CENSUS 2011

Post-enumeration survey

Results and methodology

THE SOUTH AFRICA I KNOW, THE HOME I UNDERSTAND

Statistics South Africa

De Bruyn Park Building, 170 Thabo Sehume Street, Pretoria, 0002

Private Bag X44, Pretoria, 0001, South Africa

User information service: +27(12) 310 8600, Fax: +27(12) 310 8500

Main switchboard: +27(12) 310 8911, Fax: +27(12) 321 7381

Website: www.statssa.gov.za, Email: info@statssa.gov.za

Post-enumeration Survey (PES)

Statistics South Africa

Pali Lehohla
Statistician-General

Report No. 03-01-46

Post-enumeration Survey (PES) / Statistics South Africa

Published by Statistics South Africa, Private Bag X44, Pretoria 0001

© Statistics South Africa, 2012
Users may apply or process this data, provided Statistics South Africa (Stats SA) is acknowledged as the original source of the data; that it is specified that the application and/or analysis is the result of the user's independent processing of the data; and that neither the basic data nor any reprocessed version or application thereof may be sold or offered for sale in any form whatsoever without prior permission from Stats SA.

Stats SA Library Cataloguing-in-Publication (CIP) Data
Post-enumeration Survey (PES) / Statistics South Africa. Pretoria: Statistics South Africa, 2012

Report no. 03-01-46
84 pp

ISBN 978-0-621-41394-6

A complete set of Stats SA publications is available at Stats SA Library and the following libraries:

National Library of South Africa, Pretoria Division
National Library of South Africa, Cape Town Division
Library of Parliament, Cape Town
Bloemfontein Public Library
Natal Society Library, Pietermaritzburg
Johannesburg Public Library
Eastern Cape Library Services, King William's Town
Central Regional Library, Polokwane
Central Reference Library, Nelspruit
Central Reference Collection, Kimberley
Central Reference Library, Mmabatho

This publication is available on the Stats SA website: www.statssa.gov.za

For **technical enquiries** please contact:

Nathi Gama
Tel: 012 310 8620 / 8394
Email: nathig@statssa.gov.za

For **dissemination enquiries** please contact Printing and Distribution, Statistics South Africa

Ina du Plessis
Email: inadp@statssa.gov.za

EXECUTIVE SUMMARY

Background

Statistics South Africa (Stats SA) conducted the Post-enumeration Survey (PES) for Census 2011 shortly after the completion of census enumeration, from November to December 2011. The main goal of the PES was to evaluate the coverage and content errors of the census. The PES was an integral part of the Census 2011 programme, but institutional arrangements were put in place to maintain the independence of PES operations to ensure that the results of the PES were not biased.

Fieldwork methodologies and procedures for the PES were adapted from census methodology documents since the PES replicates the census (in sampled enumeration areas (EAs)). A sample of 600 EAs was drawn from the census EAs and allocated to the provinces, based on expected standard errors which were based on those obtained in PES 2001. The PES questionnaire was a shorter version of the census questionnaire and comprised only the elements required for measuring coverage. The matching and reconciliation visit exercise commenced after the completion of fieldwork operations. This exercise involved the comparison of household and person records in census data and PES data, and re-visits to households in order to confirm or get more information that would assist in matching unresolved cases.

Evaluation of coverage error

Dual-system estimation was used to evaluate coverage errors and derive the *true population* of the country. The evaluation of coverage errors was restricted to dwelling units (DUs) in EAs that were within the scope of the PES, i.e. it only represents populations in all other EA types excluding student residences, institutions, tourist hotels/motels/inns, homes for the aged (unless they are structured into separate households), and homeless people. The EA type 'Collective Living Quarters', which includes hostels (in-scope for PES) was out of scope during sampling. However, hostels, residential hotels and homes for the aged found in the PES sample EAs were enumerated for the PES. The PES estimated the South African population to be 49,79 million people as at 09–10 October 2012. The uncorrected census count (including erroneous inclusions) was estimated to be 42,51 million, whilst the corrected count (excluding erroneous inclusions) was estimated at 42,08 million. The PES results indicate that Census 2011 omitted 6,29 million persons. The net undercount relative to the 49,79 million in the *true population* is thus 14,6%. The highest undercount was observed in Western Cape with an undercount of 18,5%.

Evaluation of content error

For the measurement of the correctness of responses between census and PES, the rate of agreement, net difference rate and index of inconsistency were used. The variables under consideration were sex, age group, population group, marital status and relationship to head of household. The variable 'sex' had the highest rate of agreement of 0,99 and lowest index of inconsistency of 2,5.

Contents

List of tables

List of figures

1. INTRODUCTION

1.1 Background

South Africa conducted its third post-1994 population census in October 2011. A population census is a total count of the country's population, where demographic, social and economic information is collected about the people who live in the country. The census yields a wealth of valuable information for analysing changes in the socio-demographic profile of the population, and for monitoring, planning, and decision-making at national and local levels, by government, business and the general community. It is also integral to the derivation of reliable post-censal population estimates and for charting future demographic trends.

Given the strategic significance of the census data and its diverse applications, Stats SA makes concerted efforts to ensure the universal coverage of its ten-yearly census. Yet, in such a large and complex exercise, it is inevitable that some people will be missed and some will be counted more than once. Some of the reasons why people may be missed include the following:

- they were travelling and were difficult to contact;
- they mistakenly thought they were counted elsewhere;
- there was insufficient space on the census questionnaire in the household where they were staying and they did not obtain additional questionnaires;
- the person completing the questionnaire thought that, for example, young babies, the elderly or visitors should not be included;
- they did not wish to be included due to concerns about the confidentiality of information or a more general reluctance to participate;
- the dwelling in which they were residing was missed because it was difficult to find (e.g. in a remote or non-residential area);
- the dwelling in which they were residing was mistakenly classified as unoccupied; and
- lost questionnaires or questionnaires that could not be processed.

Some of the reasons why people are counted more than once include the following:

- they were included on the census questionnaire at the dwelling where they usually live, even though they stayed and were counted elsewhere on census night (i.e. failure to apply the *de facto* rule accurately); and
- they were out of the country on census night and so should not have been counted at all, but were included on the census questionnaire at the dwelling where they usually live.

Usually more people are missed than over-counted during a census, so the census count of the population is lower than the *true population*. This difference is called net undercount. Rates of net undercount can vary significantly for different population groups depending on factors such as sex, age and geographic location. Stats SA obtains estimates of net undercount using information collected through a Post-enumeration Survey (PES). A PES is a special survey undertaken shortly after the census to evaluate the completeness of census coverage. It provides a concrete statistical basis for estimating census coverage, that is, the extent of undercount or overcount, and for adjusting the census data if there is evidence of significant coverage error. Stats SA used the PES to evaluate census coverage for Census '96 and Census 2001. Stats SA once again used the PES to evaluate the coverage of Census 2011. The PES was conducted in approximately 600 enumeration areas (EAs) from November to December 2011.

1.2 Objectives of the Post-enumeration Survey

The main goal of the PES was to evaluate the quality of census data by collecting high-quality data immediately after the completion of census enumeration. The specific objectives of the PES project were as follows:

- to evaluate the accuracy of census data by providing quantitative information on coverage and content error at specified levels of estimation; national, provincial and geography type (urban and non-urban);
- to provide stakeholders with quantitative information to enable determination of the overall success of the census programme;
- to identify procedural issues which require improvement in future censuses, particularly if sources and causes of errors are identified;
- to evaluate the quality of census enumeration area (EA) maps as sampling units for inter-census and household-based surveys; and
- to provide statistical basis for adjustments of census data, if and when it becomes necessary.

1.3 Planning of the PES

The PES was an integral part of the Census 2011 programme. For purposes of implementing the PES, which is supposed to be independent from the census, Statistics South Africa's Methodology and Evaluation division was mandated with managing the organisation and fieldwork operations of the PES. Planning included the drafting of strategic and operational plans for the Census 2011 PES. These documents stated the purpose of the PES, timelines of the project, as well as resource requirements (financial, human and otherwise). The schedule of activities for the PES were synchronised with census activities. The determination of timelines also took into account the fact that international best practices recommend that the PES should be conducted within a closed population, i.e. it should be carried out within a few months, preferably within six (6) months, after the completion of census fieldwork to ensure that the impact of natural population changes, such as births, deaths and migration, as well as

respondent memory lapses do not complicate the exercise. This practice is also a key assumption of the dual system of estimation.

Plans and methodology documents for the PES incorporated lessons learned during the Census 2011 PES Test that was conducted in 2009 and the Census 2011 Dress Rehearsal (CDR) PES that was conducted in 2010, which served as pilot studies for refining the methodologies and procedures for evaluating the census. Inputs and recommendations from the United Nations Population Fund (UNFPA) were incorporated into PES methodology. Activities of the PES included:

- sampling: sample design and selection;
- development of data collection methodologies: methods and procedures for data collection (publicity, listing and enumeration), including quality control measures applied during data collection;
- development of matching and reconciliation procedures and systems: guidelines for matching, including rules for determining the match status of households and individuals, as well as computer-based system for capturing household and person records for matching purposes;
- questionnaire development: selection of data items which allowed measurement of coverage and content, including layout, design and printing of questionnaire;
- data collection: publicity, listing and enumeration of households in selected enumeration areas (EAs);
- matching and reconciliation: office matching (comparison) of census and PES household and person records, and revisits to households in order to confirm or get more information that might assist in matching unresolved cases; and
- analysis and reporting: compilation of tables and the write-up of PES results.

1.4 Outline of this report

This report describes and discusses the salient features of the Census 2011 PES, including its scope of coverage, methodology, the information gathered and the key results. Discussion of the survey results also covers implications for post-censal population estimates. This report is arranged into nine (9) sections which provide details on the implementation of the PES. The overall methodology is described in Chapter 2, a description of how fieldwork operations were carried out is given in Chapter 3, and matching reconciliation operations are explained in Chapter 4. Chapter 5 outlines the procedures used for measuring coverage and content errors. Chapter 6 provides an analysis of coverage errors for persons, and Chapter 7 provides analysis of coverage errors for households. Chapter 8 focuses on the analysis of content errors. Concluding remarks and recommendations for future censuses and surveys are provided in Chapter 9.

2. METHODOLOGY

This chapter describes the methodology used as well as the assumptions used for conducting the PES. It provides a summary of the processes followed in designing the sample for the PES and the assumptions used in developing PES methodology documents and instruments.

2.1 Assumptions of the dual system of estimation in PES

A population census is an expensive and massive exercise which inevitably has inaccuracies arising from coverage and content errors. Coverage error is the error in the count of persons or housing units in the form of omissions, erroneous inclusions and duplications due to either lack of cooperation from respondents, misunderstandings, lack of quality control measures or the loss of questionnaires. Content error is an error in recording characteristics of those persons that were enumerated because of erroneous or inconsistent reporting, and/or failure to obtain or record the required data accurately. The PES is one of the methods used for measuring these types of errors in a census.

The PES is an independent survey that replicates the census in sampled enumeration areas (EAs). The records of both the census and PES are then matched (compared item by item) in order to determine census coverage and content errors, and also provide a concrete statistical basis for the adjustment of census data if there is evidence of significant coverage error. The major assumption used in the PES is that the census and the PES are independent, the estimate of the percentage missed by the PES but found by the census and the percentage missed by the census but found by the PES can be used to construct estimates of the percentage missed by both the PES and census. Despite the efforts to maintain independence, the likelihood of a person being missed in the PES may be related to whether they were missed in the census. This may result in a 'correlation bias' in the PES estimates. To minimise this bias, PES estimation takes into account the fact that different groups have a different likelihood of being missed.

Dual-system estimation was used to derive the *true population* of the country. This means that two independent sources or 'systems' were used to arrive at the estimate of the *true population*: the census and the PES. The dual system provides an estimate of the cases included in one source (PES) and excluded from the other (census), and vice versa, as well as the count of those that were enumerated in both sources. It also allows for the computation of the number, as well as the rate, of persons missed by both the census and PES using the principle of independence and probability methods. Both estimates contribute to the dual-system estimate, which is more complete than either the census or the PES estimate alone. In the end, this *true population* is compared with the *census-enumerated population* and the difference is the net *undercount* (or *overcount*). Dual-system estimation requires the following assumptions to apply:

- Closed population: migration between the census and PES is insignificant and the composition of the population remains relatively unchanged.

- There is independence between census and PES, i.e. the organisation of the census and PES, especially fieldwork operations, must be managed by different teams.
- There is absence of erroneous inclusions in either the census or PES.
- There are no incomplete matches. Any failure to match census and PES items should be due to actual omission and not the inability to match.

There are three alternative procedures for evaluating census coverage in a PES. These three procedures are known as A, B, and C (see Annexure A for definition of terms).

Procedure A

- Seeks to reconstruct the households as they existed at the time of the census.
- The respondent must identify all persons in the sample household on the census reference date.
- The aim is to match these persons (non-movers and out-movers) to the census questionnaires, and to estimate the number and percentage of matched non-movers and movers (out-movers).
- The matching of non-movers and out-movers is relatively simple and inexpensive because the search is limited to sample areas, but it is difficult and expensive to locate out-movers, especially out-mover households, given that they are no longer at the sample address (information, when available, is reported by proxy respondents), hence, there is a strong possibility of underestimation of the number of movers (out-movers).
- This leads to underestimation of the census omissions.

Procedure B

- Seeks to identify all persons in the sample household on the reference date of the PES.
- People respond for themselves; hence field enumeration is more complete than in Procedure A.
- The aim is to match these persons (non-movers and in-movers) to the corresponding census records and to estimate the number and percentage of matched non-movers and movers (in-movers).
- It provides a better estimate of the number of movers than Procedure A, but associated difficulties and costs of matching are far greater because it involves searching for in-movers in the area where they were enumerated during the census.
- It is not sure if failure to match means an actually omitted person or an incorrectly located person.
- This leads to overestimation of the census omissions.

Procedure C

- Seeks to identify all persons in the sample household on the reference date of the PES and, in addition, any other persons in the household on the reference date of the census, and to classify each person as either non-mover, out-mover, or in-mover with regard to his household presence status on the census date.
- The aim is to match to the census records only the persons present on the date of the census, that is, the non-movers and the out-movers.
- The estimate of the matched out-movers is based on the estimated matched in-movers.
- Matching rates for movers are estimated based on out-movers (as in Procedure A).

Procedure C is a combination of Procedures A and B, and takes advantage of the features of each to reduce matching difficulties and, at the same time, improves the estimation of movers. For this reason, Procedure C was chosen as the preferred method for analysis for Census 2011 PES.

2.2 Scope of coverage for the PES (target universe)

Census 2011 PES sought to estimate the total number of persons and households in dwelling units on the night of 09–10 October 2011 (census night). The units of observation were the persons who spent the census night and/or the PES night in these living quarters. The PES does not represent people in:

- homes for the aged unless they are structured into separate households;
- student residences;
- tourist hotels/motels/inns;
- institutions; or
- the homeless on the street.

2.3 PES sample design

Since PES methodology requires a two-way match of census and PES records, the principle for selecting a primary sampling unit is that the areas must have boundaries that are well-defined on geographic maps and recognisable on the ground. The boundaries for the areas selected for PES must correspond with the boundaries of census areas to allow for item-by-item comparison between the census and PES records. It is for this reason that the primary sampling units for the PES were selected from census enumeration areas (EAs).

2.3.1 Sampling process

The ultimate purpose of the PES was to evaluate the quality of the census data by providing quantitative information on coverage and content errors. The evaluation focused on the nine (9) provinces in South Africa, namely Western Cape (WC), Eastern Cape (EC), Northern Cape (NC), Free State (FS), KwaZulu-Natal (KZN), North West (NW), Gauteng (GP), Mpumalanga (MP) and Limpopo (LP). The initial sampling frame for the PES was the complete list of Census 2011 enumeration areas (EAs), amounting to 103 576 EAs. However, this frame was reduced to 89 305 EAs after the exclusion of out-of-scope EAs which amounted to 14 271 EAs (13,8% of all the Census 2011 EAs). It was desired to have included the workers' hostels in the PES sample, but the current EA type classifications did not single out workers' hostels. Instead, all institutions, including workers' hostels, etc., were put together to form a type called 'Collective Living Quarters'. Vacant EAs, parks and recreation, industrial, commercial and collective living quarters were excluded from the PES 2011 frame. The out-of-scope EA types are excluded due to the difficulty in tracking movers for matching purposes.

The stratification and sampling process was designed to allow for the provision of estimates at national, provincial, urban and non-urban levels, but estimates will only be reliable at national and provincial levels.

2.3.2 Stratification

To improve the efficiency of the PES sample design, the sampling frame was divided into homogeneous strata. For this purpose, variables correlated with coverage error, such as geographic area, were chosen, since density and accessibility affect the quality of census coverage. In addition, geographic stratification is necessary to obtain separate estimates by domain of analysis. Therefore, the first level of stratification corresponded with the geographic domains of estimation defined; namely province and urban/non-urban zones of residence.

For post-stratification, variables correlated with the extent of coverage, such as subdivisions that were well delimited and possessed a high degree of internal homogeneity with regard to socio-demographic characteristics were used. Hence, the sampling frame of EAs was implicitly stratified by geography type: urban formal, urban informal, tribal area, and rural formal.

2.3.3 Sample size and allocation

A sample of 600 EAs was selected from the total of 89 305 EAs from the Census 2011 EA frame and was allocated to the provinces based on the PES 2001 standard errors (SEs). The provincial samples were further allocated to strata using proportional allocation. International best practices recommend a sample size of 1% of the census EAs should be used for the PES in order to achieve acceptable precision levels. The budget allocated to the Census 2011 PES could only accommodate a maximum sample size of 600 EAs. The maximum SE level of 0,025 for the undercount rate was used in order to obtain a sample size of 600 EAs. This allowed for a reasonable 95% confidence level for the estimate of

the census omission rate for each province. Table 1 shows the sample size in relation to provincial standard errors.

Table 1: Sample size in relation to provincial standard errors

Province	Distribution of EAs in the frame	PES 2011 sample	Estimated SE (s2)
RSA	89 305	600	0,011
WC	9 538	81	0,0129904
EC	15 060	104	0,0166648
NC	2 572	20	0,0113137
FS	5 379	35	0,0128285
KZN	15 342	119	0,051121
NW	6 278	39	0,0134164
GP	18 197	117	0,0326377
MP	6 783	34	0,0119523
LP	10 156	51	0,0047329

2.3.4 Sample selection

The EAs within each province were stratified by urban and non-urban, and were also ordered by geography type within the strata. Explicit stratification would allow analysts to calculate domain estimates that are representative at such level (urban and non-urban). Gauteng and KwaZulu-Natal had significantly higher standard errors; hence, the sample size was increased in order to reduce the estimated standard errors. Within each stratum, units were sorted geographically (using EA code) and EAs were selected using the systematic sampling method to obtain a sample of 600 EAs nationally. Systematic selection was used in order to improve the spread of the EAs and the representativity of the sample. A SAS PROC SURVEYSELECT, which is a procedure used in SAS, was used to select 600 EAs from the Census 2011 EAs.

