelling. Answers to the survey will be strictly confidential, according to international research standards and the law on population census. These data will be used only for statistical purposes."

4 WHAT IS THE USE OF THE DWELLING UNIT TODAY?

1 ☐ Residential
2 ☐ Non residential (like office, clinic, shop)
3 ☐ Collective living quarter (like nursing home, dormitory, hostel). Write the name of the institution: _____
4 ☐ Not-inhabited (empty)

5 IS IT THE SAME USE AS IN APRIL 1ST, 2011?

To the interviewer: If the answer in Q4 is 2 or 3, please ask the neighbors.

1 ☐ Yes — GO TO 7
2 ☐ No — GO TO 6

6 IF NO, PLEASE SPECIFY WHAT IT WAS USED FOR ON APRIL 1ST, 2011?

1 ☐ Residential
2 ☐ Non residential (like office, clinic, shop)
3 ☐ Collective living quarter (like nursing home, dormitory, hostel). Write the name of the institution: _____
4 ☐ Not-inhabited (empty)

To the interviewer: If the answer to Q4 or in Q6 is 1 please continue to Q7, otherwise, end the interview here.

7 TYPE OF CONTACT WITH DWELLING UNIT:

(To be filled by the interviewer without asking the respondent)

1 ☐ Refusal (please specify the reason for refusal _____)
2 ☐ Closed
3 ☐ Open with inhabitants but not answered because of a presence of people with limited information (like minors)
4 ☐ Open with inhabitants

8 TELEPHONE NUMBER OF THE PERSON RESPONDING

To the interviewer: Telephone number of the Dwelling

_____

Page 1

ANNEX 1 - The PES Questionnaire

9 LIST OF PEOPLE RESIDING IN THIS DWELLING UNIT TODAY:

Seq. no.	First name	Last name	Father's name	Did you have a different address on the midnight between March 31st and April 1st, 2011?		Did you have two addresses on the midnight between March 31st and April 1st, 2011?	
01				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
02				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
03				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
04				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
05				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
06				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
07				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
08				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
09				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
10				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
11				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
12				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No
13				1 ☐ Yes	2 ☐ No	1 ☐ Yes	2 ☐ No

To the interviewer: If the number of people usually residing in this dwelling is more than 13 continue to continuation Questionnaire

10 TOTAL NUMBER OF PERSONS RESIDING IN THE DWELLING UNIT ☐☐

11 TO THE INTERVIEWER: WHO ANSWERED QUESTIONS Q9 AND Q10?

1 ☐ Person living in this dwelling. Name and Last name _____

2 ☐ Neighbor. Name and Last name _____

3 ☐ Others. Specify (Name and Last name) _____

12 LIST OF PEOPLE WHO LIVED IN THIS DWELLING UNIT, BUT TRAVELLED TO LIVE ABROAD AFTER APRIL 1ST 2010 AND DID NOT RETURN UNTIL MIDNIGHT BETWEEN MARCH 31ST AND APRIL 1ST 2011 (EXCLUDING VISITS AND VACATIONS):

First name	Last name	Father's name	Gender 1. M 2. F	Date of birth (dd.mm.yyyy)	Ethnicity *(Use the codes from the list)*	Marital Status *(Use the codes from the list)*	Does he/she think to stay abroad over 12 months 1. Yes 2. No
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			
				☐☐☐☐☐☐☐☐			

ETHNICITY

1. Albanian
2. Serb
3. Turkish
4. Bosniak
5. Roma
6. Ashkali
7. Egyptian
8. Goran
9. Other (specify in the box)
10. I prefer not to respond

MARITAL STATUS

1. Never married
2. Married with a marriage certificate
3. Married without a marriage certificate
4. Widowed
5. Divorced

13 TOTAL NUMBER OF PEOPLE IN THE ABOVE LIST ☐☐

Page 2

14 **TO THE INTERVIEWER: WHO ANSWERED QUESTIONS Q12 AND Q13?**

1 ☐ Person living in this dwelling. Name and Last name _____

2 ☐ Neighbor. Name and Last name _____

3 ☐ Others. Specify (Name and Last name) _____

15 **LIST OF PEOPLE WHO LIVED IN THIS DWELLING UNIT AT MIDNIGHT BETWEEN MARCH 31ST AND APRIL 1ST, 2011 AND ARE NOT LIVING HERE TODAY (INCLUDING PERSONS THAT HAVE PASSED AWAY SINCE 1ST OF APRIL 2011):**

First name	Last name	Father's name	Gender 1. M 2. F	Date of birth (dd.mm.yyyy)	Ethnicity *(Use the codes from the list)*	Marital Status *(Use the codes from the list)*	Reason for not living here today 1-Moved out – in Kosovo 2-Moved out abroad 3-Deceased