2.3.5 P sample and E sample

The PES actually involves two samples which are known as the P sample and the E sample. The P sample (or 'population' sample) consists of the PES sample EAs drawn from the same target population, but independently from the census, for the purpose of estimating census omissions when compared to census records. The E sample is the 'enumeration' sample drawn from cases already enumerated in the census, but selected for independent checks for the purpose of estimating census omissions and erroneous inclusions when compared to original census records. Not all census-enumerated cases belong in the E sample – cases that are out of scope for the PES (for example, student residences and institutions) are not included in the E sample. The estimate of erroneous inclusions provides a correction factor needed in the dual system estimate of the *true population*.

Even though theoretically the E sample may be separate from the P sample, in practice, it is better to allow it to overlap completely with the P sample to reduce costs and improve the precision of the estimates. The E sample then consists of the same EAs selected for the PES. A two-way match is conducted between the P sample and the E sample to identify both the omissions and the erroneous inclusions. The matching also produces the estimate of the 'matched population' component required in the dual system estimator.

2.4 Questionnaire development

The approach to questionnaire design focused on capturing the main elements for measuring coverage and content. Only a few elements from the Census 2011 questionnaire which were not likely to change within a short period (that is, between the census and the PES reference nights) were retained. The questionnaire was designed in line with the principles of Procedure C, and allowed for the classification of each listed person as 'non-mover', 'in-mover', 'born after census', 'out-mover', or 'out of scope', with regard to their household presence status on census night (09–10 October 2011) and PES night (06–07 November 2011). The questionnaire was printed in all eleven (11) official languages. A translation booklet was also developed and supplied to the fieldworkers. Printing of the PES questionnaire was completed in October 2011. The following data items were included in the Census 2011 PES questionnaire:

- First name and surname;
- Date of birth;
- Age;
- Sex;
- Relationship to head/acting head of household;
- Marital status as at PES night; and
- Population group.

Further questions to improve the matching process and test coverage of households and persons during Census 2011, as well as movement/migration and birth/death in the period between census data collection and the PES were included in the questionnaire. These questions covered the following:

- Presence of the household in the dwelling unit (DU) with respect to whether the household was already living in DU by census night;
- Enumeration of members of the household during census with respect to whether persons were enumerated in the household and DU;
- Death (reference period being census night) after census night;
- Birth after census night; and
- Migration (reference period being census night and PES night) with respect to a person's presence in the household and DU.

Questions for determining moving status included the following:

P-00	a) Please tell me the name and surname of the head/acting head of the household and of each person who was present in the household at midnight between 6 and 7 November including members who were absent on 6 and 7 November but were working on night shift or travelling back home.
	b) In addition, tell me the names and surnames of any persons who did not spend the night between 6 and 7 November, but who spent the night between 9 and 10 October in this household

- P-02: Presence on PES night
- P-03: Presence on census night
- P-04: Counted
- P-05: Where counted

Moving status for persons listed in the PES questionnaire (based on responses to the above-mentioned questions) was as follows:

1. non-mover (present on PES night, and also on census night)
2. in-mover (present on PES night, but absent on census night)
3. born after census
4. out-mover (absent on PES night, but present on census night)
9. PES out of scope

The PES questionnaire also provided for the collection of a barcode sticker that had been left with the census-enumerated household by the Census Fieldworker. Census Fieldworkers used a pair of PES barcode stickers per household by pasting one of the stickers on the front door or other feature of the dwelling unit and giving a loose sticker to the household to keep for PES Fieldworkers. The barcode stickers carried the same tracking barcode as the census questionnaire used to enumerate the household. PES Fieldworkers were expected to ask respondents if they had a sticker given to them by the Census Fieldworker; and if they had the sticker, the PES Fieldworker would place this sticker directly on the PES questionnaire or transcribe the barcode sticker number onto the PES questionnaire (if the sticker had been pasted on the door or other item in the DU). The barcode sticker was used to aid the matching process by providing a direct link between census and PES questionnaires used to enumerate each household.

2.5 Fieldwork methodology

The PES replicated the census in the sampled EAs, which meant that all methodologies and procedures for data collection were based on census methodologies and procedures. Extra measures were put in place to ensure that the PES conducted a more complete audit of the census, for example, extensive probing was conducted during the listing exercise in order to correctly identify and classify all structures and dwelling units, including the number of households and the number of persons per household. PES fieldwork was split into the following three (3) phases; publicity and listing, enumeration and mop-up operations.

- Publicity and listing were conducted at the same time. Publicity focused on informing and educating respondents and relevant stakeholders about the purpose of the PES to ensure successful coverage of all dwelling units (DUs) in selected EAs. Listing involved the recording of all structures (including all DUs, number of households in DUs and number of persons in households) in the sampled EAs in the EA Summary Books.

- Enumeration involved interviewing respondents and recording responses in the fields provided in the PES questionnaire. Self-enumeration for the PES was discouraged, but was used in instances where the respondent insisted on self-enumeration.

- Mop-up operations were conducted in the form of follow-up visits by senior field staff to households that could not be contacted during the enumeration period.

Detailed instructions on data collection and other fieldwork procedures, e.g. concepts and definitions, lines of communication, quality assurance procedures, etc., were outlined in the manuals provided during training. The following manuals were developed for fieldwork purposes:

- Census 2011 PES Logistics Manual (procedures for distribution and return of materials);
- Census 2011 PES Fieldworker Manual (procedures for publicity, listing and data collection);
- Census 2011 PES Fieldwork Supervisor Manual (procedures for supervision of staff and reporting);
- Census 2011 PES District Survey Coordinator Manual (procedures for management and coordination of the survey at district level and reporting);
- Census 2011 Map Reading Manual (guidelines on how to use maps);
- Census 2011 PES Quality Assurance Manual (procedures on implementation of quality assurance measures);
- Census 2011 PES Trainee Workbook on Publicity and Listing; and
- Census 2011 PES Trainee Workbook on Enumeration.

2.6 Matching and reconciliation methodology

The matching exercise involved the comparison of household and person records in census data and PES data. A two-way case-by-case matching was conducted using the two sources: PES questionnaires and census questionnaires.

Matching plays an integral role in the dual-system methodology:

- It provides an account of the persons included in both sources, and of the persons included in one source and excluded from the other, based on direct observation. (The PES does not simply rely on people reporting that they were or were not enumerated in the census.)
- It also enables the discovery and removal of erroneous inclusions (fabrications, duplications, out of scope, geographic misallocations) in either source.

Reconciliation visits were conducted in order to confirm or get more information that would assist in matching unresolved cases, i.e. households or individuals enumerated in the census that did not correspond with households or individuals enumerated in the PES. Guidelines for matching, including rules for determining the match status of households and individuals, were developed. A computer-assisted manual matching system was developed for the capturing of data for matching purposes. Two automated matching systems were also deployed as additional quality measures on the computer-assisted manual matching exercise. These automated systems were also being tested as future alternatives to the computer-assisted manual matching system.

3. DATA COLLECTION

PES data collection commenced immediately after the completion of census fieldwork. The PES is a much smaller scale operation (and hence easier to control) than the census. These features enable the PES to deliver a more accurate estimate of the percentage of people and dwellings missed by the census. PES data collection (field operations) was independent from census operations and the following measures were taken to maintain the operational independence of the PES:

- independent listing of enumeration areas (EAs) in the PES sample;
- using separate/independent office staff in the PES and census where possible;
- ensuring the PES interviewers were not employed as census field staff in the same area, and vice versa; and
- maintaining the confidentiality of the PES sample so that census field and office staff were not aware which areas are included in the PES.

This chapter summarises the processes that were carried out in preparation for, and during PES fieldwork operations.

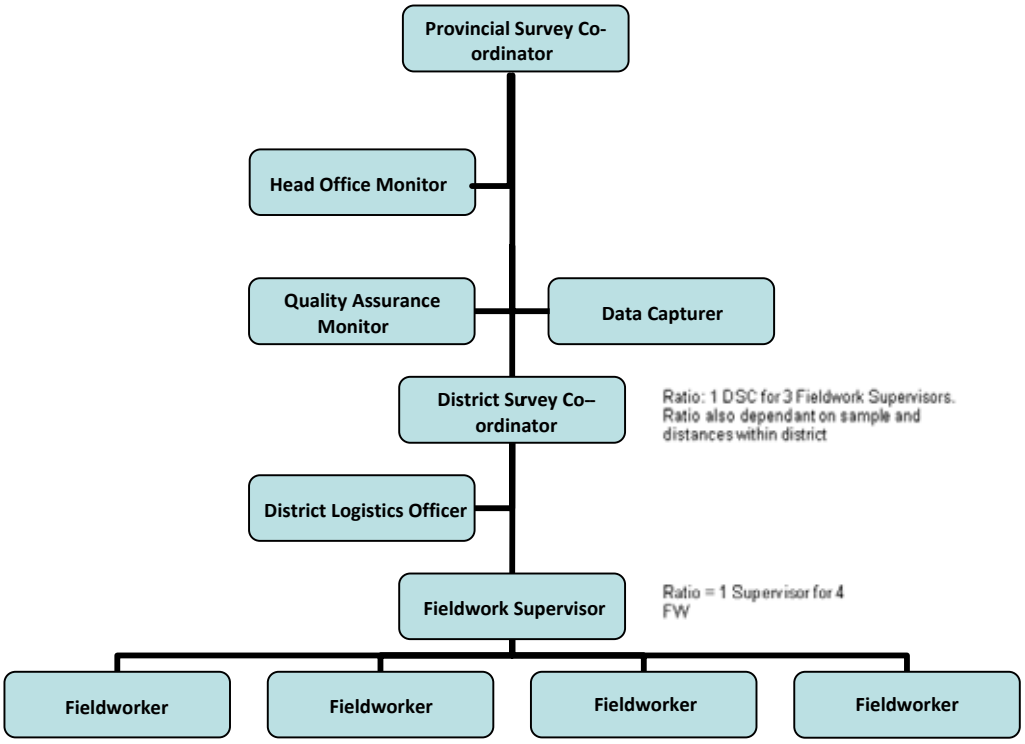
3.1 Logistical arrangements

Logistical arrangements involved the procurement of fieldwork materials (quantities were based on sample size and distribution), which included print material (questionnaires, training manuals and publicity material), field gear, stationery, vehicles and other resources. Once these materials had been received from service providers, they were stored and packaged into consignments in preparation for dispatch to the provinces. Training manuals and fieldwork materials, including questionnaires and field gear, were dispatched to the provinces in preparation for provincial training and data collection by the courier company. Questionnaires and EA Summary Books were captured on the Census and Survey Administration System (CSAS) for tracking purposes.

3.2 Recruitment and training of field staff

For fieldwork purposes, temporary personnel (Fieldworkers and Fieldwork Supervisors) were recruited from the EAs/districts in which they would be working and underwent rigorous training on fieldwork procedures to ensure that they deliver work of high quality at the end of the fieldwork phase. Experienced permanent staff members from Household Surveys (based in provincial offices) were seconded to the project for the duration of data collection in supervisory positions to ensure high-quality data and minimise costs. These staff members fulfilled the roles of Provincial Survey Coordinator, District Survey Coordinator, Quality Assurance Monitor and District Logistics Officer. The reporting structure for data collection in each province is illustrated in the diagram below:

PES fieldwork structure in provinces



Logistics training for District Logistics Officers (DLOs), District Survey Coordinators (DSCs), Data Capturers (DCaps) and Provincial Survey Coordinators (PSCs) was conducted from 26 to 30 September 2011. Training was on forward and reverse logistics, including the use of CSAS.

National training on PES fieldwork methodologies and procedures was conducted from 10 to 21 October 2011. This level of training addressed Provincial Survey Coordinators (PSCs), District Survey Coordinators (DSCs) and Quality Assurance Monitors (QAMs). Training covered procedures for publicity, listing and enumeration. There was a two-day field practice exercise which was conducted in selected enumeration areas (EAs) around Pretoria; the first day addressed practical training on publicity and listing, and the second day addressed practical training on questionnaire completion. A feedback session was conducted after each exercise.

Provincial training on PES fieldwork methodologies and procedures commenced on 24 October 2011 and was completed on 04 November 2011. This level of training addressed potential Fieldworkers (FWs) and Fieldwork Supervisors (FWSs). There was 20% over-recruitment in order to increase the pool of candidates to choose from. Trainees were appointed upon completion of training, and on the basis of scores obtained in assessments administered during training. Candidates were expected to obtain an overall score of at least 60% in order to be considered for appointment. All provinces were able to recruit the required number of staff for PES fieldwork. Those who resigned during training were replaced. Recommendations for appointment (on the basis of scores obtained during training) were signed by the Project Manager and Provincial Executive Managers.

The PES followed the integrated approach towards fieldwork; whereby one Fieldworker conducted publicity, listing and enumeration in one EA. This required the appointment of 608 people across the country to ensure that all EAs in the PES sample were allocated a Fieldworker. A ratio of one (1) Fieldwork Supervisor for four (4) Fieldworkers was applied; which meant that 152 Fieldwork Supervisors were to be appointed, but due to the spread of the sample in various districts, this ratio could not always be applied; hence, a total of 160 Fieldwork Supervisors were appointed.

3.3 Publicity and listing

Publicity and listing commenced on 14 November 2011, and was completed on 22 November 2011. Publicity was conducted at gatekeeper level, EA level and then at dwelling unit (DU) level simultaneously with listing. Posters and information sheets were used as publicity instruments. Listing involved the recording of all structures (including all DUs, number of households in DUs and number of persons in households) in the sampled EAs in the EA Summary Books.

Quality assurance measures were applied during this process, including the use of global positioning system (GPS) devices to verify EA boundaries. The approach to quality assurance also involved the identification of areas in methodology that proved challenging for staff to implement correctly, e.g. variables in the EA Summary Book which Fieldworkers found difficult to populate correctly. The supervisory layers were expected to check each Fieldworker's EA Summary Book(s) on a daily basis to ensure that all listings were done according to prescribed methodology using the provided quality assurance checklists. Once these areas of improvement were identified by the Fieldwork Supervisor, Quality Assurance Monitor, District Survey Coordinator and Head Office Monitors, Fieldworkers were re-trained to ensure that they were able to implement the methodology correctly. Quality checks were intensified at the beginning of listing to ensure that staff understood the implementation of procedures from the onset. The quality of work from each Fieldworker improved as the survey progressed.

Three (3) EAs were confirmed as vacant; one in Northern Cape, one in Western Cape and one in KwaZulu-Natal. One EA in Western Cape was also confirmed as an institution upon completion of listing. Enumeration could therefore not be conducted in these EAs since institutions are out of scope for the PES. All EAs were verified by supervisory staff and Head Office Monitors before being assigned a final status.

3.4 Enumeration

Enumeration, which started on 23 November 2011 in all provinces, was completed on 15 December 2011, with mop-up operations being completed on 22 December 2011. Enumeration was conducted in 608 (initially 600 but increased to 608 due to split EAs) enumeration areas (EAs). This process involved the enumeration of all persons and households who were present on census night (09–10 October 2011) and PES night (06–07 November 2011) within the boundaries of the sampled EAs. The PES questionnaire was used for interviewing respondents for each household and recording responses.

Based on the outcomes of the interviews, i.e. whether the interview was conducted successfully or not, the final result code was recorded on the cover page of the questionnaire. The final result codes were as follows:

Result code	Response details
11	Completed
12	Partly completed
21	Non-contact
22	Refusal
31	Unoccupied
32	Vacant
33	Demolished
34	New dwelling under construction

The overall response rate was 94,8% (see table below).

Table 2: Summary of result codes for response rate

Province	Response Rate
Western Cape	96,5
Eastern Cape	98,1
Northern Cape	97,6
Free State	99,3
KwaZulu-Natal	97,1
North West	95,6
Gauteng	86,0
Mpumalanga	99,2
Limpopo	99,8
Total	**94,8**
Urban	93,2
Non-urban	98,3
Total	**94,8**

Table 3: Summary of final result codes

	PES_FINAL_RESULT_CODE									
		0	11	12	21	22	31	32	33	34
Province	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**	**N**
Western Cape	1	5	11 558	10	339	75	1 135	453	16	100
Eastern Cape	2	5	10 108	7	168	30	922	655	46	247
Northern Cape	.	4	2 142	3	46	7	216	200	6	35
Free State	1	3	4 229	4	25	5	409	339	3	101
KwaZulu-Natal	.	34	15 868	10	356	125	1 255	472	66	364
North West	.	7	5 441	6	231	19	740	277	10	53
Gauteng	3	38	17 199	85	2 375	430	1 182	665	26	123
Mpumalanga	.	7	4 345	2	15	18	493	164	2	109
Limpopo	3	11	6 563	11	3	13	744	428	20	163
Total	**10**	**114**	**77 453**	**138**	**3 558**	**722**	**7 096**	**3 653**	**195**	**1 295**
Urban	9	82	52 650	123	3 142	698	4 285	1 512	95	505
Non-urban										
Total	**10**	**114**	**77 453**	**138**	**3 558**	**722**	**7 096**	**3 653**	**195**	**1 295**

Households were enumerated through interviews, and responses were recorded on the PES questionnaire. The quality assurance approach used during publicity and listing was also used during enumeration, i.e. the identification of areas in methodology that proved challenging for staff to implement correctly, e.g. variables in the questionnaire which Fieldworkers found difficult to populate correctly. The supervisory layers were expected to check each Fieldworker's questionnaires on a daily basis to ensure that all questionnaires were completed correctly and according to prescribed methodology. Once these areas of improvement were identified by the Fieldwork Supervisor, Quality Assurance Monitor, District Survey Coordinator and Head Office Monitors, Fieldworkers were re-trained to ensure that they were able to implement the methodology correctly. Quality checks on questionnaires were conducted throughout the enumeration phase.

The highest number of refusals was observed in Gauteng, representing 57% of all refusals in the country. Unoccupied dwellings were mostly observed in formal residential areas and traditional residential areas. Unoccupied dwellings were high in formal residential areas in Western Cape, Gauteng and KwaZulu-Natal, as well as traditional residential areas in Eastern Cape, KwaZulu-Natal and Limpopo. The possible reasons for the extremely high cases of non-response include the following:

- Knock-on effects of delayed field activities due to extension of census non-response follow-up;
- Problems relating to logistics;
- Timing of PES fieldwork operations (holiday season);
- High-walled areas/gated communities (especially in Gauteng);
- No media coverage for PES;
- Concerns regarding safety;
- Farms (in Eastern Cape) that are no longer occupied;

- People migrating to urban areas (rural-urban migration) in search of employment;
- People moving to Reconstruction and Development Programme (RDP) houses and leaving shacks behind;
- Seasonal employment in farms (people were absent from dwellings on farms during PES since there was no employment at the time);
- Inability to complete construction due to lack of funds (NDUCs); and
- Migration of students after end-of-year examinations.

3.5 Reverse logistics

After the completion of PES data collection in December 2011, all PES questionnaires were supposed to be captured onto CSAS (as per result code on the questionnaires). As per PES sample (including split EAs), 608 EA questionnaire boxes were supposed to be captured. All questionnaires were received at the Data Processing Centre (from provinces) by 20 January 2012 and the capturing of these questionnaires on CSAS commenced on 26 January 2012. Capturing and storage of questionnaires was completed by 30 January 2012. A total of 608 EA questionnaire boxes were received, with a total of 95 355 questionnaires.

3.6 Challenges experienced during data collection

Logistical arrangements

Not all materials were received in time from service providers, e.g. the questionnaires, which meant that the team had to revise the timelines for forward logistics. Since the PES could not appoint staff for key positions in the project structure, the few staff members available had to work long hours creating consignments for dispatch to each province. Some field staff members received airtime (minutes or credit for making cellphone calls) very late, or did not receive airtime at all, which made it difficult for staff to communicate during fieldwork.

It was initially planned that vehicles would only be issued to supervisory layers, i.e. Provincial Survey Coordinators, District Survey Coordinators, Quality Assurance Monitors and Fieldwork Supervisors, but due to vast distances and other issues (e.g. accessibility) in some districts, vehicles had to be provided to a number of Fieldworkers to ensure an acceptable level of progress. Most of the vehicles allocated to PES field staff had been used during census data collection, and some of them had mechanical problems or were in dire need for service by the time the PES staff received them. In some instances, service providers took more than 48 hours to replace vehicles, and this hampered progress. In Gauteng, PES funds were used for other Stats SA projects, and the funds were not available by the time the PES needed money for fuelling (before the arrival of petrol cards). The unavailability of funds, including the late delivery of petrol cards, also had a negative impact on progress. A number of vehicles were also involved in accidents.

Recruitment and training of field staff

The recruitment strategy for the PES stated that there would be over-recruitment by 20% in each province in order to increase the pool of candidates to choose from at the completion of training. This was not always possible to achieve since the field staff database was exhausted in most provinces by the time PES started recruiting. In some instances, people were headhunted from the areas in which the PES was conducted, but this was not always successful. Some trainees also dropped out of training, which also reduced the number of candidates to choose from. Human Resource Management (HRM) processes also had a negative impact on progress in instances whereby field staff had to be replaced. Numerous follow-ups had to be made regarding the signing of documents for the appointment of staff.

The secondment of permanent staff also proved to be a challenge in instances where some of them refused to work over weekends (which was a requirement for census and PES fieldwork due to timeframes), or they undermined the authority of contract staff who were in higher positions. This had a negative impact on staff morale because it meant that one person had to do the work of two people over the weekends.

Publicity and listing

Publicity and listing was initially planned for 07 to 22 November 2011, but due to the large presence of census field staff in the field and a decision by senior management, the start date for publicity and listing was moved to 14 November 2011. This meant that the period for this exercise was shortened to at least a week, leaving very little time for the team to conduct comprehensive quality assurance on listings. This resulted in the team having to do numerous updates on the EA Summary Books during enumeration for the structures that had been missed during listing. Despite the postponement of PES publicity and listing, census field staff was still present in some EAs that were in the PES sample, which compromised the independence of the PES.

Pamphlets and posters were only available in English, which meant that Fieldworkers had to spend a lot of time explaining the messages in these posters and pamphlets in the language of the respondent. The quality of the posters was also poor (posters were easily ruined by wind and rain, and had a small font size). The fact that the census had not communicated to the public about the PES (despite the fact that the PES had requested census to include messages on PES in census publicity material) also meant that Fieldworkers had to spend a lot of time explaining the purpose of their visit to respondents, which resulted in refusals in some cases.

Enumeration

The PES questionnaire was fairly short and had been translated into all the official languages, which made the interviewing process fairly easy. However, the challenge was the different dialects of each language from place to place, e.g. the question on relationship to the head of the household in the Sotho translation was interpreted differently in certain areas. The question on age was also cause for concern among field staff since some of them thought they were supposed to link a person's age to the reference night. The period of enumeration for the PES presented a challenge for most of field teams, especially in Gauteng and the coastal provinces, since a significant number of respondents could not be contacted since they had already left for the holidays.

It was difficult to enumerate in high-walled areas since some gatekeepers only allowed Fieldworkers to work in the area at specified times. The sticker barcode methodology was also not used correctly by census enumerators in some instances, e.g. some households reported that they had not been enumerated during census but had a sticker barcode on their doors, some households had multiple stickers and some households had not been issued with stickers. Some respondents refused to be enumerated during the PES since census enumerators had informed them that they were being enumerated for the last time by Stats SA, and therefore should not allow anyone else to enumerate them.

Reverse logistics

The inability to capture questionnaires on CSAS at the provincial office (due to network problems) for purposes of reverse logistics meant that it was difficult to confidently state the quantities and state of the questionnaires that should be expected at head office upon delivery since CSAS could not be used as a tracking mechanism. A manual form was used. Arrangements also had to be made for the capturing of questionnaires (checking in) at the Data Processing Centre in order to prepare for data processing.

4. MATCHING AND RECONCILIATION

This chapter describes the methodology that was used to match census and PES records. It provides a summary of the activities that were implemented from the processing of questionnaires, initial matching, reconciliation visits and final matching.

4.1 Data processing

The processing of PES data was done by the Census Data Processing team and was completed in March 2012. This process involved the scanning of PES questionnaires in order to store the data in the form of electronic images and text files. The scanning of census questionnaires for the EAs that are in the PES sample, including surrounding EAs, were prioritised to ensure that both census and PES questionnaires were made available in time for matching and reconciliation. The PES sample consisted of 600 EAs, but 596 EA questionnaire boxes were scanned for the PES since four EAs had been identified as out of scope after the completion of listing (three EAs had been identified as vacant and one as a special dwelling institution (SDI)). The Census Data Processing Centre (CDPC) handed over the data to the PES team in March 2012.

4.2 Recruitment and training of staff

Training on matching procedures

Training on matching procedures was conducted over a period of one week. Theoretic training was carried out over a period of five days (12–16 March 2012) and practical training was conducted over a period of two days (17–18 March 2012) and was computer-based (on the computer-assisted manual matching system). Human Resource Management was requested to recruit 360 candidates in order to increase the pool of candidates to choose from (20% over-recruitment). 300 staff members were selected from a pool of 335 trainees upon completion of training as well as scores obtained for assessments administered during training. Candidates were expected to obtain an average score of 60% for all four (4) assessments administered during training in order to be considered for appointment. The 300 candidates were appointed for the following positions:

- 180 Matching Clerks,
- 60 Matching Supervisors, and
- 60 Quality Assurance Monitors.

Training on reconciliation visit procedures

Training on reconciliation visits procedures was conducted from 26–28 March 2012 in Pretoria. Training included a practical exercise on the use of global positioning system (GPS) devices to ensure that reconciliation visits are carried out within the correct enumeration area (EA) boundaries, and make a clear distinction between households that were correctly or erroneously enumerated (boundary-related inclusions/exclusions). The Reconciliation Visits team was made up of 27 staff members who had

occupied supervisory positions during PES fieldwork (November–December 2011) from all nine (9) provinces (three from each province). The team was complemented by three staff members from head office who were responsible for the management and coordination of fieldwork operations for the PES.

4.3 Matching operations

The matching process involved the comparison of household and person records in census data and PES data. Guidelines for matching, including rules for determining the match status of households and individuals, incorporated lessons learnt during the Census Dress Rehearsal PES. Data was captured onto a computer-assisted manual matching system. The computer-assisted manual matching system was initially developed and tested during the Census Dress Rehearsal PES, but was improved for the main PES (incorporating lessons learnt). An automated system was also developed, but never tested during the Census Dress Rehearsal PES, hence it was not recommended as the main tool for matching. The automated system applied the same rules for matching as the manual system, but using scanned data. This system was used to complement the computer-assisted manual matching system to ensure that all households and persons were correctly matched. A SAS-based automated system was also used to complement these two systems in ensuring matches for persons and households. This system used a weighted scoring procedure to determine best matches for persons and households. The main phases in the matching process were:

4.3.1 Initial matching

Initial matching involved searching through census records in order to find the corresponding cases from PES enumeration records, and vice-versa (a two-way match). The matching of households required staff to take one household from the PES and locate its corresponding household from the census using the barcode sticker, physical identification of the dwelling unit and at least one 'matched' person. The following conditions were applied for identifying matched households:

Table 4: Conditions for identifying matching PES and census households

Physical ID of DU	Sticker barcode number	Person match (1 or more matched)	HH group
✓	✓	✓	Household match
✓	X	✓	Household match
X	✓	✓	Household match
✓	✓	X	Possible match
✓	X	X	Possible match
X	✓	X	Possible match
X	X	✓	Possible match
X	X	X	Non-match

Persons appearing in both census and PES questionnaires for the same household were matched according to their primary variables, i.e. sex, date of birth, age and population group. Tolerance limits for age for matched persons were applied in the following manner:

Table 5: Tolerance limits for age

Age (years)	Tolerance
Under 10	-2 to +2 years
10–19	-3 to +3 years
20–39	-5 to +5 years
40–59	-7 to +7 years
60+	-10 to +10 years

4.3.2 Capturing and assigning of initial match status

Capturing involved the capturing of PES and census information on a capturing tool which formed part of the computer-assisted manual matching system. Information for non-matched households and persons was also captured. Quality assurance measures applied during this phase included the checking of what had been matched and captured by the Matching Clerk against the information on PES and census questionnaires. Inspection of captured information was done by the Matching Supervisor and Quality Assurance Monitor. The process of assigning an initial match status and moving status for households and persons was automated, and done simultaneously with the capturing process. The computer-assisted matching phase produced three matching outcomes for households and seven for persons.

Households

Household Initial Match Status

1 = Matched
2 = Possible match
3 = Non-match

All 'Possible match' and 'Non-match' cases were considered 'unresolved' and were reviewed in order to determine if they required field follow-up (reconciliation) to determine their final status.

Persons

Person Initial Match Status

1 = Matched

2 = Possible match

In PES not in census:

3 = in PES not in census – definite non-match

4 = in PES not in census – insufficient or unclear information

5 = in-mover

6= born after census

7 = in census not in PES

Cases of 'possible match', 'in PES not in census – insufficient or unclear information', and 'in census not in PES' were considered unresolved and required field follow-up (reconciliation visit) to determine their final match status. Additionally, other cases were sent to reconciliation if they had an unclear or insufficient person presence. Furthermore, some cases of 'in PES not in census – definite non-match', that is, persons in 'non-match' households were sent for field follow-up in order to confirm if they could not be matched with cases of 'in census not in PES'.

4.3.3 Reconciliation visits

The **reconciliation visits operation** consisted of (field) follow-up visits to households in the PES sampled EAs. The purpose of the reconciliation visits was to collect relevant information in order to determine the final match status of unresolved cases identified during initial matching, specifically to:

- resolve the final match status for 'Possible Match' cases;
- determine whether households and/or persons enumerated in the census but not in the PES were correctly or erroneously enumerated in the census;
- determine whether households and/or persons enumerated in the PES but not in the census were correctly or erroneously enumerated in the PES;
- clarify doubtful cases or cases with insufficient or unclear information;
- clarify cases with unclear or insufficient person presence; and
- investigate EAs where boundary or enumeration quality problems are suspected.

4.3.4 Final matching

Final matching involved the use of the results obtained from the reconciliation visits and initial matching phases to assign a definite match status to each RV case. The table below illustrates the expected outcomes from final matching.

1. matched

In PES not in census:

2. missed in census
3. PES erroneous inclusion – cases in PES not in census that were outside the EA boundaries or otherwise erroneously included in PES
4. PES insufficient information – cases in PES not in census for which a final match status cannot be assigned due to insufficient information
5. in-mover
6. born after census

In census not in PES:

7. correctly enumerated in census, missed in PES
8. Census erroneous inclusion
9. Census insufficient information – cases in census not in PES for which a final match status cannot be assigned due to insufficient information

4.4 Challenges during matching

The major challenge experienced during matching was the fact that not all census EA questionnaire boxes (for census EAs in the PES sample) were received by the matching team at the beginning of the matching exercise, and also by the proposed end-date of matching. This resulted in the extension of the matching exercise to allow for the thorough search of EA questionnaire boxes that were relevant to the PES at the Census Data Processing Centre.

5. ESTIMATION PROCEDURES

5.1 Sampling weights

5.1.1 Base weights

The sampling frame consisted of 89 305 EAs (after deleting the vacant, institution, recreational and industrial EAs) from the Census EA database of June 2011. The sample allocation is indicated in Section 2.3. The EAs within each province were stratified by urban and non-urban. Within each stratum, units were sorted geographically (using EA code) and then EAs were selected using the **systematic sampling method** using the SAS procedure called 'surveyselect' to obtain the sample of 600 EAs nationally.

The base weight of a sample EA was equal to the inverse sampling rate, calculated as follows:

The sampling rates vary from stratum to stratum. The base weight for the sample persons and households in stratum h is:

$$W_h = \frac{N_h}{n_h},$$

where,

N_h = total number of EAs in the frame for stratum h

n_h = number of sample EAs selected in stratum h for the PES 2011

The base weight is given by total number of EAs in the stratum divided by the number of sample EAs in that stratum. Within each EA, the weight for each household and each person was equal to the EA sampling weight, since their probability of selection, given the selection of the EA, was equal to one. It should be noted that during data collection there was no *substitutions of selected EA due to field or frame related problems* and no *adjustments to base weights*.

5.1.2 Application of base weights to survey estimates

The PES 2011 sample estimates were inflated to represent the entire population, hence it was necessary to multiply the data by a sampling weight, called expansion factor. These expansion factors were applied at EA level.

There was no *non-response adjustment* at EA level or household level for PES 2011. Hence the inverse selection probabilities were applied as calculated from the 'PROC SURVEYSELECT' without any further adjustments.

5.2 Coverage evaluation: Calculation of dual-system estimates for persons

Coverage measures were calculated only for cases belonging to the universe of interest (Section 2.2: Scope of the PES).

1. The initial estimates – weighted estimates of total from the sample – using Procedure C, are (also see Figures 1 and 2):

    a. total number of non-movers in the universe (P sample);
    b. total number of out-movers in the universe (P sample);
    c. total number of in-movers in the universe (P sample);
    d. total number of matched non-movers in the universe (P sample);
    e. total number of matched out-movers in the universe (P sample);
    f. estimated total number of matched in-movers in the universe;
       Note: in Procedure C, the number of matched in-movers cannot be calculated directly, given that matching the in-movers in the sample is not attempted. But the 'out-movers' and the 'in-movers' constitute the same group in the universe: the 'movers', assuming a closed population. Therefore, an assumption can be made that, in the universe, the match rate for in-movers would be the same as that for out-movers (estimated by e/b). Hence, the total number of matched in-movers in the universe is estimated indirectly by [(e/b)*c].
    g. total number of census erroneous inclusions in the population (E sample);
    h. total number of cases correctly enumerated in the census but missed in the PES (E sample);
    i. total number and percentage of census persons with insufficient information (E sample); and
    j. total number and percentage of PES erroneous inclusions and PES insufficient information cases (P sample).

Figure 1: Initial derivations in dual-system estimation

	Parameter	Derivation
I1	Estimated no. of non-movers and % of total population represented by non-movers	NM / PES Pop
I2	Estimated no. of out-movers and % of total population represented by out-movers	OM / PES Pop
I3	Estimated no. of in-movers and % of total population represented by in-movers	IM / PES Pop
I4	Estimated no. and rate of matched non-movers	Matched NM / NM
I5	Estimated no. and rate of matched out-movers	Matched OM / OM
I6	Estimated no. of matched in-movers	I5 rate * I3 total
I7	Estimated no. of census erroneous inclusions	weighted sum
I8	Estimated no. of census correctly enumerated persons missed in PES	weighted sum
I9	Estimated no. and % of census persons with insufficient information	divide by A1a
I10	Estimated no. and % of PES erroneous inclusions cases	divide by A2a
I11	Estimated no. and % of PES insufficient information cases	divide by A2a

2. The 'matched' population is given by the total number of matched non-movers plus the estimated total number of matched in-movers in the universe. The match rate is calculated as a percentage of the PES population.

MATCHED POP = MATCHED NON-MOVERS + ESTIMATED MATCHED IN-MOVERS

Figure 2: Analysis derivations in dual-system estimation

	Parameter	Derivation
A1a	Census population (uncorrected for erroneous inclusion and insufficient information)	(I4 + I6) + I7 + I8 + I9
A1b	Census population (corrected for erroneous inclusion and insufficient information)	(I4 + I6) + I8
A2a	PES population	I1 + I3
A2b	Matched population	I4 + I6
A3	PES persons missed in Census – Total	A2a – (I4 + I6)
	PES persons missed in Census – Rate	divided by A2a
	Coverage rate	[1 – A3 rate]
A4	Census correctly enumerated missed in PES	I8
	Census correctly enumerated missed in PES – Rate	divided by A1b
A5	Census erroneous inclusions – Total	I7
	Census erroneous inclusions – Rate	divided by A1a
A6	Preliminary dual-system estimate of true population	(A1b * A2a) / Matched pop
A7	Net error (net undercount) – Total	A6 – A1a
	Net error (net undercount) – Rate	divided by A6
A8	Gross error – Total	A3 rate * A6
	Gross error - Rate relative to true population	A3 rate
A9	'Adjustment factor' for Census	A6 / A1a
	Final dual-system estimate of true population	A9* census count

3. The E-sample estimate of the population enumerated in the census *[Uncorrected CENS_POP]* 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 cases with insufficient information.

   CENS_POP_UNCORR = MATCHED_POP + CORR_ENUM + ERR_INCL + INSUFF_INFO

   The census population corrected for erroneous inclusions and insufficient information *[Corrected CENS_POP]* is calculated without adding these last two categories.

   CENS_POP_CORR = MATCHED + CORR_ENUM

4. The P-sample estimate of the total population [*PES_POP*] is the sum of the non-movers and in-movers in the population.

   PES_POP = NON-MOVERS + IN-MOVERS

5. The PES-enumerated population missed in the census is calculated by subtracting the matched population from the PES estimate of the total population to obtain:

   PES_POP_MISSED_IN_CENSUS = PES_POP – MATCHED_POP

   The rate of PES population missed in the census is the missed population above relative to the PES estimate of total population.

6. The estimated total number of census erroneous inclusions *ERR_INCL* is the same as calculated in the initial tables. It includes fabrications, duplications, and geographic misallocations, etc. As mentioned, the main purpose of the E sample is to provide an estimate for this variable in order to permit a correction in the dual-system estimate of the true population.

   The census erroneous inclusion rate is equal to the total number of persons erroneously included in the census relative to the E-sample estimate of the census population.

7. The preliminary dual-system estimate of the *true population [TRUE_POP]* is the population estimated from the PES multiplied by the population estimated from the census (after correcting for erroneous inclusions and insufficient information) and divided by the population found in both:

$$TRUE\_POP = \frac{PES\_POP \times Corrected\,CENS\_POP}{MATCHED}$$

8. The net coverage error – universally known as the 'net omission rate' or the 'undercount' – is the difference between what should have been counted (true population) and what was counted (census population). The net coverage error represents the undercount still remaining in the census figures even after the partial cancellation caused by the overcount.