16 **SUM OF PEOPLE IN THE ABOVE LIST** ☐☐

17 **TO THE INTERVIEWER: WHO ANSWERED QUESTIONS Q15 AND Q16?**

1 ☐ Person living in this dwelling. Name and Last name _____

2 ☐ Neighbor. Name and Last name _____

3 ☐ Others. Specify (Name and Last name) _____

COMMENTS OF THE INTERVIEWER:

_____

_____

_____

_____

_____

_____

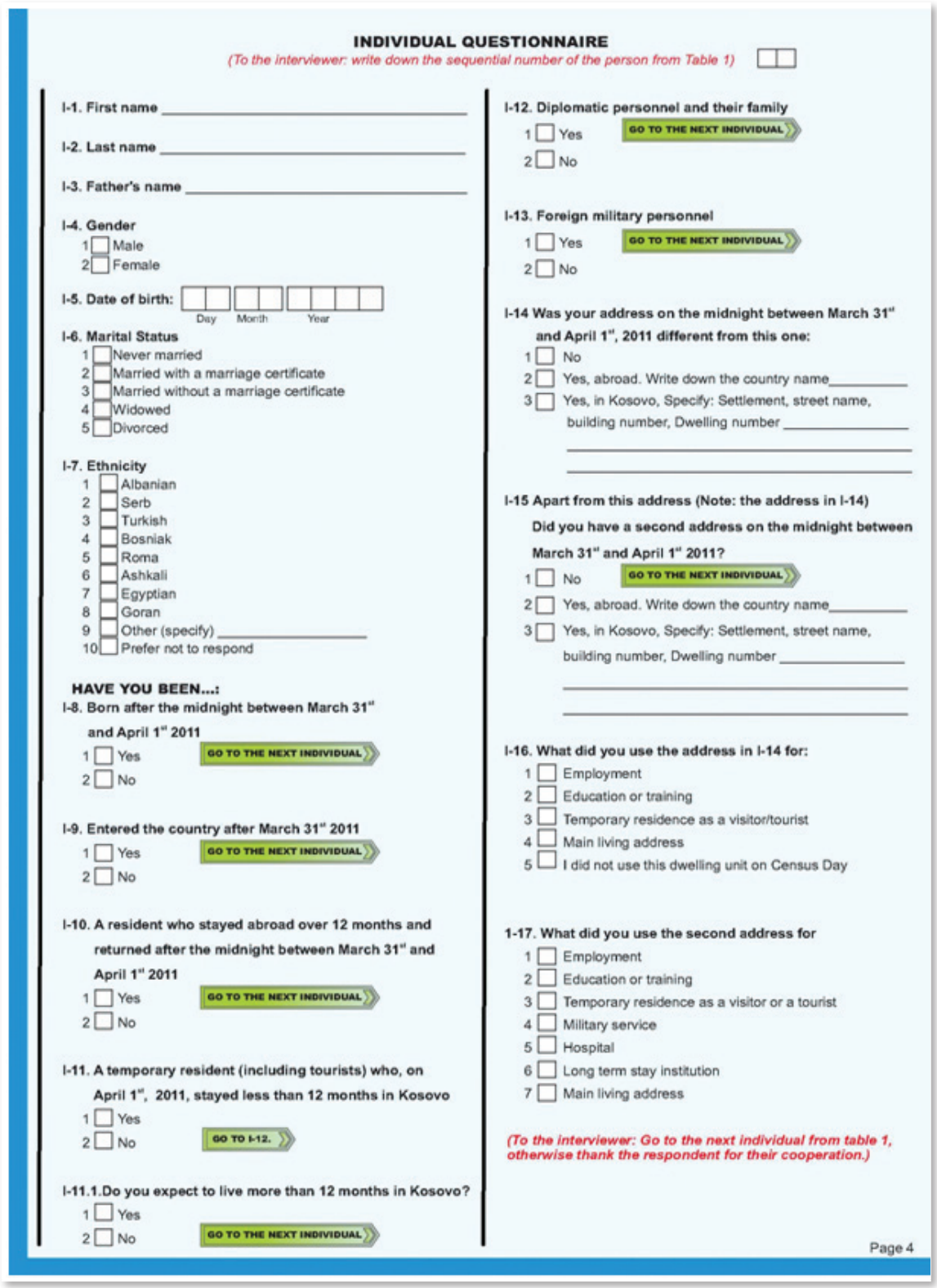
_____

_____

_____

Page 3

ANNEX 1 - The PES Questionnaire



INDIVIDUAL QUESTIONNAIRE

(To the interviewer: write down the sequential number of the person from Table 1) ☐☐