   Net Undercount = TRUE_POP – CENS_POP_UNCORR

   The net coverage error rate – the 'net omission rate' or the rate of 'undercount' – is the total net error relative to the dual-system estimate of the true population; that is, divided by TRUE_POP. **This measure constitutes the single most important indicator of the quality of the census coverage.**

9. The gross coverage error – the 'gross omission' – is, as defined in this context, what the census truly missed without taking into account the overcount. It is the gross omission relative to the true population, as opposed to the net omission, that is, without being offset by the erroneous inclusions.

   Gross Coverage Error= *Population Found in PES Missed in census*
   + Population Missed in Both census and PES

$$= PESPersonsMissedinCensus + \frac{(TruePop - CensPopCorr) \times (TruePop - PESPop)}{TruePop}$$

$$\text{Gross Error Rate} = \frac{\text{Gross Error}}{\text{True Population}}$$

   Equivalently:

   Gross Error Rate *= 1 – Matched Pop/PES Pop*
   = rate of PES persons missed in census (Table A3)

   which means the total gross error can be calculated as:

Gross Error Total = *Rate PES persons missed in census* $\times$ *True Pop*

10. The final dual-system estimate of the *True Population*, which corresponds to the 'Adjusted Population', is obtained through the use of a ratio estimator of total, which is superior in accuracy to the preliminary estimate, by reducing both variance and bias.

$$\left[\frac{Preliminary\ TRUE\_POP}{CENS\_POP\_UNCORR}\right] * Actual\ Census\ Count$$

    where the ratio inside the bracket represents the 'adjustment factor' for the census count.

11. The relation between the undercount rate and the adjustment factor is the following:

$$ADJ\_FACT = \frac{1}{1 - undercount\ rate}$$

In other words, the adjustment factor is the reciprocal of the complement of the undercount rate.

For example, an undercount rate of 2% implies an adjustment factor of 1,0204. Likewise, an undercount rate of 8% implies an adjustment factor of 1,0870, and an undercount rate of 14% implies an adjustment factor of 1,1628, and so forth.

12. Another way of viewing the adjustment factor is the following:

$$\text{Adjustment Factor} = \frac{\frac{PES\_POP \times CENS\_POP\_CORR}{MATCHED\_POP}}{CENS\_POP\_UNCORR}$$

If we consider the quantity $\frac{MATCHED\_POP}{PES\_POP}$ as the 'Coverage Rate', then:

$$\text{Adjustment Factor} = \left[\frac{1}{Cov\ Rate}\right] \times \left[\frac{CENS\_POP\_CORR}{CENS\_POP\_UNCORR}\right]$$

While the quantity inside the first bracket is clearly a correction for under-enumeration, the quantity in the second bracket – which is the proportion of the census population that was correctly enumerated, i.e., not erroneously included – serves as a correction for over-enumeration.

Hence, the final adjusted population is in effect calculated as follows:

Adjusted Population = *under-enumeration correction factor* $\times$ *over-enumeration correction factor* $\times$ *census count*

Also note that the under-enumeration correction factor is always ≥ 1 and the over-enumeration correction factor is always ≤ 1. The overall factor can theoretically fall on either side of 1, depending on which is higher, the undercount or the overcount.

13. The probabilities of inclusion and omission of a person are calculated as follows:

Figure 3: Derivation of probabilities of inclusion

P (included in Census)	=	Census Population Corrected / True Population
P (included in PES)	=	PES Population / True Population
P (included in both Census and PES)	=	P (included in Census) * P (included in PES) *per independence assumption*
P (included in Census, but missed in PES)	=	P (included in Census) * [1 - P (included in PES)]
P (included in PES, but missed in Census)	=	P (included in PES) * [1 - P (included in Census)]
P (missed in both Census and PES)	=	[1 - P (included in Census)] * [1 - P (included in PES)] *per independence assumption*

14. The distribution of the true population – based on the preliminary dual-system estimate, after removing erroneous inclusions and insufficient information cases in census – is the following:

Figure 4: Derivation of population distribution estimates

		Census Enumeration		
		Included	**Omitted**	**Total**
PES Population	**Included**	MATCHED POP	in PES, missed in Census	PES POP
	Omitted	in Census, missed in PES	Missed in both	
	Total	CENSUS POP CORR	GROSS OMISSION	TRUE POP

It is given by:

Census pop corrected for err incl & insuff. info	=	P(included in census)	× dual-sys estimate of pop
PES pop (excludes err. incl. & insuff. info)	=	P(included in PES)	× dual-sys estimate of pop
Pop included in both census and PES	=	P(included in both census and PES)	× dual-sys estimate of pop
Pop included in census, missed in PES	=	P(included in census, but missed in PES)	× dual-sys estimate of pop
Pop included in PES, missed in census	=	P(included in PES, but missed in census)	× dual-sys estimate of pop
Pop missed in both census and PES	=	P(missed in both census and PES)	× dual-sys estimate of pop

5.3 Coverage evaluation for households

A working definition for households first had to be established. A PES 'household' was defined as a parent questionnaire, including the continuation questionnaires. The total number of matched households was calculated as the total number of matched questionnaires from the P-sample.

Next, the 'moving status' and 'match status' were defined for each household as follows: If at least one person in the questionnaire was matched, then the household was considered matched. If all persons in the questionnaire were missed, then the household was considered as a miss (in the census or in the PES).

5.4 Formation of adjustment classes

The overall coverage estimates when broken down (disaggregated) by geographic or demographic variables (such as province, sex, age group or population group) could be skewed due to the fact that persons and households are not evenly missed over such subgroups of the population. Homogeneous adjustment classes, i.e., classes within which coverage rates are more or less the same, are thus formed and a single adjustment factor is then calculated in each of the adjustment classes independently. The adjustment classes were obtained by using the Chi-square Automatic Interaction Detection (CHAID) technique.

The dependent variable was defined to be the “matched population” variable, where “0” indicates non-matched,“1” indicates matched (non-mover and out-mover) and a rate between “0” and “1” represented the probability of being matched as an in-mover. For persons, the predictors used (per province) were: geography type, sex, age group, and population group. For households, the predictors used were: province, geography type, and size of the household.

The CHAID model was used to determine combinations of the predictors that were statistically significant in modelling the coverage probability. The characteristics defined by the CHAID branches (i.e., the different branches in the dendrogram (decision tree) created by CHAID) were then taken as the adjustment classes. General PES 2011 specifications for all the decision trees at provincial levels were as follows:

- Significance level = 0,05 or 5%
- Leaf Size = 200
- Node Sample = minimum (5000), depending on the size (density) of the province: NC, NW and FS a Node Sample of 3000 was used while for EC, KZN, WC, MP, GP and LP a minimum Node Sample of 5000 was used.

After the creation of these various adjustment classes, a separate adjustment factor was calculated for each class, using the formulas described in Section 5.2. Due to the fact that the factors are ratios, the population when adjusted at the national or at the provincial level is not equal to the summation of the adjusted population over all adjustment classes. This is an inherent mathematical inequality – a difference between totals produced using combined ratios vs. separate ratios – and not a calculation error.

A decision had to be made on whether the national adjusted population should be the separate ratio estimate of total (summing up the adjusted population across adjustment classes) or the combined ratio estimate of total using the national adjustment factor. The separate-ratio estimate produced a lower variance (because of homogeneous classes with high heterogeneity among them, with a sufficient number of observations in each class) but it has a higher bias than the combined-ratio estimate. The combined-ratio estimate had higher variance but its bias is lower than that of the separate-ratio estimate

due to the consistency property of ratio estimators (which makes the bias diminish as *n* gets larger). Nevertheless, since each class had a large number of observations, the separate-ratio estimate was chosen.

As a result of this 'bottom-up' approach, the undercount rate was re-calculated in each publication cell as:

Adjusted in-scope population – Unadjusted in-scope population

5.5 Application of adjustment factors to census data

The adjusted population is obtained by multiplying the appropriate adjustment factor to the actual census count in the corresponding census adjustment class, and then summing across classes. In practice, this is equivalent to using a standard weighting procedure where the 'weight' corresponds to the adjustment factor.

As mentioned in the discussion of 'PES target universe' in Section 2.2, the PES was limited to a large subset of the population. Because the coverage rates in the balance of population are unknown, no adjustment was made for these persons.

Hence, as a first step in the application procedure, the total universe for the census was partitioned into two sets: 'Population within in-scope sub-universe' and 'Balance of population'. Each person or household was first determined to be in or out of the target population based on EA type, living-quarters type, and questionnaire type. Only eligible cases, i.e. cases in the in-scope sub-universe, received the designated adjustment factors. Non-eligible cases, i.e. balance-of-population cases, received an adjustment factor of 1.

The eligible person was then assigned, on an individual level, the adjustment factor corresponding to the adjustment class he belonged to, according to province, geography type, sex, age group, and population group. Similarly, each household was assigned, on an individual level, the adjustment factor corresponding to the adjustment class it belonged to, according to province, geography type, size of the household, and population group of the head of the household.

Census counts, both unadjusted and adjusted, were then calculated separately for the two population subsets:

Unadjusted census population =

Unadjusted 'Population within in-scope sub-universe'

\+ *Unadjusted 'Balance of population'.*

Adjusted census population =

Adjusted 'Population within in-scope sub-universe'

\+ *Unadjusted 'Balance of population'.*

It is worth noting that PES adjustment factors were based on the original geographic and demographic classifications of persons. For geography type and EA type, 'original' referred to the classification in the Census 2011 frame, before EA type changes occurred. For living-quarters type and for demographic variables, 'original' refers to these variables as originally reported in the census.

Therefore, to maintain compatibility between the distribution of PES cases and census cases, the **original** classifications (i.e., unedited or 'raw' data before editing and imputation) were used to decide which factor a person would receive. Thus, census persons received the adjustment factor corresponding to their original geography type and EA type, original living-quarters type, and original sex/age group/population group cell. Once the adjustment factors were applied, persons and households were permitted to shift to post-editing classification cells (which render census data more accurate and more meaningful), but they carried their original adjustment factors individually into their new cells.

5.6 Content evaluation for persons

Content analysis is discussed in Section 8.1. The following must be noted regarding the use of the PES for the measurement of content error:

- It is limited to matched cases.
- It is limited to the in-scope sub-universe, consisting of dwelling units and hostels within in-scope EA types.
- The PES is not assumed to provide the 'truth'; therefore, response bias is not measured, only response variance.
- Comparison is of unedited PES and census socio-demographic responses. (PES socio-demographic data are not subject to edit; census data are, but these edits take place outside the PES.)
- Unlike the census and PES questionnaires in the PES sample, data capture for the full census was not by key-from-paper but by scanning with rigorous quality control. In addition, census data were later subject to an intensive edit and automatic-correction process. Hence, to a certain extent, the data quality in the published census results is improved over what is indicated by the content analysis.

It was also noted in Section 8.1 that the estimated person totals shown in the content analysis tables do not coincide with the final census totals for each characteristic because:

- they are based on the sample of census records in the PES and are, therefore, subject to sampling variability;
- they include only matched cases, not the full sample;
- they are unedited while the census characteristics are edited;
- they include only the in-scope sub-universe while the final census totals include the full universe; and

- they are unadjusted while the final census totals are adjusted.

The sole purpose of these totals is to compare the census responses with the PES responses and to calculate the measures of consistency; they are not for socio-demographic analytical purposes. It is further noted that, although the content tables were supposed to include all matched persons, about 11% of the person records did not have their matching companion because of barcode/person number errors. They were thus omitted from the content error calculations. To the extent that these 11% might reflect different consistency patterns, the content error measures might be somewhat biased.

Variability is measured by means of four different indicators: the net difference rate, the index of inconsistency (simple and aggregate), the gross difference rate, and the rate of agreement.

5.6.1 Net difference rate (NDR)

The net difference rate is the difference between the number of cases in the census and the number of cases in the PES that fall under each response category, relative to the total number of matched persons in all response categories. The NDR formula for the *i-th* category is:

$$NDR = \frac{Y_{\bullet i} - Y_{i\bullet}}{n} \times 100$$

for i = 1, ..., C

where: $Y_{.i}$ = *unweighted number of census cases in i-th category*
$Y_{i.}$ = *unweighted number of PES cases in i-th category*
n = *unweighted number of matched cases*
C = *total number of response categories for characteristic 'y'*

5.6.2 Index of inconsistency

The index of inconsistency is the relative number of cases for which the response varied between the census and the PES. It is the ratio of the simple response variance to the total variance of the characteristic, including its variability in the population.

It is calculated for each response category 'i' according to the following formula:

$$I \equiv \frac{(y.i - yi. - 2yii)}{\frac{1}{n}[y.i(n - yi.) + yi.(n - y.i)]} * 100$$

(i = 1,...,C)

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

The following formula is used to calculate the **aggregate** index of inconsistency (that is, for all the response categories of the characteristic as a whole):

$$\hat{I}_{AG} = \frac{(n - \sum_{i}^{C} Y_{ii})}{(n - \frac{1}{n}\sum_{i}^{C} Y_{.i} Y_{i\bullet})} \times 100$$

5.6.3 Gross difference rate (also off-diagonal proportion)

The gross difference rate (GDR) 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.

$$GDR = \frac{(n - \sum_{i}^{c} Y_{ii})}{n} \times 100$$

5.6.4 Rate of agreement

The rate of agreement is the complement of the gross difference rate. A low rate of agreement indicates a high degree of variability, and vice-versa.

$$Rate\ of\ \ Agreement = \frac{\sum_{i}^{c} Y_{ii}}{n} \times 100$$

6. COVERAGE EVALUATION OF PERSONS

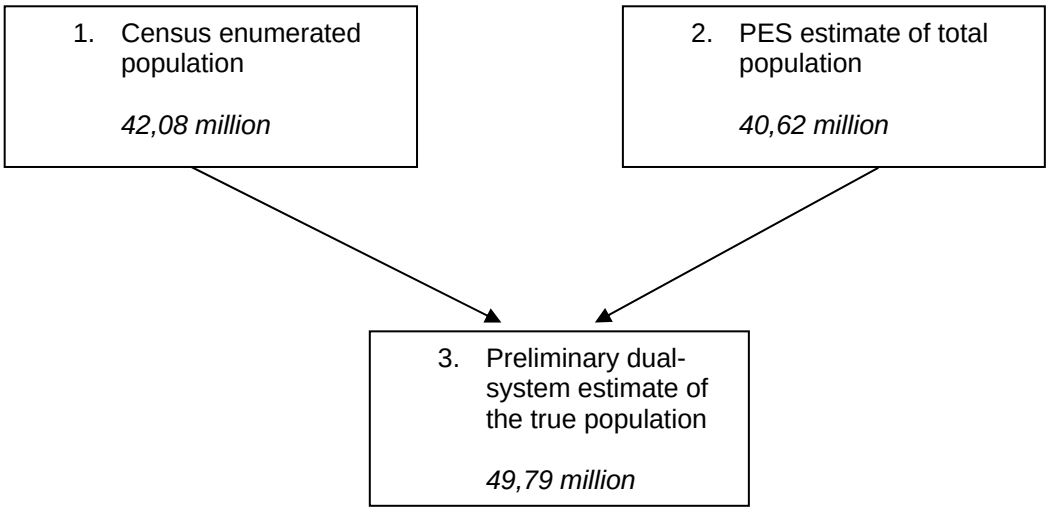
6.1 Estimate of *true population*

Two independent sources or 'systems' are used to arrive at the estimate of the *true population*: the census and the PES. The first attempt at measuring the true population yields the *census-enumerated population*, based on an exhaustive enumeration. The second attempt yields the *PES estimate of the total population*, based on sampling techniques.

Instead of assuming that one or the other is better, both of these estimates are used to derive a third, composite estimate of the *true population* called the 'dual-system estimate of the true population' (see Section 5.2 for estimation formulas). The dual system provides an estimate of the cases included in one source (PES) and excluded from the other (Census), and vice versa. Both estimates contribute to the dual system estimate, which is more complete than either the census or the PES estimate alone.

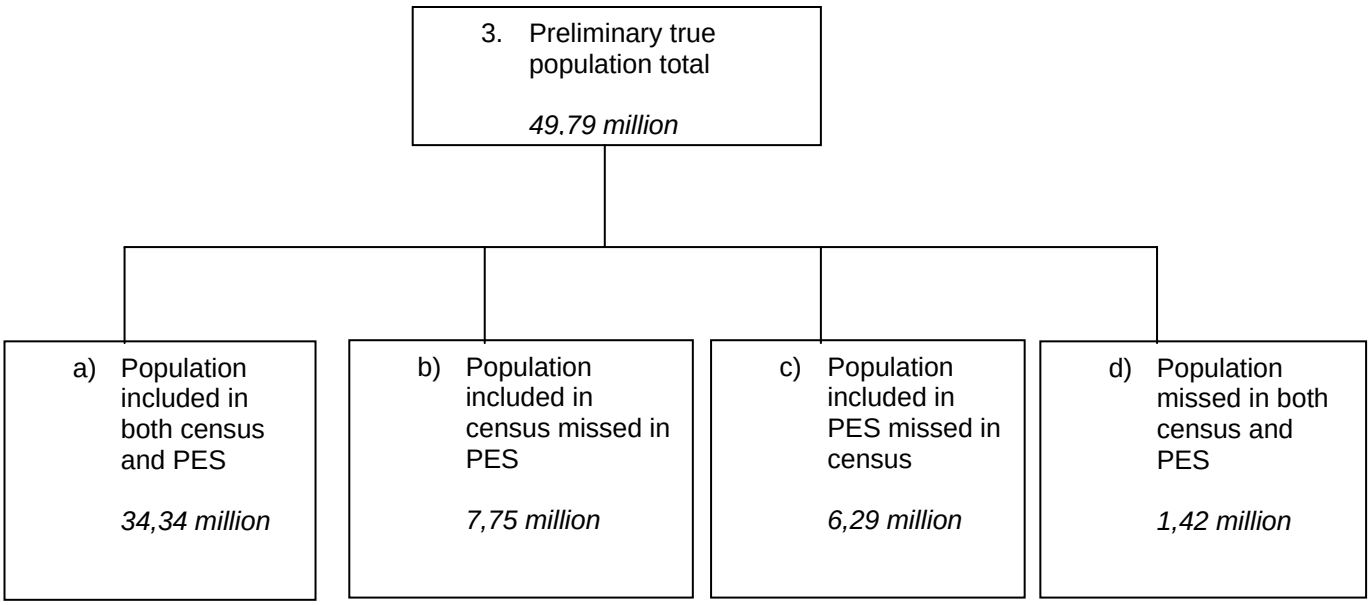
In the end, this *true population* is compared with the *census-enumerated population* and the difference is the net *undercount* (or *overcount*).

Figure 5: Estimates of total population from the individual systems and from the dual system in-scope sub-universe



In the in-scope sub-universe, the separate census and PES enumerations produced 42,08 million and 40,61 million persons, respectively. Using the dual system estimation method, the *true population* of South Africa in the in-scope sub-universe was estimated at 49,79 million. Four components together make up the dual-system estimate of the true population.