I-1. First name _____

I-2. Last name _____

I-3. Father's name _____

I-4. Gender
1 ☐ Male
2 ☐ Female

I-5. Date of birth: ☐☐ ☐☐ ☐☐☐☐
Day Month Year

I-6. Marital Status
1 ☐ Never married
2 ☐ Married with a marriage certificate
3 ☐ Married without a marriage certificate
4 ☐ Widowed
5 ☐ Divorced

I-7. Ethnicity
1 ☐ Albanian
2 ☐ Serb
3 ☐ Turkish
4 ☐ Bosniak
5 ☐ Roma
6 ☐ Ashkali
7 ☐ Egyptian
8 ☐ Goran
9 ☐ Other (specify) _____
10 ☐ Prefer not to respond

HAVE YOU BEEN...:

I-8. Born after the midnight between March 31st and April 1st 2011
1 ☐ Yes → GO TO THE NEXT INDIVIDUAL
2 ☐ No

I-9. Entered the country after March 31st 2011
1 ☐ Yes → GO TO THE NEXT INDIVIDUAL
2 ☐ No

I-10. A resident who stayed abroad over 12 months and returned after the midnight between March 31st and April 1st 2011
1 ☐ Yes → GO TO THE NEXT INDIVIDUAL
2 ☐ No

I-11. A temporary resident (including tourists) who, on April 1st, 2011, stayed less than 12 months in Kosovo
1 ☐ Yes
2 ☐ No → GO TO I-12.

I-11.1.Do you expect to live more than 12 months in Kosovo?
1 ☐ Yes
2 ☐ No → GO TO THE NEXT INDIVIDUAL

I-12. Diplomatic personnel and their family
1 ☐ Yes → GO TO THE NEXT INDIVIDUAL
2 ☐ No

I-13. Foreign military personnel
1 ☐ Yes → GO TO THE NEXT INDIVIDUAL
2 ☐ No

I-14 Was your address on the midnight between March 31st and April 1st, 2011 different from this one:
1 ☐ No
2 ☐ Yes, abroad. Write down the country name_____
3 ☐ Yes, in Kosovo, Specify: Settlement, street name, building number, Dwelling number _____

I-15 Apart from this address (Note: the address in I-14) Did you have a second address on the midnight between March 31st and April 1st 2011?
1 ☐ No → GO TO THE NEXT INDIVIDUAL
2 ☐ Yes, abroad. Write down the country name_____
3 ☐ Yes, in Kosovo, Specify: Settlement, street name, building number, Dwelling number _____

I-16. What did you use the address in I-14 for:
1 ☐ Employment
2 ☐ Education or training
3 ☐ Temporary residence as a visitor/tourist
4 ☐ Main living address
5 ☐ I did not use this dwelling unit on Census Day

1-17. What did you use the second address for
1 ☐ Employment
2 ☐ Education or training
3 ☐ Temporary residence as a visitor or a tourist
4 ☐ Military service
5 ☐ Hospital
6 ☐ Long term stay institution
7 ☐ Main living address

(To the interviewer: Go to the next individual from table 1, otherwise thank the respondent for their cooperation.)

Page 4

APPENDIX 1

Post Enumeration Survey Estimation Methodology

This appendix describes the various steps used to estimate the coverage errors for PES. The methodology described in (Census 2004) was adopted for Kosovo PES with modifications, as necessary.

Base Weights

The sample of EA was selected using a probability sample. The probability of selection for each EA in the sample was provided. The base sampling weight, W_i represents the weight of household *i*, is the inverse of its probability of selection of each EA since all household were in sample in a selected EA.

Non-Interview Adjustments

Below are discussed non-interviews and adjustment procedures to deal with them.

Unit non-response

Unit non-response occurred when sample households that were not interviewed in the PES because the interviewee (household respondent) declined to respond. A unit (household) was defined as an interview (for the given reference day – Census Day or Interview Day) if there was at least one person with sufficient information for matching and who possibly or definitely was a resident of the dwelling unit on the given reference day. The PES day response is primarily used to apply procedure C in DSE estimation since it provides a better estimate of number of matched person (households).

Non-Interview Adjustment Procedures

Non-response adjustment was performed only on the P sample. Vacant or demolished Dwelling units were deleted from the list of PES sample. The households from these deleted dwelling units did not factor into the census or interview day non-interview adjustment. The non-interview adjustment based on census day was used to adjust the weights of person non-movers and person out-movers. The non-interview adjustment based on PES interview day was used to estimate the number of in-movers.

Census day non-interview adjustment factor - $f_c^{\bullet}$ is defined as

$$f_c^{\bullet} = \frac{\Sigma\ Census\ day\ interviews^{\ wi} + \Sigma\ Census\ day\ non\text{-}interviews^{\ wi}}{\Sigma\ Census\ day\ interviews^{\ wi}}$$

PES interview day non-interview adjustment factor -

$$f_{\hat{p}}^{\bullet} = \frac{\Sigma\ PES\ day\ interviews^{\ wi} + \Sigma\ PES\ day\ non\text{-}interviews^{\ wi}}{\Sigma\ PES\ day\ interviews^{\ wi}}$$

Where W_i represents the weight of household *i,* i.e. the inverse of its probability of selection of a household.

In practice, when a person responded in Kosovo PES, he or she answered all PES questions. Due to complete responses to the PES questionnaires in responded households, the non-interview adjustments for both reference periods (census and PES day) were the same.

Characteristic (Person level) Imputation

Due to complete responses to the PES questionnaires, person has the sufficient information for matching and forming post-strata for estimation. Therefore, P- and E- sample data did not have any missing demographic characteristics for PES estimation. Thus, imputation was not needed for PES estimation.

Probability for Unresolved Cases

Even after reconciliation visit, the sufficient data was not available to determine residence, match or correct enumeration status for some individuals. Hence, it was necessary to compute probabilities of Persons with the following unresolved status:

- Unresolved census day residence status (P-Sample),
- Unresolved Final match status (P-Sample) or
- Unresolved Final correct enumeration status (E-Sample)

These probabilities ideally should have been estimated by calculating weighted ratios based on persons with resolved final status after field reconciliation within each post-stratum. However, due to small sample in a post-stratum or no resolved cases after reconciliation visit, the probabilities were computed based on the pre-reconciliation data. The probabilities were to be estimated for about 50 and 99 out of about 6,000 persons in the entire P- and E-sample, respectively and its affect on estimates was expected to be small.

The non interview adjustment factor was not incorporated into the person weights when calculating these probabilities. This is because the non-interview adjustment is designed to spread the weight of non-interviewed households over the interviewed households.

APPENDIX 1 - Post Enumeration Survey Estimation Methodology

Unresolved Census Day Residence Status in the P Sample

All persons in the P-Sample who were eligible to be matched to the census (person non-movers and person out-movers) were classified into three types according to their resident status in their sampled EA at the time of Census: Census day residents, Census day non-residents, and unresolved persons

All the census day resident person non-movers and out-movers were included with census day resident probability of 1 while the non-resident person non-movers and out-movers were included with census day resident probability of 0.

Assigning the Probability of being a census day resident for unresolved persons

Probabilities for persons with unresolved final census day residence status were estimated by calculating weighted ratios based on persons with resolved final status.

The residence probability (P_r) for unresolved persons needing follow-up was calculated using the following formula;

$$P_r = \frac{Resident\ persons\ before\ reconciliation}{Persons\ \ with\ resolved\ census\ day\ resident\ status\ before\ reconciliation}$$

Unresolved Match Status (P sample)

Some person non-movers and person out-movers had unresolved match status. Computing the dual system estimator required measuring the total number of P-sample people who were matched to persons included in the census. After follow-up activities were completed, each confirmed or possible (unresolved) Census Day resident in the P sample was determined to fall into one of the following three categories; Match, Non-match or Unresolved (that is, persons for whom match status could not be determined after field follow-up).

For each confirmed census day resident, the persons whose match status was a match were assigned a census day resident probability of 1 while those residents whose final match status was a non-match were assigned a census day resident probability of zero (0).

Assigning the Probability of being a match for unresolved persons

The match probability for the persons with unresolved match status is the proportion of matches among person non-movers and person out-movers with resolved final match status (excluding confirmed census day non-residents)

Formula;

$$P_m = \frac{Mached\ persons\ before\ reconciliation}{Persons\ \ with\ resolved\ match\ status\ persons\ before\ reconciliation}$$

Unresolved Enumeration Status (E-sample)

The dual estimator requires the total number of correct enumeration in E-sample. After follow up, each person in the E-sample was assigned one of the following enumeration statuses; correct enumeration, erroneous enumeration, and unresolved enumeration.

The probability of correct enumeration for those determined to be correctly enumerated was one (1) while that of those determined to be erroneously enumerated was zero (0)

Assigning the Probability of being a Correct Enumeration for Unresolved Persons

For E-sample, the persons with unresolved enumeration status, the correct enumeration probability is the proportion of correct enumerations (among persons with resolved enumeration status).

$$P_{ce} = \frac{Persons\ correctly\ enumerated\ before\ reconciliation}{Persons\ \ with\ resolved\ enumerated\ status\ before\ reconciliation}$$

Dual System Estimation (DSE)

PES used the Dual System Estimation methodology which is based on capture and recapture methodology (Chandrasekaran-Deming estimator (1949)). The methodology estimates the total population. The DSE model is conceptualize that each person has a probability of being either included in the census or not included in the census as well as either included or not included in PES. This can be described as in the following table.