Figure 6: Breakdown of dual-system estimate of population total-in-scope sub-universe (in millions rounded to two decimals)



Components (a), (b), and (c) are obtained through a **matching** process, based on direct observation. Component (d) is a mathematical derivation, based on an assumption of independence. Component (a), the population included in both the census and the PES, was estimated at 34,34 million persons; component (b), the population included in the census but missed in the PES, was estimated at 7,75 million; component (c), the population included in the PES but missed in the census, was estimated at 6,28 million; and component (d), the population missed in both the census and the PES, was estimated at 1,42 million (derivations can be found in Figure 6).

Table 6 shows that, of the 42,51 million persons counted in the census for the in-scope sub-universe, 42,08 million are estimated to be correctly enumerated. Of these, the PES included 34,34 million and missed 7,75 million. The census erroneous inclusions (fabrications, duplications, and geographic misallocations) are estimated to be 0,42 million or approximately 1,0% of the census total.

Table 6: Coverage distribution of Census population – in-scope sub-universe (in millions rounded to two decimals)

	Census enumeration
Total excluding erroneous inclusions	42,08
Included in PES	34,34
Omitted from PES	7,75
Erroneous inclusions	0,42
Total including erroneous inclusions	**42,51**

*Sums are subject to rounding error.

It is estimated that the census omitted 7,71 million persons in total, 6,29 million of which were correctly enumerated in the PES and another 1,42 million of which were missed in the PES as well as the census

(Table 7). This total omission does not take into account the erroneous inclusions. When it is offset by the 0,42 million erroneous inclusions, the net undercount is 7,28 million. The net undercount relative to the 49,79 million in the *true population* is thus 14,6% (Table 10).

Table 7: Coverage distribution of true population – in-scope universe (in millions rounded to two decimals)

		Census enumeration		
		Included	**Omitted**	**Total**
PES Population	Included	34,33	6,29	40,62
	Omitted	7,75	1,42	9,17
Total excluding erroneous inclusions		**42,08**	**7,71**	**49,79**

*Sums are subject to rounding error.

While the PES estimated the total population in the in-scope sub-universe at 40,62 million, it omitted 7,75 million persons who were correctly enumerated in the census, and another 1,42 million who were missed in both the census and the PES, for a total omission of 9,17 million (Table 7).

The total South African population of 51,77 million persons was calculated by adding the census-enumerated 1,79 million persons in the other collective living quarters and the out-of-scope EA types to the adjusted census count of 49,98 million in the in-scope sub-universe (Table 8).

Table 8: Adjusted census population – full universe (in millions rounded to two decimals)

	Persons in dwelling units and hostels within in-scope EA types	**Persons in other collective living quarters and other EA types**	**Total population**
Adjusted	49,98	1,79	51,77

The overall empirical probabilities of inclusion and omission of a person in the census or in the PES are shown below in Table 9 (derivations can be found in Figure 3). According to the enumeration results, a member of the in-scope sub-universe had approximately an 84,5% chance of being enumerated in the census, an 81,6% chance of being enumerated in the PES, and a 68,9% chance of being enumerated in both. Conversely, the person had approximately a 15,6% chance of being included in the census but missed in the PES, a 12,6% chance of being included in the PES but missed in the census, and a 2,85% chance of being missed in both.

Table 9: Probabilities of inclusion and omission of a person – in-scope sub-universe

Probability of being included in Census	0,845193
Probability of being included in PES	0,815745
Probability of being included in both Census and PES	0,689462
Probability of being included in Census, but missed in PES	0,155731
Probability of being included in PES, but missed in Census	0,126283
Probability of being missed in both census and PES	0,028524

6.2 Estimation of the net undercount rate

The net undercount (or overcount) is the difference between the estimated true population (dual-system estimate) and the estimated census population. The rate is the net undercount expressed as a percentage of the estimated true population. The net undercount rates, together with their absolute errors and confidence intervals, are shown in the following tables for geographic and demographic groups. The confidence interval is formed around the estimate by adding or subtracting the absolute error. It must be noted that high absolute errors indicate that the estimate is not statistically reliable and confidence intervals are very wide as a result.

In Table 10, it can be observed that the net undercount rate at the national level was estimated at 14,6%, with confidence intervals ranging from 14,34 and 14,86. When comparing rates for different sets of persons, the confidence intervals must be taken into account. Before concluding that a 'differential' undercount exists, for example, that the undercount rate for one group is in fact higher (or lower) than that of another group, the two confidence intervals in question must not overlap. (This is equivalent to a two-tailed hypothesis test at the 0,05 level of significance.) An overlap in the intervals indicates that – except for a 5% chance of erring in the conclusion – the difference observed is not statistically significant due to random error; in other words, that there is no evidence of a real difference. A 'floating bars' chart is useful for visualising the intervals (see Figure 7).

Table 10: Net undercount rate for persons by province – in-scope sub-universe (values expressed in percentage points rounded to one decimal)

Province	Net undercount rate	Standard error	95% confidence interval limits	
			Lower	Upper
Western Cape	18,5	0,542	17,46	19,58
Eastern Cape	12,9	0,196	12,52	13,29
Northern Cape	13,4	0,318	12,75	13,99
Free State	10,1	0,362	9,39	10,81
KwaZulu-Natal	16,7	0,379	15,98	17,47
North West	14,9	0,532	13,84	15,92
Gauteng	14,7	0,174	14,35	15,03
Mpumalanga	15,5	0,473	14,53	16,38
Limpopo	10,0	0,135	9,72	10,25
All provinces	14,6	0,132	14,34	14,86
Geography type	**Net undercount rate**	**Standard error**	**95% confidence interval limits**	
			Lower	**Upper**
Urban	14,7	0,144	14,40	14,96
Non-urban	14,5	0,247	13,97	14,94
All areas	14,6	0,132	14,34	14,86

*Subject to rounding error

Among the provinces, the highest undercount was observed in Western Cape, KwaZulu-Natal and Mpumalanga (18,5%, 16,7%, and 15,5%, respectively) (see Table 10 and Figure 7). Undercounts in the provinces of North West (14,9%) and Gauteng (14,7%) are not significantly different from one another. The lowest undercount was observed in Limpopo (10,0%).

Figure 7: Graphic representation of confidence intervals for persons undercount rate – provinces

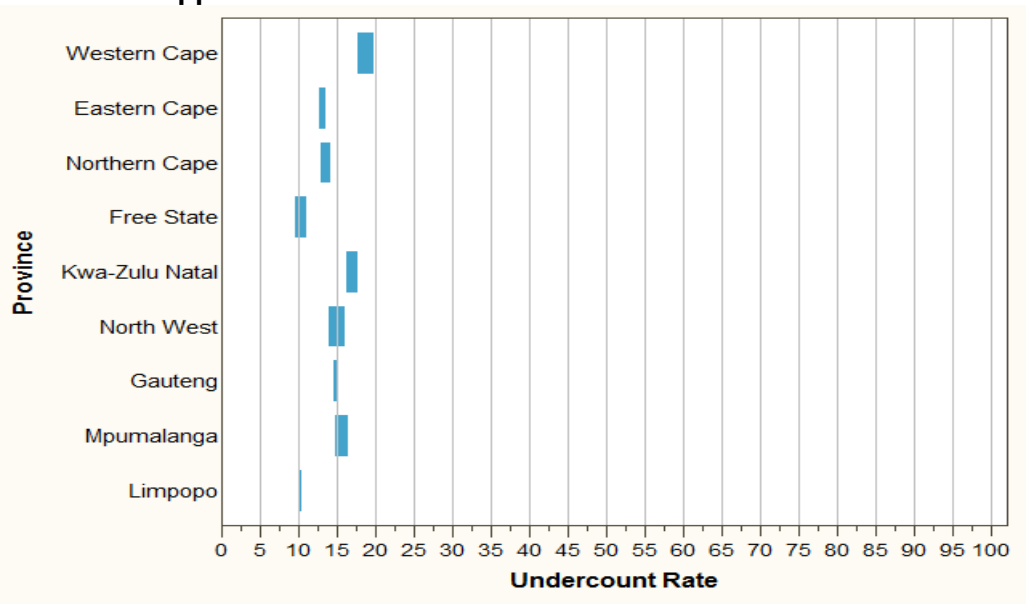


Table 11: Net undercount rate for persons by demographic group – in-scope sub-universe; single-variable classifications (values expressed in percentage points rounded to one decimal)

			95% confidence interval limits	
Category	**Net undercount rate**	**Standard error (+ or -)**	**Lower**	**Upper**
All persons	14,6	0,32	14,34	14,86
Population group				
Black African	9,9	0,138	9,58	10,12
Coloured	12,8	0,604	11,66	14,03
Indian or Asian	11,5	0,528	10,50	12,57
White	15,6	0,367	14,86	16,30
Other	23,2	1,165	20,88	25,45
Sex				
Male	15,9	0,135	15,63	16,16
Female	13,4	0,133	13,11	13,63
Age group				
Under 5 years	15,1	0,156	14,81	15,42
5-9 years	11,4	0,173	11,10	11,78
10-14 years	11,1	0,150	10,79	11,37
15-19 years	12,8	0,146	12,47	13,04
20-29 years	18,1	0,147	17,85	18,43
30-44 years	16,9	0,139	16,60	17,14
45-64 years	12,5	0,165	12,16	12,80
65+ years	9,8	0,193	9,46	10,22

*Subject to rounding error
*The undercount rate for undetermined population group is excluded from this table

The undercount for males (15,9%) is higher than that for females (13,4%). When population groups are compared, persons under the population group 'Other' were undercounted at a significantly higher rate than the other population groups, with a rate of 23,2%, followed by Whites with a rate of 15,6% (see Figure 8 below). There is no significant difference in undercount among the coloured and Indian/Asian groups. The lowest undercount was observed among Black Africans (9,9%).

Figure 8: Graphic representation of confidence intervals for persons undercount rate by population group

Lower and upper limits

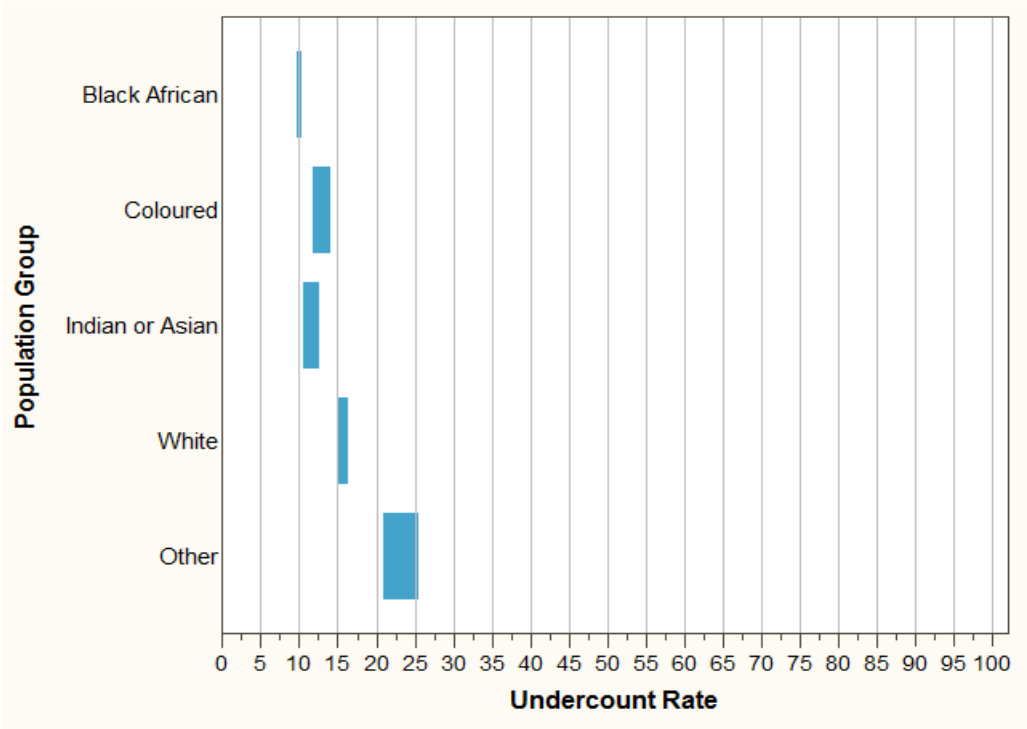
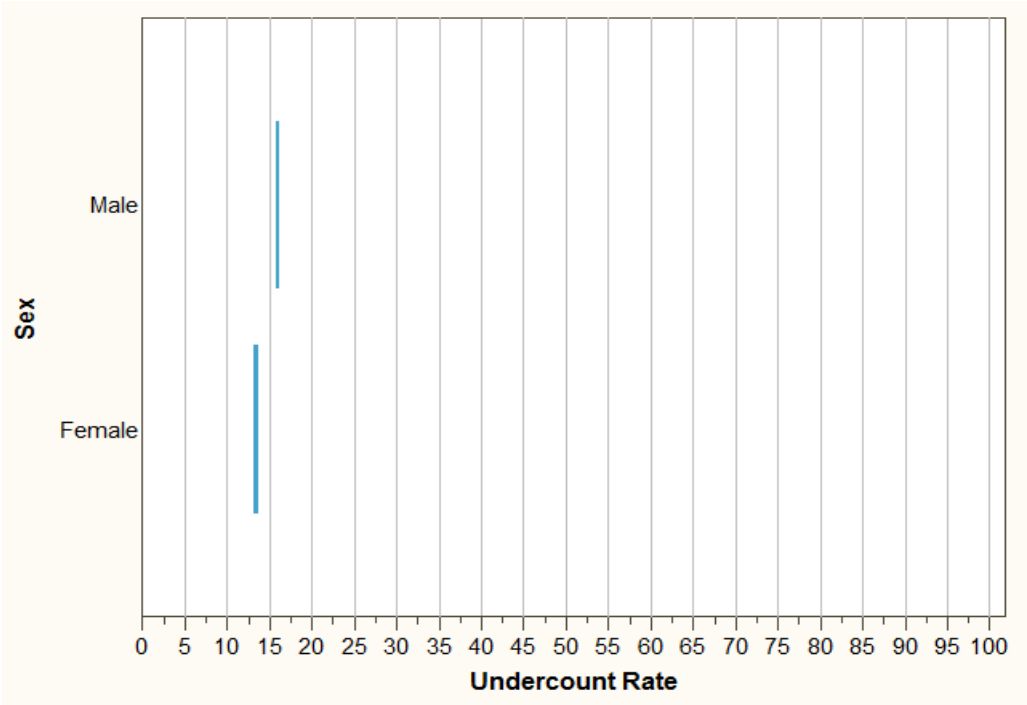


Figure 9: Graphic representation of confidence intervals for persons undercount rate by sex

Lower and upper limits



Except for the 20–29 years age group and the 30–44 years age group, no claim of a differential undercount among age groups can be made, that is, the undercount rate is in the same range for all the other age groups (see Figure 10 below). While these two age groups are significantly more undercounted than the other groups, their undercounts are not significantly different from each other.

Figure 10: Graphic representation of confidence intervals for persons undercount by age group

Lower and upper limits

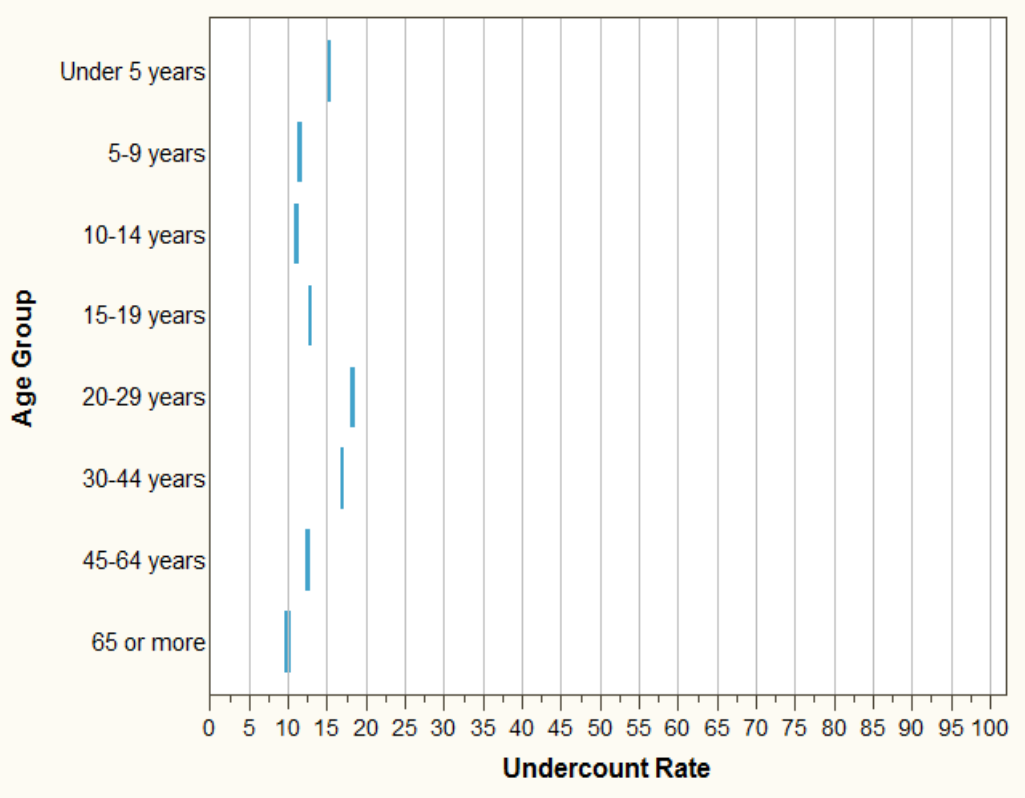


Table 12: Net undercount rate for persons by demographic group – in-scope sub-universe; two-variable classifications (values expressed in percentage points rounded to one decimal)

				95% confidence interval limits	
		Net undercount rate	Standard error (+ or -)	Lower	Upper
Population group by sex					
Black African	Male	10,5	0,142	10,21	10,76
	Female	8,6	0,140	8,34	8,88
Coloured	Male	13,0	0,609	11,77	14,16
	Female	11,9	0,602	10,68	13,04
Indian or Asian	Male	11,9	0,591	10,77	13,08
	Female	10,3	0,500	9,34	11,30
White	Male	15,6	0,370	14,84	16,29
	Female	14,5	0,368	13,79	15,23
Other	Male	24,0	1,002	22,07	26,00
	Female	19,8	1,562	16,76	22,88
Population group by age group					
Black African	Under 5 years	11,5	0,150	11,25	11,83
	5-9 years	8,2	0,191	7,83	8,58
	10-14 years	7,7	0,162	7,36	8,00
	15-19 years	9,4	0,157	9,07	9,69
	20-29 years	14,1	0,156	13,77	14,38
	30-44 years	12,8	0,147	12,46	13,04
	45-64 years	8,5	0,162	8,16	8,80
	65 or more	5,6	0,179	5,26	5,96
Coloured	Under 5 years	14,8	0,615	13,63	16,04
	5-9 years	14,1	0,633	12,89	15,38
	10-14 years	14,5	0,629	13,22	15,69
	15-19 years	13,9	0,648	12,63	15,17
	20-29 years	14,4	0,632	13,17	15,65
	30-44 years	14,6	0,622	13,39	15,83
	45-64 years	12,4	0,652	11,08	13,64
	65 or more	11,2	0,741	9,75	12,66
Indian or Asian	Under 5 years	14,4	0,723	12,98	15,82
	5-9 years	12,7	0,737	11,23	14,12
	10-14 years	11,5	0,759	10,06	13,04
	15-19 years	8,8	0,633	7,59	10,07
	20-29 years	16,4	0,553	15,33	17,50
	30-44 years	14,3	0,675	13,01	15,66
	45-64 years	8,8	0,597	7,63	9,97
	65 or more	10,0	0,594	8,88	11,20
White	Under 5 years	16,8	0,547	15,69	17,83
	5-9 years	16,4	0,487	15,41	17,32
	10-14 years	18,1	0,452	17,19	18,97
	15-19 years	18,3	0,396	17,52	19,07
	20-29 years	20,0	0,499	18,99	20,94
	30-44 years	18,1	0,398	17,37	18,93
	45-64 years	15,8	0,403	15,04	16,62
	65 or more	13,3	0,527	12,26	14,33