DSE Model

	In Census	Not in Census	Total
In PES	N_{11}	N_{12}	N_{1+}
Not in PES	N_{21}	N_{22}	N_{2+}
Total	N_{+1}	N_{+2}	N_{++}

In the above table, all the cells are observable except N_{22} and all the marginal that included N_{22}. The model assumes independence between the census and the PES. Hence the probability of being in ij^{th} cell Pij is the product of two marginal probabilities that contains ij^{th} cell. Thus, under the independence assumption, the estimate of total population is

$$DSE = N_{++} = \frac{(N_{+1})\,(N_{1+})}{N_{11}}$$

Thus the total population can be written as function of number included in census, in PES and those included in both. This model is applied within each estimation domain usually called post-stratum.

APPENDIX 1 - Post Enumeration Survey Estimation Methodology

In practice, the components of DSE are estimated from samples. In the table above, $N+1$ is not the census counts. Thus, the census counts should be corrected for erroneous enumeration. Also, the persons with insufficient information for matching could not be matched with PES enumeration. Hence, census counts must be corrected for enumeration with insufficient information to match with PES enumeration. Thus, DSE uses the following formula:

$$DSE = DD * \frac{CE}{N_e} * \frac{N_p}{M} = DD * \frac{CE}{N_e} \left[\frac{N_n + N_i}{(M_n + (\frac{M_e}{N_e}) N_i)} \right]$$

Where;

DD = the number of census data-defined persons eligible and available for PES matching,

CE = the estimated number of correct enumerations from the E-sample,

$$= \sum_{j \varepsilon E\ Semple} P\ ce, j\ W_j^{\blacklozenge}$$

N_e = the estimated number of people from the E sample

$$= \sum_{j \varepsilon E\ Semple} W_j^{\blacklozenge}$$

N_p = the estimated total population from the P sample,

$$= Nonmovers\ (Nn) + Inmovers\ (N_i)$$

N_n = Weighted total population for non-movers for the post-stratum from the P-sample. The weight for each person j is the product of three values:

$$= \sum_{j \varepsilon Nonmovers} f_{c,j}^{\blacklozenge} P_{r,j} W_j^{\blacklozenge}$$

M_n = Estimated number of p-sample non-mover matches

$$= \sum_{j \varepsilon Nonmovers} P_{m,j} F_{c,j}^{\blacklozenge} P_{rj} W_j^{\blacklozenge}$$

N_i = Weighted total population for in-movers for the post-stratum from the P-Sample. The weight for each person j is the product of two values:

$$= \sum_{j \varepsilon Inmovers} F_{p,j}^{\blacklozenge} W_j^{\blacklozenge}$$

N_o = Estimated number of P-sample out-movers

$$= \sum_{j \varepsilon \, Outmovers} f^{\blacklozenge}_{c,j} P_{r,j} W^{\blacklozenge}_{j}$$

M_o = Estimated number of P-sample out-mover matches

$$= \sum_{j \varepsilon \, Outmovers} P_{m,j} f^{\blacklozenge}_{c,j} P_{r,j} W^{\blacklozenge}_{j}$$

Where;

P_{ce} = probability of correct enumeration for person j
$W^{\blacklozenge}_{j}$ = Adjusted weight. Inverse of the P-sample selection probability
$f^{\blacklozenge}_{c,j}$ = non-interview adjustment based on Census Day interview status for person j
$P_{r,j}$ = probability of Census Day residence for person j.
$f^{\blacklozenge}_{p,j}$ = non-interview adjustment based on PES interview day status for person j
$P_{m,j}$ = probability of a match for person j

Coverage Correction factor

The coverage correction factor (CCF) is a measure of the net over-count or net under-count of the household population within the census. The coverage correction factor for a post-stratum was calculated by dividing the DSE for that post-stratum by its census count.

$$CCF = \frac{DSE}{C}$$

Where; C = Census household population count.

Computation of Estimates

All the estimates were calculated at the following levels; National, Region, and Urban and Rural at national level, and all 38 post-strata levels.

Census Population estimate

The E-sample estimate of the population enumerated in the census is the sum of the matched population, the population erroneously included in the census, the population correctly enumerated in the census but missed in the PES, and the census insufficient-information cases.

$$Cesus\ population = Matched\ population + correctly\ Enumerated + erroneously\ enumerated$$

$$= \sum_{j \varepsilon \, E\ Semple} W^{\blacklozenge}_{j}$$

APPENDIX 1 - Post Enumeration Survey Estimation Methodology

PES Population Estimate

The P-sample estimate of the total population is the sum of the non-movers and in-movers in the population.

$$PES\ Population = nonmovers + inmovers$$

True Population

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. The true population, also known as DSE, is calculated for each post-stratum.

$$True\ Population = DD * \frac{CE}{N_e} \left[\frac{N_n + N_i}{(M_n + (\frac{M_o}{N_o}) N_i)} \right]$$

Coverage Errors

Coverage error refers to either an under-count or over-count of units owing to omissions of persons/ dwelling units or duplication/erroneous inclusion, respectively.

Net Census Under-count Error

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

$$Net\ coverage\ error = True\ population - Census\ count$$

Net Census Under-count Rate

This 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.

$$Net\ coverage\ error\ rate = \frac{True\ population - Census\ count}{True\ population\ estimate} * 100$$

$$= \frac{Net\ census\ cover\ error}{True\ population\ estimate} * 100$$

Census Omission

The formula used to estimate the census omissions is;

$$Census\ Omissions = True\ population - Correct\ Enumerations$$

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

$$Omissions\ rate = \frac{Census\ Omissions}{True\ population\ estimate} * 100$$

Census Erroneous Inclusion

The erroneous inclusions include fabrications and duplications, and geographic misallocation. The erroneous inclusions are computed as:

$$Erroneous\ inclusions = (Census\ count + Census\ Omissions) - True\ Population$$

Census Erroneous Inclusion Rate

$$Erroneous\ inclusion\ rate = \frac{Erroneous\ inclusions}{True\ population\ estimate} * 100$$

Coverage Correction factor

The coverage correction factor (CCF) is a measure of the net over-count or net under-count of the household population within the census. The coverage correction factor for a post-stratum was calculated by dividing the DSE for that post-stratum by its census count.

$$CCF = \frac{DSE}{C}$$

Where; C = Census household population count.

Estimate of DSE for A Population Group

DSE_j for a group j that can be formed by combining post-strata is given by

$$DSE_j = \sum_{i}^{s} DSE_{ij}$$

Where summation I is over all those strata that include the desired group.

DSE_j for any combination of cross-section of post-strata is given by

$$DSE_j = \sum_{i}^{s} Correction\ factor_j\ (C_{ij})$$

Where C_{ij} is the census count for j^{th} group in i^{th} post-stratum, and correction factor is defined as .

APPENDIX 2

Variance Estimation Methodology

The Jackknife variance estimator was used for 2011 Kosovo Census PES. The method is briefly explained below. A short discussion of how to implement the methodology is also given below. For further discussion of the methodology, see Wolter (1985)

The Jackknife variance estimator obtains estimates of interest from each of many subsamples of the population and then computes variance from the variability between the subsample estimates. For simplicity, one could consider that a sample of n units is already selected. Thus, n may be the number of block clusters, housing units, households or EAs with housing units. Instead of selecting many new subsamples, one could use the existing sample and select a subsample of the population by dropping one element or a group of elements (say EA) from the existing sample.

Let X_{ik} be the estimate of interest by excluding k^{th} element (group or ED) where k=1,2,3, ..., n. Then, the standard error for an estimate is given by:

$$Standart\ Error\ (X_i) = \sqrt{\frac{n-1}{n} \sum_{i=1}^{n} (X_{ik} - \bar{X}_{1k})^2}$$

Where n is the total number of subsamples.

The estimate (statistic) X_i was defined as dual system estimate (DSE), net error, erroneous inclusions, or any other estimate of interest that was computed for PES analysis.

There is no reweighing or recalculation of correct enumeration or match probabilities. The mean (average) estimate could be used based on the entire population in estimation domain. For example, average (mean) DSE could be based on the entire sample in an estimation domain. This estimate of DSE has already been computed. The methodology to compute variances at a different geographical level or for a subpopulation group is the same and was repeated at the desired geographic level.

Because of the data limitations, the variances were computed only for the estimates at the national level. The estimates for the urban/rural level at the national level were also obtained but expected to be less stable. Variances or confidence intervals for all other estimates in the report were not computed. The average of X_{ik} was used for the entire sample in the post-strata to simplify the calculations with little or no loss in accuracy of the estimate.

APPENDIX 3

Estimation of Content Error Indices

Content error measures the discrepancy between the census and the PES data. Content error is only estimated for the matched persons and for the selected variables in order to measure the inconsistency between the data captured answers from census and PES to the same questions. It is important to know the inconsistency between data from the two sources (census and PES) for those items that are used for PES matching operations and forming post-strata for estimation. The inconsistencies in these items lead to bias in DSE results. Ideally, the data from the two sources should be the same. However, this is not the case is practice.