Table 12: Net undercount rate for persons by demographic group – in-scope sub-universe; two-variable classifications (values expressed in percentage points rounded to one decimal) (concluded)

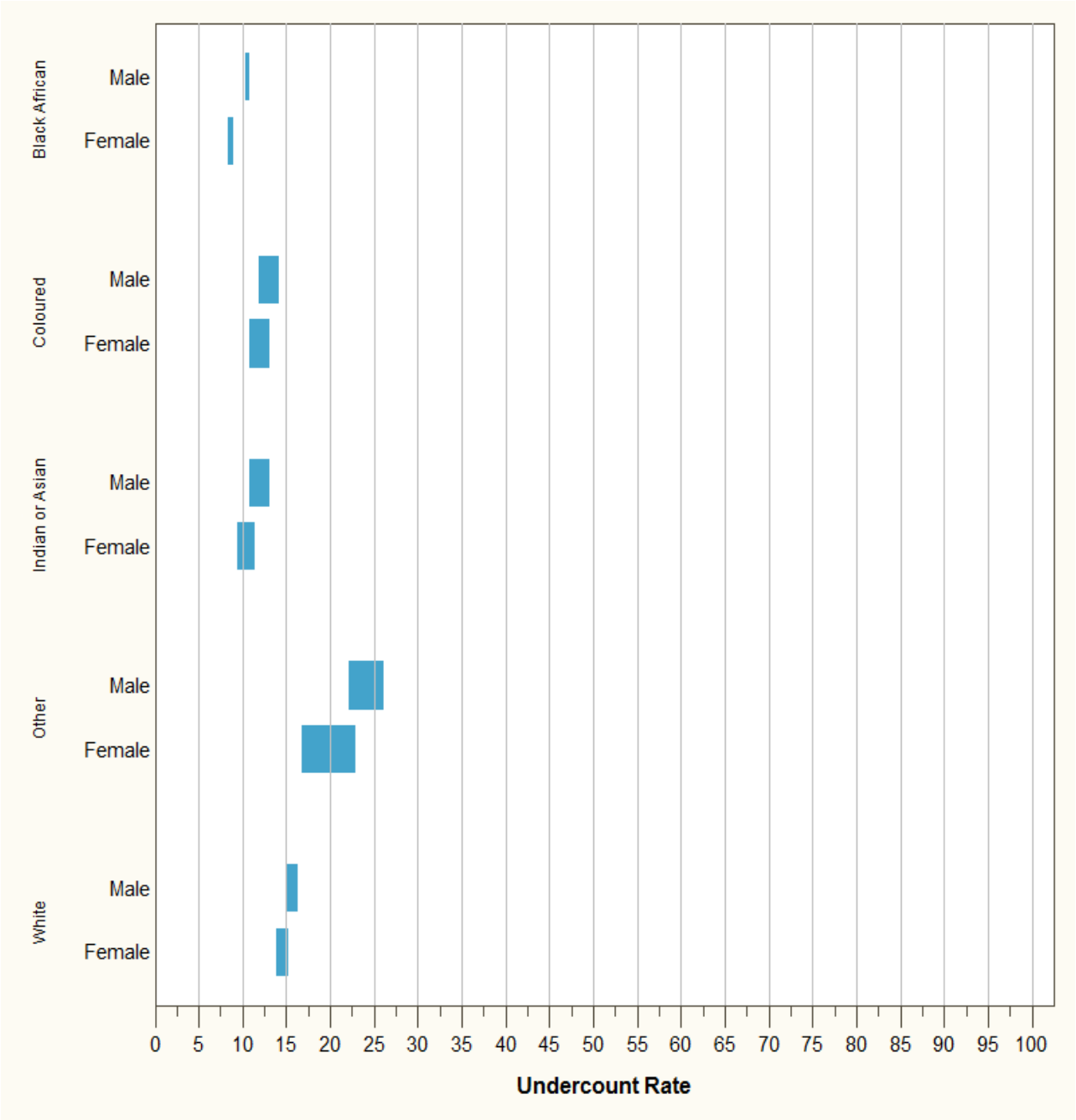
				95% confidence interval limits	
		Net undercount rate	Standard error (+ or -)	Lower	Upper
Population group by sex					
Other	Under 5 years	28,5	2,319	23,92	33,01
	5-9 years	29,3	1,427	26,47	32,06
	10-14 years	10,9	1,091	8,76	13,04
	15-19 years	15,0	1,850	11,40	18,65
	20-29 years	25,9	1,235	23,43	28,27
	30-44 years	26,3	1,225	23,93	28,73
	45-64 years	27,9	2,811	22,44	33,46
	65 or more	22,7	4,369	14,18	31,31
Sex by age group					
Male	Under 5 years	15,2	0,161	14,92	15,55
	5-9 years	11,3	0,183	10,97	11,69
	10-14 years	10,8	0,162	10,52	11,16
	15-19 years	12,8	0,169	12,48	13,14
	20-29 years	19,7	0,168	19,33	19,99
	30-44 years	19,3	0,159	19,02	19,65
	45-64 years	14,9	0,180	14,58	15,28
	65 or more	11,2	0,217	10,75	11,60
Female	Under 5 years	14,9	0,174	14,58	15,26
	5-9 years	11,5	0,185	11,14	11,86
	10-14 years	11,3	0,171	10,93	11,60
	15-19 years	12,6	0,153	12,35	12,94
	20-29 years	16,6	0,151	16,29	16,89
	30-44 years	14,5	0,137	14,28	14,82
	45-64 years	10,5	0,170	10,15	10,82
	65 or more	9,1	0,206	8,66	9,47

*Subject to rounding error
*The undercount rates for undetermined age groups are excluded from this table

Table 12 and Figures 7 to 10 will allow the reader to make many different comparisons based on population group by sex, population group by age, or sex by age. For example, the lack of a differential undercount between males and females holds across all population groups (Figure 11). The same figure also shows that the undercount for both white males and white females is significantly higher than that for all other population groups by sex. White males are significantly more undercounted than all other population group/sex groups, except white females. Finally, except for the difference between black males and coloured females as well as black males and coloured males, there are no other significant differences in undercount among the population group/sex groups.

Figure 11: Graphic representation of confidence intervals for persons undercount by population group and sex

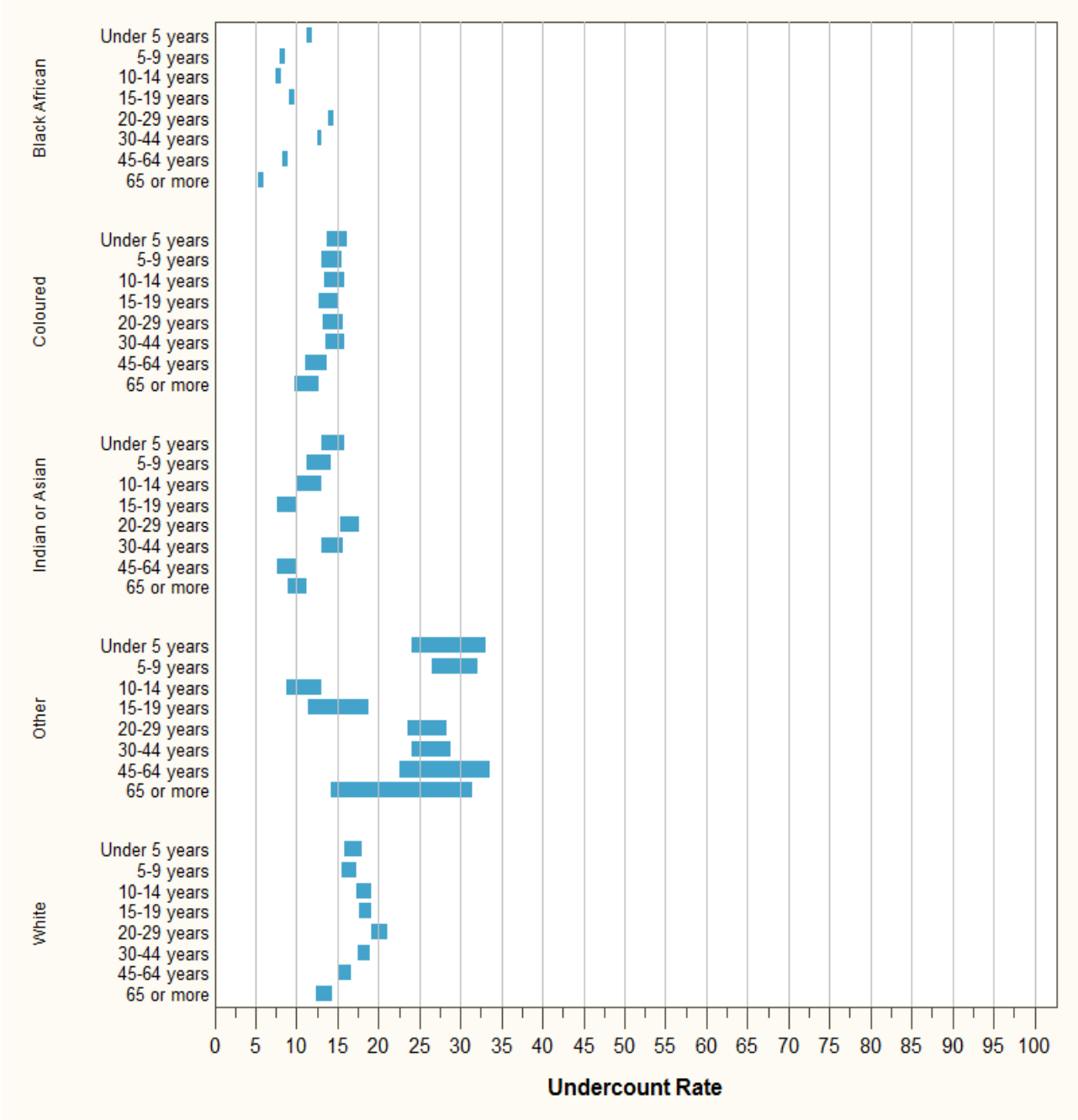
Upper and lower limits



The population group 'Other' has the highest undercount for all the variables under consideration, i.e. population group, sex and age group. The undercount for whites 20–29 years old (20%) is significantly higher than that for Indians or Asians 20–29 years (16,4%), Coloureds 20–29 years (14,4%) and black Africans 20–29 years (14, 1%). Likewise, the undercount for males 20–29 years (19,7%) is significantly higher than that for females 20–29 years (16,6%).

Figure 12: Graphic representation of confidence intervals for persons undercount by population group and by age group

Lower and upper limits



Within both males and females (Figure 13), the undercounts for the age groups 20–29 years and 30–44 years are not significantly different from each other but they are both significantly higher than for the other age groups. However, where population groups are concerned (Figure 12), the finding that the age group 20–29 years has a significantly higher undercount than all other age groups (except 30–44 years) holds for black Africans and Indians or Asians.

Figure 13: Graphic representation of confidence intervals for persons undercount by sex and age group

Lower and upper limits

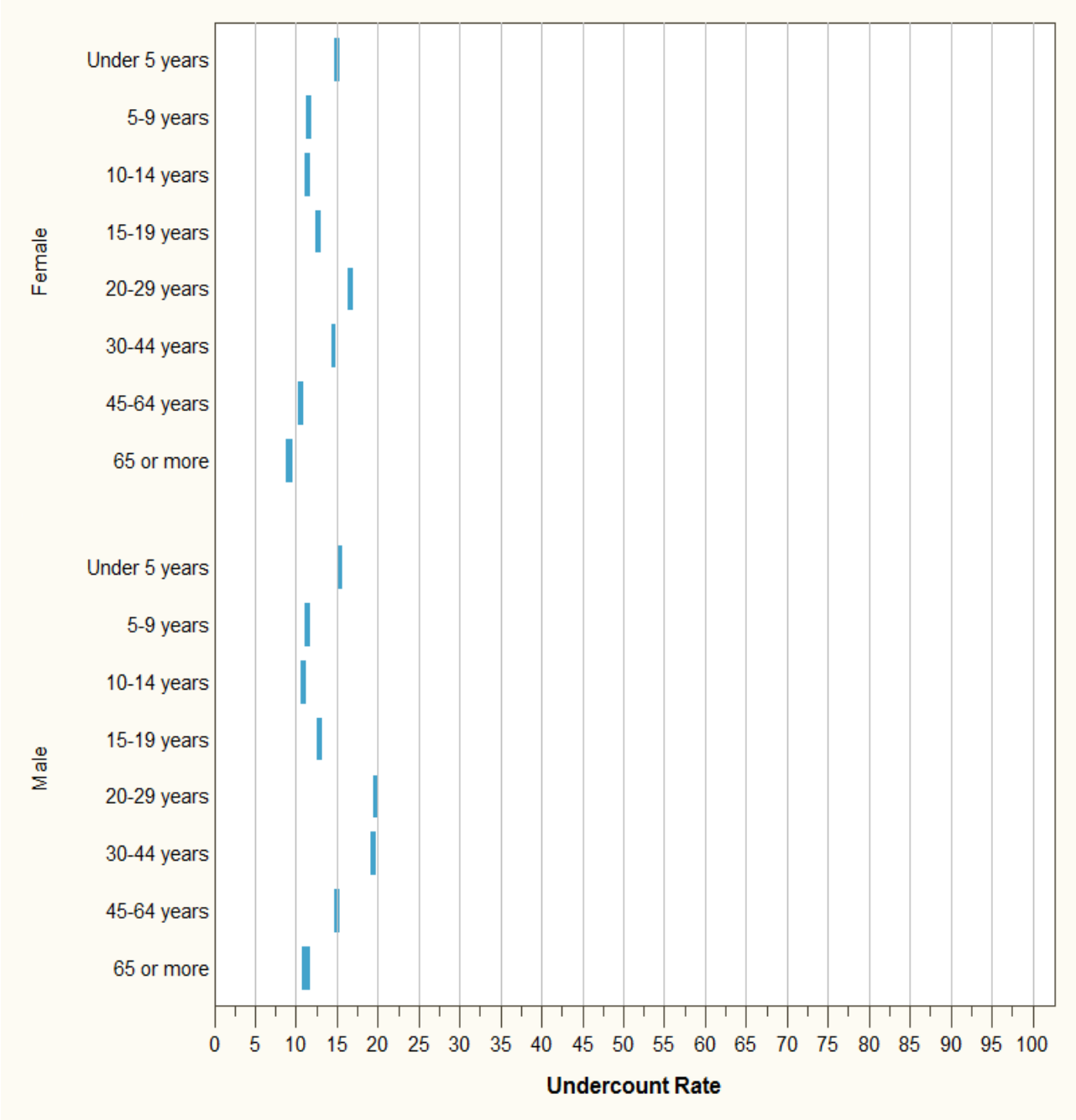


Table 13: Undercount rate for persons by demographic group – in-scope sub-universe; three-variable classifications (values expressed in percentage points rounded to one decimal)

Population group	Sex	Age group	Net undercount rate	Standard error (+ or -)	95% confidence interval limits	
					Lower	Upper
Black African	Male	Under 5 years	12,6	0,156	12,32	12,93
		5-9 years	8,9	0,202	8,52	9,32
		10-14 years	8,2	0,176	7,84	8,52
		15-19 years	10,3	0,184	9,99	10,71
		20-29 years	16,9	0,181	16,50	17,20
		30-44 years	16,4	0,177	16,02	16,71
		45-64 years	11,8	0,180	11,40	12,10
		65 or more	7,2	0,213	6,79	7,63
	Female	Under 5 years	12,6	0,169	12,26	12,92
		5-9 years	9,0	0,202	8,60	9,39
		10-14 years	8,6	0,184	8,24	8,97
		15-19 years	10,1	0,165	9,83	10,47
		20-29 years	14,0	0,161	13,64	14,28
		30-44 years	11,8	0,141	11,48	12,03
		45-64 years	7,4	0,175	7,04	7,72
		65 or more	5,6	0,208	5,20	6,02
Coloured	Male	Under 5 years	17,2	0,645	15,98	18,51
		5-9 years	15,4	0,696	14,07	16,80
		10-14 years	15,6	0,617	14,35	16,77
		15-19 years	14,3	0,653	13,05	15,61
		20-29 years	16,1	0,638	14,89	17,40
		30-44 years	17,0	0,641	15,72	18,23
		45-64 years	15,3	0,686	13,96	16,65
		65 or more	13,0	0,769	11,53	14,54
	Female	Under 5 years	15,2	0,653	13,87	16,43
		5-9 years	15,5	0,628	14,22	16,69
		10-14 years	16,0	0,675	14,71	17,35
		15-19 years	16,0	0,678	14,68	17,34
		20-29 years	15,4	0,664	14,06	16,66
		30-44 years	15,0	0,617	13,83	16,25
		45-64 years	12,0	0,635	10,79	13,27
		65 or more	11,7	0,734	10,30	13,18

Table 13: Undercount rate for persons by demographic group – in-scope sub-universe; three-variable classifications (values expressed in percentage points rounded to one decimal) (continued)

Population group	Sex	Age group	Net undercount rate	Standard error (+ or -)	95% confidence interval limits	
					Lower	Upper
Indian or Asian	Male	Under 5 years	15,3	1,181	12,96	17,59
		5-9 years	11,7	0,701	10,34	13,09
		10-14 years	12,1	0,981	10,16	14,00
		15-19 years	9,8	0,910	7,98	11,54
		20-29 years	20,4	0,643	19,15	21,67
		30-44 years	17,4	0,768	15,89	18,90
		45-64 years	10,3	0,539	9,24	11,36
		65 or more	10,7	0,846	9,02	12,33
	Female	Under 5 years	16,2	0,643	14,95	17,47
		5-9 years	16,0	1,055	13,91	18,05
		10-14 years	13,1	0,775	11,60	14,64
		15-19 years	9,5	0,828	7,90	11,15
		20-29 years	15,2	0,665	13,86	16,47
		30-44 years	13,8	0,714	12,38	15,18
		45-64 years	9,0	0,701	7,63	10,38
		65 or more	11,2	0,566	10,08	12,29
White	Male	Under 5 years	19,2	0,568	18,13	20,36
		5-9 years	17,3	0,519	16,32	18,35
		10-14 years	20,4	0,454	19,55	21,33
		15-19 years	21,1	0,485	20,13	22,04
		20-29 years	22,1	0,533	21,08	23,17
		30-44 years	20,9	0,437	20,01	21,72
		45-64 years	17,8	0,437	16,93	18,64
		65 or more	14,4	0,481	13,44	15,32
	Female	Under 5 years	17,3	0,705	15,94	18,70
		5-9 years	18,5	0,62	17,26	19,69
		10-14 years	19,1	0,619	17,88	20,31
		15-19 years	18,9	0,489	17,91	19,83
		20-29 years	21,5	0,533	20,45	22,54
		30-44 years	18,8	0,405	18,03	19,62
		45-64 years	16,8	0,393	16,07	17,61
		65 or more	14,6	0,584	13,5	15,79

Table 13: Undercount rate for persons by demographic group – in-scope sub-universe; three-variable classifications (values expressed in percentage points rounded to one decimal) (concluded)

Population group	Sex	Age group	Net undercount rate	Standard error (+ or -)	95% confidence interval limits	
					Lower	Upper
Other	Male	Under 5 years	27,5	2,528	22,52	32,43
		5-9 years	37,8	0,993	35,87	39,76
		10-14 years	14,8	1,359	12,14	17,46
		15-19 years	11,7	1,010	9,71	13,67
		20-29 years	29,8	1,251	27,32	32,23
		30-44 years	31,6	1,254	29,19	34,10
		45-64 years	36,6	2,085	32,55	40,73
		65 or more	17,1	6,726	3,93	30,30
	Female	Under 5 years	36,6	2,174	32,37	40,89
		5-9 years	26,8	2,881	21,12	32,41
		10-14 years	9,1	1,064	7,03	11,20
		15-19 years	19,9	1,458	17,01	22,73
		20-29 years	25,2	2,025	21,21	29,15
		30-44 years	23,1	1,737	19,67	26,48
		45-64 years	24,5	3,843	17,01	32,07
		65 or more	34,3	3,961	26,53	42,05

*Subject to rounding error
*The undercount rates for undetermined age groups are excluded from this table

6.3 The adjustment

The actual adjustment procedure consisted of creating homogeneous adjustment classes with similar coverage rates within province – based on geography type, population group, sex, and age group – and calculating a common population, undercount rate, and adjustment factor, for each class separately. The national adjusted population was obtained by summing the adjusted classes. Only the population within the scope of the PES received adjustment factors. The totals for the balance of population (namely, people living in collective living quarters other than hostels, the homeless on the street, and those living in out-of-scope EAs) were not adjusted (see Sections 5.5 and 5.6).

Table 14 shows the adjusted total population by geographic classifications and the corresponding confidence intervals, which reflect the sampling error around the estimate. The confidence interval was obtained by adding the unadjusted balance of population to the lower and upper limits of the confidence interval for the adjusted population in the in-scope sub-universe.

Table 14: Adjusted total population – full universe

			95% confidence interval limits	
Category	**Estimate**	**Standard error (+ or -)**	**Lower**	**Upper**
All persons	51 770 560	997 560	49 815 342	53 725 778
Province				
Western Cape	5 822 734	40 830	5 742 707	5 902 761
Eastern Cape	6 562 053	125 810	6 315 465	6 808 641
Northern Cape	1 145 861	82 466	9 842 27.6	1 307 494
Free State	2 745 590	117 567	2 515 159	2 976 021
KwaZulu-Natal	10 267 300	109 994	10 051 712	10 482 888
North West	3 509 953	166 754	3 183 115	3 836 791
Gauteng	12 272 263	106 023	12 064 458	12 480 068
Mpumalanga	4 039 939	219 299	3 610 113	4 469 765
Limpopo	5 404 868	251 244	4 912 430	5 897 306

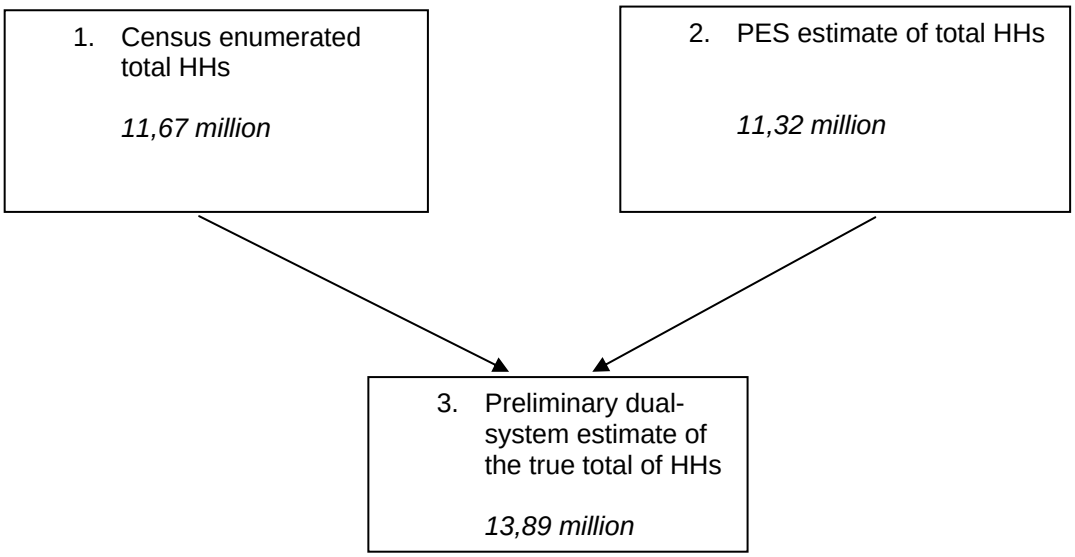
7. COVERAGE EVALUATION OF HOUSEHOLDS

7.1 Estimation of true population

The same dual-system estimation procedure described for persons in Section 5.1, and explained in detail in Section 5.3, was applied to households.

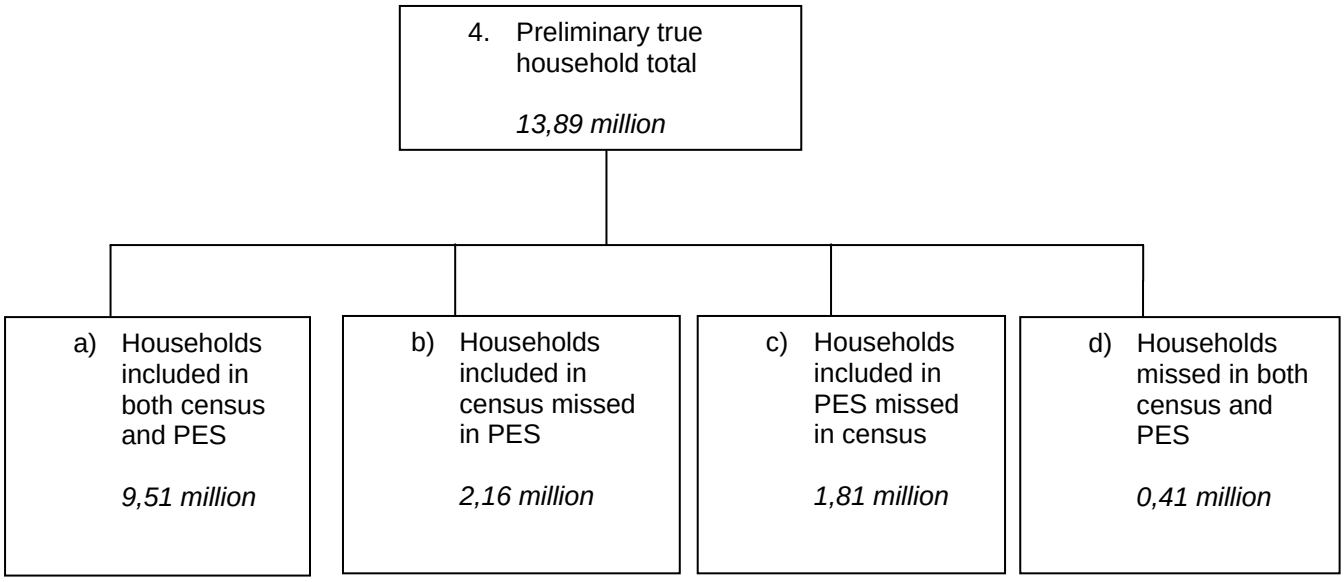
In Census 2011, a 'household' corresponds to the collection of persons in one set of questionnaires (including continuation questionnaires). 'Questionnaire' (parent questionnaire coupled with continuation questionnaires) and 'household' thus refer to the same set of persons. Even though the basic definition for household is similar in both the census and PES, there are conceptual differences because the 'questionnaire' is not a fixed entity in the universe: the number of questionnaires completed for one housing unit can vary from interview to interview, especially in *de facto* enumerations which are based on presence rather than usual place of residence.

Figure 14: Estimates of total households from individual systems and from the dual system – in-scope sub-universe



In the sub-universe in scope for the PES, the separate census and PES enumerations produced 11,32 million and 11,67 million households, respectively. Using the dual-system estimation method, the true household total of South Africa in the in-scope sub-universe was estimated at 13,89 million. Four components together make up the dual-system estimate of the true population.

Figure 15: Breakdown of dual-system estimate of household total-in-scope sub-universe



Component (a), the households included in both the census and the PES, was estimated at 9,51 million; component (b), the households included in the census but missed in the PES, was estimated at 2,16 million; component (c), the households included in the PES but missed in the census, was estimated at 1,81 million; and component (d), the households missed in both the census and the PES, was estimated at 0,41 million.

Table 15 shows that, of the 11,76 million households counted in the census for the in-scope sub-universe, 11,67 million are estimated to be correctly enumerated. Of these, the PES enumerated 9,52 million and missed 2,15 million. The census erroneous inclusions are estimated to be 0,78 million.

Table 15: Coverage distribution of Census household totals – in-scope sub-universe (in millions rounded to two decimals)

	Census enumeration
Total excluding erroneous inclusions	11,67
Included in PES	9,52
Omitted from PES	2,15
Erroneous inclusions	0,78
Insufficient information	0,15
Total including erroneous inclusions	**11,76**

*Sums are subject to rounding error

It is estimated that the census omitted 2,22 million households in total, 1,81 million of which were correctly enumerated in the PES, and another 0,41 million of which were missed in the PES as well as in the census (Table 16). This total omission does not take into account the erroneous inclusions.

Table 16: Coverage distribution of true household total – in-scope universe (in millions rounded to two decimals)

		Census enumeration		
		Included	Omitted	Total
PES Population	Included	9,51	1,81	11,32
	Omitted	2,16	0,41	2,57
Total excluding erroneous inclusions		11,67	2,22	13,89

*Sums are subject to rounding error

While the PES estimated the household total in the in-scope sub-universe at 11,32 million, it omitted 2,16 million households that were correctly enumerated in the census, and another 0,41 million that were missed in both the census and the PES, for a total omission of 2,57 million (Table 13).

The true household total, estimated at 15,07 million, was calculated by adding the census-enumerated 0,44 million households in the balance of universe to the dual-system estimate of 14,63 million in the in-scope sub-universe (Table 17).

Table 17: Adjusted census household totals – full universe (in millions rounded to two decimals)

	Households in dwelling units and hostels within in-scope EA types	Households in other collective living quarters and other EA types	Total households
Adjusted	14,63	0,44	15,07

The overall empirical probabilities of inclusion and omission of a household in the census or in the PES are shown below in Table 18. A household in the in-scope universe had approximately an 84% chance of being enumerated in the census, an 81,4% chance of being enumerated in the PES, and a 68,4% chance of being enumerated in both. Conversely, the household had approximately a 15,5% chance of being included in the census but missed in the PES, a 13% chance of being included in the PES but missed in the census, and a 3% chance of being missed in both.

Table 18: Probabilities of inclusion and omission of a household – in-scope sub-universe

Probability of being included in Census	0,83994
Probability of being included in PES	0,81481
Probability of being included in both Census and PES	0,68439
Probability of being included in Census, but missed in PES	0,15555
Probability of being included in PES, but missed in Census	0,13042
Probability of being missed in both census and PES	0,02964

7.2 Estimation of the net undercount rate

The net undercount (or overcount) is the difference between the estimated true household total and the census-enumerated household total. The rate is the net undercount expressed as a percentage of the estimated true total. Net undercount rates, together with their absolute errors and confidence intervals, are shown in Table 19 for households by geographic classification. The confidence interval is formed around the estimate by adding or subtracting the absolute error. See Section 5.2 for notes concerning confidence intervals and absolute errors, and their use in determining any differential undercounts.

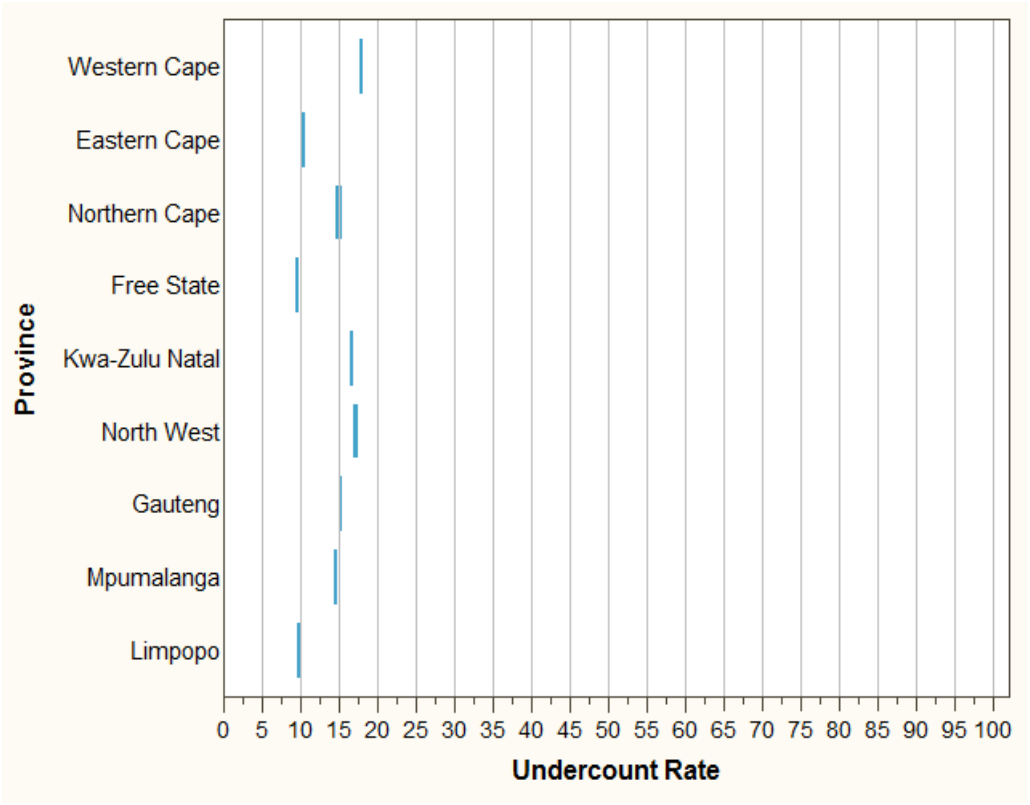
In Table 19, it can be observed that the net undercount rate for households at the national level was estimated at 14,3%.

Table 19: Net undercount rate for households by province – in-scope sub-universe (values expressed in percentage points rounded to one decimal)

			95% confidence interval limits	
Category	**Net undercount rate**	**Standard error (+ or -)**	**Lower**	**Upper**
Province				
Western Cape	17,8	0,110	17,58	18,01
Eastern Cape	10,3	0,104	10,08	10,49
Northern Cape	14,8	0,204	14,36	15,16
Free State	9,4	0,143	9,09	9,65
KwaZulu-Natal	16,5	0,096	16,32	16,70
North West	17,0	0,124	16,75	17,24
Gauteng	15,2	0,081	15,01	15,33
Mpumalanga	14,4	0,137	14,18	14,72
Limpopo	9,6	0,078	9,48	9,78
All households	14,3	0,037	14,25	14,40
			95% confidence interval limits	
Geography type	**Net undercount rate**	**Standard error**	**Lower**	**Upper**
Urban	14,8	0,045	14,74	14,92
Non-urban	13,2	0,066	13,11	13,37
All areas	14,3	0,037	14,25	14,40

*Subject to rounding error

Figure 16: Graphic representation of confidence intervals for household undercount rate, by province



Among the provinces, the highest household undercount was observed in Western Cape, North West and KwaZulu-Natal (17,8%, 17,0% and 16,5% respectively) (see Table 19 and Figure 16). There is no significant undercount difference between Western Cape and North West. The lowest undercount was observed in Free State (9,4%), which closely followed by Limpopo which had an undercount of 9,6%.

7.3 The adjustment

The adjustment procedure for households was similar to the adjustment procedure for persons. It consisted of creating homogeneous adjustment classes with similar coverage rates – based on geography type, province, household size, and population group of head of household – and calculating a common population, undercount rate and adjustment factor, for each class separately. The national adjusted household total was obtained by summing across the adjustment classes. Only the households in the in-scope sub-universe received adjustment factors. The balance of the households (i.e., in non-institutional collective living quarters other than hostels and in the out-of-scope EAs) were not adjusted (see Sections 5.5 and 5.6).

Table 20 shows the adjusted total number of households by geographic classification and the corresponding confidence intervals, which reflect the sampling error around the estimate. This estimate includes households in **housing units** only. It includes the housing units in the in-scope EAs and the out-of-scope EAs; however, it excludes hostels and other collective living quarters. The confidence interval for this estimate was obtained by adding the unadjusted households in out-of-scope housing units to the lower and upper limits of the confidence interval for the adjusted household total in in-scope housing units.

Table 20: Adjusted household total – dwelling units universe

			95% confidence interval limits	
Category	**Estimate**	**Standard error (+ or -)**	**Lower**	**Upper**
All households	15 066 999	18 177	15 031 372	15 102 626
Province				
Western Cape	1 705 919	5 319	1 695 494	1 716 344
Eastern Cape	1 756 138	6 132	1 744 119	1 768 157
Northern Cape	313 810	5 788	302 466	325 154
Free State	839 869	5 496	829 097	850 641
KwaZulu-Natal	2 635 999	3 616	2 628 912	2 643 086
North West	1 098 315	9 515	1 079 666	1 116 964
Gauteng	4 165 297	2 641	4 160 121	4 170 473
Mpumalanga	1 102 970	5 481	1 092 227	1 113 713
Limpopo	1 448 682	7 755	1 433 482	1 463 882

8. CONTENT EVALUATION – PERSONS ONLY

8.1 Nature of content analysis

Content error, also known as response error, is defined as the deviation of the obtained value from the true value for a given characteristic. Depending on whether essential or transient conditions are involved, response error can be divided into response bias (systematic error) and response variance (variable error). The PES is regarded as a replication, an independent re-interview of a sample from the census for the purpose of estimating variable error, not bias. The PES content error analysis measures **consistency**, not which answers are right or wrong, i.e. it measures how **differently** answers are reported between the census and the PES.