This inconsistency is measured by means of four indicators: the net difference rate; index of inconsistency (simple and aggregated), the gross difference rate; and the rate of agreement [for details, see UNSD (2010)]. It is also desirable to review bivariate table to determine item grouping for matching and estimating coverage errors.

The content error estimation and analysis was done at the national level for sex, age, ethnicity, and marital status. The indices computed for content errors evaluation are described below.

Net Difference Rate (NDR)

The NDR approximates the level of difference in captured PES and Census data for each response in the Census and the PES relative to the total number of matched persons in all response categories. It can be interpreted as a measure of the bias in census data only if the PES is considered to be truth. The NDR is calculated using the formula:

$$NDRi = \frac{Y_{.i} - Y_{i.}}{n} x\ 100 \qquad \text{for } i = 1,2,\ldots\ldots C$$

where:

$Y_{.i}$ = unweighted census number of cases in the i^{th} category

$Y_{i.}$ = unweighted PES number of cases in the i^{th} category

n = unweighted number of matched cases

c = total number of response categories for characteristics for "y". For example, sex has two categories – male, and female

APPENDIX 3 - Estimation of Content Error Indices

Index of Inconsistency

The index of Inconsistency is the ratio of the Simple Response Variance (SRV) to the total variance for a given item. It is relative number of cases for which the response varied between the census and the PES. It is computed for each response category *i* according to the following formula:

$$\hat{I}i = \frac{\left(Y_{.i} + Y_{i.} - 2Y_{ii}\right)}{\frac{1}{n}\left[Y_{.i}\left(n - Yi.\right) + Y_{i.}\left(n - Y_{.i}\right)\right]} x\ 100 \quad \text{for } i = 1,2,\ldots\ldots C$$

where:

Yii = number of cases where category i was given as response in both the census and the PES

Aggregate Index of Inconsistency (AII)

The Aggregate Index of Inconsistency (AII) is a summary measure of the index of Inconsistency (that is for all the response categories of the characteristic as a whole). The computation formula is as follows:

$$\hat{I}_{AG} = \frac{\left[n - \sum_{i}^{C} Y_{ii}\right]}{\left[n - \frac{1}{n}\sum_{i}^{C} Y_{.i}Y_{i.}\right]} x\ 100$$

Gross Difference Rate

The Gross difference rate (GDR) measures the percentage of responses reported/recorded differently to questions asked by Census enumeration and PES. The formula used is:

$$GDR = \frac{\left[n - \sum_{i}^{C} Y_{ii}\right]}{n}$$

Rate of agreement

Rate of agreement (RA) is the complement of the gross difference rate (GDR).The rate of agreement indicates the level at which the information given in the Census matches that given during the PES. A low rate of agreement indicates a high degree of variability and vice-versa. The rate of agreement is therefore a good measure of the gross error for an item.

The Rate of Agreement (RA) is given by the following formula:

$$RA = \frac{\sum_{i}^{C} Y_{ii}}{n}$$

Interpretation of Content Error Indices

The content error indices can be analyzed using the criteria presented in the table below.

Interpretation of Indices for Content Errors

Interpretation of Indices for Content Errors			
	Level		
Measure	**Low**	**Medium**	**High**
Index of inconsistency	<20	20 - 50	>50
Aggregate index of inconsistency	<20	20 - 50	>50
Absolute Value of NDR	<0.01	0.01 -0.05	>0.05

Source: UNSD (2010), Post Enumeration Survey, Operational Guidelines

APPENDIX 4

Additional Tables from Post Enumeration Survey

This appendix presents additional tables to provide more information on PES data results. The coverage errors in these tables were briefly discussed in the main report. Tables in this appendix give additional information. Due to limited sample size and high variability, these results can only be used as guidance for future usages. However, one should exercise caution about making definite conclusions about the coverage errors based on these tables.

Content errors were evaluated only at the national levels and had sufficient data to analyze them. The content error tables provided here covers the characteristics that have about the same content errors indices as those discussed in the main report and were not discussed in detail.

Additional coverage error tables are given below in this appendix.

Table A 4.1.: Omissions in Census by Post-strata

Omission (Persons)					
Region	Urban/ Rural	Albanian			Other
		Age group			
		0-19	20-45	46+	Empty
Gjakovë/Đakovica	Urban	2780	1487	619	Empty
	Rural	3433	985	582	Empty
Gjilan/Gnjilane	Urban	1469	1622	768	105
	Rural	Empty	Empty	Empty	3206
Mitrovicë/Mitrovica	Rural	1074	0	0	5560
Pejë/Peć	Rural	1129	2887	391	1155
Prizren	Rural	0	1392	0	949
Prishtinë/Priština	Urban	1504	6727	1637	29
	Rural	0	1060	4426	2958
Ferizaj/Uroševac	Urban	571	667	834	Empty
	Rural	1784	4949	791	6375

NOTE: Prishtinë/Priština corresponds to both the capital city and the Prishtinë/Priština region. The urban corresponds to the capital city and the rural corresponds to the region outside the capital city.

Table A 4.2: Erroneous Inclusions in Census by Post-strata

Erroneous Inclusion (Persons)					
Region	Urban/ Rural	Albanian			Other
		Age group			
		0-19	20-45	46+	Empty
Gjakovë/Đakovica	Urban	0	0	269	Empty
	Rural	1093	5167	502	Empty
Gjilan/Gnjilane	Urban	131	285	107	176
	Rural	Empty	Empty	Empty	0
Mitrovicë/Mitrovica	Rural	0	1013	506	484
Pejë/Peć	Rural	693	1967	1467	0
Prizren	Rural	0	678	0	2225
Prishtinë/Priština	Urban	224	608	3	271
	Rural	1302	1332	0	338
Ferizaj/Uroševac	Urban	228	233	142	Empty
	Rural	0	177	0	265

NOTE: Prishtinë/Priština corresponds to both the capital city and the Prishtinë/Priština region. The urban corresponds to the capital city and the rural corresponds to the region outside the capital city.

Content Error Tables

The additional content error tables are given below.

The content error indices can be analyzed using the criteria presented in the table below.

Table A 4.3 Interpretation of Indices for Content Errors

Interpretation of Indices for Content Errors			
	Level		
Measure	Low	Medium	High
Index of inconsistency	<20	20 - 50	>50
Aggregate index of inconsistency	<20	20 - 50	>50
Absolute Value of NDR	<0.01	0.01 -0.05	>0.05

Source: UNSD (2010), Post Enumeration Survey, Operational Guidelines

APPENDIX 4 - Additional Tables from Post Enumeration Survey

TABLE A 4.4 P-and E-sample Persons' Response Captured Data by Gender

E-sample	P sample			
	Gender	Male	Female	Total
	Male	2760	64	**2824**
	Female	67	2756	**2823**
	Total	**2827**	**2820**	**5647**

TABLE A 4.5 Content Error Indices for Gender

	NDR	Index of inconsistency
Male	-0.000531256	4.639632681
Female	0.000531256	4.639632681
Aggregate Inconsistency Index	4.639632681	
Gross difference rate	0.023198158	
Rate of agreement	97.68018417	

TABLE A 4.6 P-and E-sample Persons' Response Captured Data by Marital Status

E sample	P sample						
	Marital status	Never married	Married with certificate	Married without certificate	Widowed	Divorced	**Total**
	Never married	2676	60	6	8	1	**2751**
	Married with certificate	76	2133	68	27	2	**2306**
	Married without certificate	10	75	196	1	1	**283**
	Widowed	10	73	4	127	2	**216**
	Divorced	3	2	3	2	13	**23**
	Total	**2775**	**2343**	**277**	**165**	**19**	**5579**

TABLE A 4.7 Content Error Indices for Marital Status

	NDR	Index of inconsistency
Never married	-0.004301846	6.238124541
Married with certificate	-0.006632013	14.12185233
Married without certificate	0.001075462	31.58501317
Widowed	0.009141423	34.48992099
Divorced	0.000716974	38.23786433
Aggregate Inconsistency Index	13.47120446	
Gross difference rate	0.077791719	
Rate of agreement	92.22082811	

TABLE A 4.8 P- and E-sample Persons' Response Captured Data by Ethnic Group

		P sample		
		Albanian	Others	Total
E sample	Albanian	4881	17	4898
	Others	25	720	745
	Total	4906	737	5643

TABLE A 4.9 Content Error Indices for Ethnic Groups

	NDR	Index of inconsistency
Albanian	-0.001417686	3.2623903
Other	0.001417686	3.2623903
Aggregate Inconsistency Index=	3.2623903	
Gross difference rate=	0.00744285	
Rate of agreement=	99.25571505	

BIBLIOGRAPHY

CHANDRASEKARAN, Chidambara and DEMING, Edwards (1949); "On a method of estimating birth and death rates and the extent of registration." *Journal of the American Statistical Association*, 44(245): 101-115

UNSD (2010); "Post Enumeration Survey, Operational Guidelines"

U.S. CENSUS BUREAU (2004); "Census 2000, Accuracy and Coverage Evaluation of Census 2000: Design and Methodology"

WOLTER, Kirk (1985); "Introduction to Variance Estimation", *New York: Spinger Verlag*

WOLTER, Kirk (1986); "Some coverage Error Models for Census Data", *Journal of American Statistical Association,* 81, 338-346