The following characteristics were selected for content error analysis:

- Sex
- Age group
- Relationship to head of household
- Marital status
- Population group

To ensure comparability between the PES and the census, the same wording, response categories and precodes, and also the same concept definitions, were maintained in the PES. First, estimated totals from the census and the PES, as reported in the census and as reported in the PES, are compared for matched persons for the selected characteristics. The number of cases in agreement in the universe is observable along the diagonal.

Variability between the census and the PES is then measured by means of four different indicators: the net difference rate, the index of inconsistency (simple and aggregate), the gross difference rate, and the rate of agreement. These measures and their confidence intervals are presented for the selected characteristics.

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

- **Index of Inconsistency.** The index of inconsistency is the relative number of cases for which the response varied between the census and the PES. It is the ratio of the simple response variance to the total variance of the characteristic, including its variability in the population. It is calculated for each response category.

- **Gross Difference Rate (also Off-Diagonal Proportion).** The gross difference rate (GDR) 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.** The rate of agreement is the complement of the gross difference rate. A low rate of agreement indicates a high degree of variability, and vice versa.

Table 21: Standards for the interpretation of the different content error measures

Measure	Low	Moderate	High
Index of inconsistency	<20	20–50	>50
Aggregate index of inconsistency	<20	20–50	>50
Absolute value of NDR relative to mean or proportion (NDR/P)	<0,01	0,01–0,05	>0,05

Source: 'Evaluating censuses of Population and Housing', ISP-TR-5, US Census Bureau, 1985

Important note

The estimated person totals shown in the content analysis tables do not coincide with the final census totals for each characteristic because:

- they are based only on the sample of census records in the PES and are, therefore, subject to sampling variability;
- they include only matched cases, not the full sample;
- the data are unedited, while the data in the final census totals are edited;
- they include only the in-scope sub-universe (consisting of housing units and hostels within in-scope EA types) while the final census totals include the full universe; and
- they are unadjusted while the final census totals are adjusted for coverage error.

The sole purpose of these totals is to compare the census responses with the PES responses for consistency/variability analysis purposes. They are not intended for socio-demographic analysis purposes; final census results should be used for such purposes. The data quality in the final census results is, to a certain extent, greatly improved over what the content analysis indicates due to more accurate data capturing (by automated scanning with rigorous quality control systems) and to sophisticated editing procedures.

8.2 Content analysis for sex

Is *(name)* male or female?

Table 22: Sex as reported in the census the PES

	Sex (PES)			
Sex (Census)	**Male**	**Female**	**Undetermined**	**Total**
Male	105 108	950	47	106 105
Female	904	118 717	43	119 664
Undetermined	426	411	8	845
Total	**106 438**	**120 078**	**98**	**226 614**

Table 23: Net difference rate, index of inconsistency, and gross difference rate for sex

				Net Difference Rate			Index of Inconsistency		
Response category	**Total consistent cases**	**Total in Census**	**Total in PES**	**Rate**	**95% LCL**	**95% UCL**	**Index**	**95% LCL**	**95% UCL**
Male	105 108	106 105	106 438	-0,15	-0,19	-0,11	2,06	1,98	2,15
Female	118 717	119 664	120 078	-0,18	-0,22	-0,14	2,04	1,96	2,13
Undetermined	8	845	98	0,33	0,3	0,36	98,38	92,25	104,92
Total	**223 833**	**226 614**	**226 614**	**0**	**-0,06**	**0,06**			
Aggregated Index of Inconsistency = 2,5 Gross difference Rate = 1,23 Rate of Agreement = 0,99									

The characteristic sex shows a **low** level of inconsistency or variability (index < 20%) and can be expected to be reported more or less reliably and consistently from survey to survey.

8.3 Content analysis for age group

What is *(name's)* age in completed years?

Table 24: Age group as reported in the census and the PES

Age group (Census)	Age group (PES)								
	Under 5 years	5-14 years	15-19 years	20-29 years	30-44 years	45-64 years	65 or more	Undetermined	Total
Under 5 years	22 752	1 004	43	45	44	27	13	192	24 120
5-14 years	790	43 156	925	180	35	12	9	418	45 525
15-19 years	26	597	21 193	726	28	6	7	185	22 768
20-29 years	21	190	654	38 026	934	45	14	473	40 357
30-44 years	7	34	41	794	40 201	978	41	544	42 640
45-64 years	4	6	7	61	887	34 779	513	447	36 704
65 or more	7	17	15	33	65	488	12 263	142	13 030
Undetermined	140	224	133	258	300	258	116	41	1 470
Total	**23 747**	**45 228**	**23 011**	**40 123**	**42 494**	**36 593**	**12 976**	**2 442**	**226 614**

Table 25: Net difference rate, index of inconsistency, and gross difference rate for age group

				Net Difference Rate			Index of Inconsistency		
Response category	Total consistent cases	Total in Census	Total in PES	Rate	95% LCL	95% UCL	Index	95% LCL	95% UCL
Under 5 years	22 752	24 120	23 747	016	0,12	0,21	5,52	5,3	5,75
5-14 years	43 156	45 525	45 228	0,13	0,07	0,19	6,12	5,94	6,3
15-19 years	21 193	22 768	23 011	-0,11	-0,16	-0,06	8,24	7,97	8,53
20-29 years	38 026	40 357	40 123	0,1	0,05	0,16	6,69	6,5	6,89
30-44 years	40 201	42 640	42 494	0,06	0	0,12	6,84	6,65	7,04
45-64 years	34 779	36 704	36 593	0,05	0	0,1	6,09	5,89	6,28
65 or more	12 263	13 030	12 976	0,02	-0,01	0,06	6,04	5,74	6,35
Undetermined	41	1 470	2 442	-0,43	-0,48	-0,38	98,7	95,63	101,88
Total	**212 411**	**226 614**	**226 614**	**0**	**-0,15**	**0,15**			
Aggregated Index of Inconsistency = 7,4 Gross Difference Rate = 6,27 Rate of Agreement = 0,94									

In both the PES and the census, age was derived from the date of birth; in other words, age derived from the date of birth was preferred over reported age when the two were inconsistent. The characteristic age (as derived) shows a **low** level of inconsistency or variability (index < 20%) and can be expected to be reported more or less reliably and consistently from survey to survey.

8.4 Content analysis for population group

How would *(name)* describe him/herself in terms of population group?

Table 26: Content analysis for population group

Population group (Census)	Population group (PES)						Total
	Black African	Coloured	Indian or Asian	White	Other	Undeter-mined	
Black African	174 127	808	211	62	102	98	175 408
Coloured	700	22 941	143	221	89	20	24 114
Indian or Asian	60	156	5 696	16	44	20	5 992
White	79	149	15	15 353	35	23	15 654
Other	292	114	74	52	129	0	661
Undetermined	3 620	650	116	370	26	3	4 785
Total	**178 878**	**24 818**	**6 255**	**16 074**	**425**	**164**	**226 614**

Table 27: Net difference rate, index of inconsistency, and gross difference rate for population group

Response category	Total consistent cases	Total in Census	Total in PES	Net Difference Rate			Index of Inconsistency		
				Rate	95% LCL	95% UCL	Index	95% LCL	95% UCL
Black African	174 127	175 408	178 878	-1,53	-1,6	-1,46	7,8	7,6	8
Coloured	22 941	24 114	24 818	-0,31	-0,36	-0,26	6,99	6,74	7,24
Indian or Asian	5 696	5 992	6 255	-0,12	-0,14	-0,09	7,18	6,71	7,67
White	15 353	15 654	16 074	-0,19	-0,21	-0,16	3,46	3,26	3,68
Other	129	661	425	0,1	0,08	0,13	76,42	71,39	81,8
Undetermined	3	4 785	164	2,04	1,98	2,1	100,02	97,27	102,85
Total	**218 249**	**226 614**	**226 614**	**0**	**-0,11**	**0,11**			
Aggregated Index of Inconsistency = 9,9 Gross Difference Rate = 3,69 Rate of Agreement = 0,96									

The characteristic 'population group' also exhibits a low degree of inconsistency and variability among the characteristics measured. It seems to be quite robust and reliable from one measurement to another. At the individual response category level, the two categories 'other' and 'undetermined' do show great inconsistency from census to PES. However, they only occur in a few cases and in the final census results they are edited out.

8.5 Content analysis for relationship to head of household

What is *(name's)* relationship to the head or acting head of the household?

Table 28: Content analysis for relationship to head of household

Relationship to the Head of the Household (Census)	Relationship to the Head of the Household (PES)						
	Head/ Acting Head	Husband/ Wife/ Partner	Son/ Daughter	Adopted Son/ Daughter	Stepc hild	Brother/ Sister	Parent (Mother/ Father)
Head/Acting Head	53 980	3 450	937	6	12	757	346
Husband/Wife/Partner	4 358	21 840	652	3	5	120	47
Son/Daughter	866	373	68 726	168	556	1 137	65
Adopted Son/Daughter	14	4	316	103	16	11	.
Stepchild	16	4	569	7	152	24	.
Brother/Sister	766	105	1 177	7	18	3 617	13
Parent (Mother/Father)	431	97	553	.	11	24	378
Parent-in-law	50	27	44	1	.	17	51
Grand/ Great Grandchild	112	23	3 047	49	47	163	16
Son/ Daughter-in-law	80	95	285	5	4	40	4
Brother/Sister-in-law	84	44	120	3	3	199	1
Grandmother/Father	74	10	328	2	2	19	57
Other relative	368	128	1,342	64	40	461	31
Non-related person	310	71	125	31	10	52	8
Undetermined	472	177	796	7	9	89	15
Total	**61 981**	**26 448**	**79 017**	**456**	**885**	**6 730**	**1 032**

Table 29: Content analysis for relationship to head of household (continued)

Relationship to the Head of the Household (Census)	Relationship to the Head of the Household (PES)								
	Parent-in-law	Grand/ Great Grand-child	Son/ Daughter-in-law	Brother/ Sister-in-law	Grand-mother/ Father	Other relative	Non-related person	Un-deter-mined	Total
Head/Acting Head	37	125	71	46	39	429	228	18	60481
Husband/Wife/ Partner	12	54	91	24	8	132	71	7	27 424
Son/Daughter	20	2 977	128	53	43	2 202	160	37	77 511
Adopted Son/Daughter	1	170	3	.	1	182	43	.	864
Stepchild	2	53	11	2	5	111	5	.	961
Brother/Sister	5	169	9	169	4	801	91	6	6 957
Parent (Mother/Father)	70	52	10	8	27	75	5	.	1 741
Parent-in-law	126	32	65	4	13	62	9	.	501
Grand/Great Grandchild	10	27 532	51	12	163	2 146	67	17	33 455
Son/Daughter-in-law	51	199	703	72	2	284	59	3	1 886
Brother/Sister-in-law	3	23	77	218	2	215	37	3	1 032
Grandmother/ Father	22	1 687	1	2	43	175	3	7	2 432
Other relative	31	1 063	77	111	20	3 473	339	9	7 557
Non-related person	4	57	21	14	2	273	615	2	1 595
Undetermined	14	389	20	8	4	172	39	6	2 217
Total	**408**	**34 582**	**1 338**	**743**	**376**	**10 732**	**1 771**	**115**	**226 614**

Table 30: Net difference rate, index of inconsistency, and gross difference rate for relationship to head of household

				Net Difference Rate			Index of Inconsistency		
Response category	**Total consistent cases**	**Total in Census**	**Total in PES**	**Rate**	**95% LCL**	**95% UCL**	**Index**	**95% LCL**	**95% UCL**
Undetermined	6	2 217	115	0,93	0,89	0,97	99,58	95,61	103,72
Head/Acting Head	53 980	60 481	61 981	-0,66	-0,77	-0,56	16,23	15,96	16,49
Husband/Wife/Partner	21 840	27 424	26 448	0,43	0,34	0,52	21,47	21,06	21,89
Son/Daughter	68 726	77 511	79 017	-0,66	-0,78	-0,55	18,62	18,35	18,88
Adopted Son/Daughter	103	864	456	0,18	0,15	0,21	84,62	79,79	89,73
Stepchild	152	961	885	0,03	0	0,07	83,87	79,79	88,17
Brother/Sister	3 617	6 957	6 730	0,1	0,03	0,17	48,61	47,44	49,82
Parent(Mother/Father)	378	1 741	1 032	0,31	0,27	0,35	73,16	70,03	76,42
Parent-in-law	126	501	408	0,04	0,02	0,06	72,42	67,09	78,17
Grand/Great Grandchild	27 532	33 455	34 582	-0,5	-0,6	-0,4	22,43	22,05	22,82
Son/Daughter-in-law	703	1 886	1 338	0,24	0,2	0,28	56,78	54,23	59,45
Brother/Sister-in-law	218	1 032	743	0,13	0,1	0,16	75,73	71,78	79,89
Grandmother/Father	43	2 432	376	0,91	0,86	0,95	97,22	93,63	100,94
Other relative	3 473	7 557	10 732	-1,4	-1,49	-1,31	64,55	63,37	65,75
Non-related person	615	1 595	1 771	-0,08	-0,12	-0,04	63,93	61,28	66,7
Total	**181 512**	**226 614**	**226 614**	**0**	**-0,26**	**0,26**			
Aggregated Index of Inconsistency = 25,9 Gross Difference Rate = 19,9 Rate of Agreement = 0,8									

The characteristic 'relationship to head of household' shows a **moderate** level of inconsistency or variability (20% < index < 50%). It may not be reported consistently from survey to survey. In the case of Census 2011 and PES 2011, the inconsistencies are most likely due to the fact that the person referred to as head was not necessarily the same in both cases. Given the *de facto* enumeration rule (based on presence), the person could have been the 'acting head' in the absence of the normal head of household. It may not be possible to ensure more consistent responses for this variable in future surveys unless a *de jure* rule (usual residence) is used. With a *de jure* rule, the head of household generally remains the same, even when temporarily absent from the household.

8.6 Content analysis for marital status

What is *(name's)* PRESENT marital status?

Table 31: Content analysis for marital status

Marital status (Census)	Marital status (PES)								Total
	Un-deter-mined	Married	Living toge-ther	Never married	Widower/ widow	Sepa-rated	Divor-ced	Not applica ble	
Undeter-mined	29	661	228	13 385	197	19	58	151	14 728
Married	22	39 753	2 302	2 282	1 425	280	165	22	46 251
Living together	3	2 560	9 141	2 108	81	46	89	8	14 036
Never married	85	2 960	3 175	130 216	1 096	428	525	280	138 765
Widower/ widow	3	1 153	118	1 003	6 328	82	159	12	8 858
Separated	0	213	32	429	76	324	139	3	1 216
Divorced	2	205	133	450	186	150	1 620	3	2 749
Not applicable	0	2	.	8	.	.	.	1	11
Total	**144**	**47 507**	**15 129**	**149 881**	**9 389**	**1 329**	**2 755**	**480**	**226 614**

Table 32: Net difference rate, index of inconsistency, and gross difference rate for marital status

				Net Difference Rate			Index of Inconsistency		
Response category	Total consistent cases	Total in Census	Total in PES	Rate	95% LCL	95% UCL	Index	95% LCL	95% UCL
Undetermined	29	14 728	144	6,44	6,33	6,54	99,74	98,14	101,35
Married	39 753	46 251	47 507	-0,55	-0,66	-0,45	19,16	18,85	19,48
Living together	9 141	14 036	15 129	-0,48	-0,57	-0,39	39,88	39,14	40,63
Never married	130 216	138 765	149 881	-4,91	-5,05	-4,76	26,85	26,54	27,16
Widower/widow	6 328	8 858	9 389	-0,23	-0,3	-0,17	31,92	31,1	32,77
Separated	324	1 216	1 329	-0,05	-0,09	-0,01	74,96	71,66	78,41
Divorced	1 620	2 749	2 755	0	-0,04	0,04	41,64	39,96	43,39
Not applicable	1	11	480	-0,21	-0,23	-0,19	99,6	91,16	108,83
Total	187 412	226 614	226 614	0	-0,24	0,24			
Aggregated Index of Inconsistency = 31,7 Gross Difference Rate = 17,3 Rate of Agreement = 0,83									

The characteristic 'marital status' shows a moderate level of inconsistency or variability (20% < index < 50%). It may not be reported consistently from survey to survey. At the level of each response category, most categories show an even **greater** degree of inconsistency.

8.7 Summary of content error analysis

The characteristics 'population group', 'age' and 'sex' show a low level of variability between the census and the PES. They can be expected to be measured reliably from survey to survey. The variables 'relationship to head of household', 'marital status' (combining categories)' show a moderate level of variability, which might be indicative of a need for clearer concept definitions and wording, and more probing.

Table 33: Characteristics ranked from lowest to highest

Characteristic	Aggregated index of consistency	Interpretation
Sex	2,5	Low
Age group	7,4	Low
Population group	9,9	Low
Relationship to head of household	25,9	Medium
Marital status	31,7	Medium

9. CONCLUSIONS AND RECOMMENDATIONS

This chapter gives a summary of the lessons learnt during the execution of the 2011 PES as well as recommendations for future censuses and surveys.

9.1 Lessons learnt

- The fact that the PES was an integral part of the Census 2011 Programme ensured that the PES was able to conduct extensive planning and allowed for the testing of methodologies and procedures well in advance of the main PES.
- The operational timelines must be aligned with census activities to allow for thorough and timely completion of PES activities.

9.2 Recommendations

- The EA type 'Collective Living Quarters' should be decomposed, i.e. Workers' Hostels and other institutions should be separated (for sampling purposes for PES and household surveys).
- Large EAs should be split for sampling and administrative purposes (i.e. appointment of staff and workload allocation).
- Listing during census should include the collection of household information (for each DU).
- The timing of listing should be reviewed (the lag between listing and enumeration should be shortened to minimise changes within EA).
- Quality assurance should be strengthened and intensified during listing.
- There should be a structured way for checking completion of both listing and enumeration (e.g. close-off checklist).
- Monitoring of progress reports (from fieldwork) should be strengthened to allow for intervention if and when required.
- Publicity for PES should be included in census publicity campaign.
- Funding for PES should be improved to allow for appointment of a larger core team.

ANNEXURE A: CONCEPTS AND DEFINITIONS

CONCEPT	DEFINITION
Born after the Census	Babies who were present in the household on the PES reference night (6–7 November), but who were not yet born as of midnight between 9 and 10 October (census reference night).
Census reference night	Census reference night is the midnight between **09 and 10 October**. It is in the PES questionnaire.
Dwelling	Any structure intended or used for human habitation.
Dwelling unit	Structure or part of a structure or group of structures occupied or meant to be occupied by one or more than one household.
Enumeration	Enumeration is the process of counting all the members of a defined population and collecting demographic and other information about each person. This counting takes place by means of administering a PES questionnaire to all households in the sampled EA.
Enumeration area	An enumeration area (EA) is the smallest geographical unit (piece of land) into which the country is divided for census 2011 enumeration purposes. Each EA is expected to have clearly defined boundaries.
EA number	The EA number is a unique ID number given to an EA for record-keeping and coding purposes. The first digit indicates the province. The next two digits indicate the municipality and the last five distinguish among the different EAs within the municipality.
EA Summary Book	The EA Summary Book is a register of mapping and listing information pertaining to a particular EA. The summary book identifies an EA by province, local municipality, main place name and sub-place name. Maps and/or aerial photographs of the EA and adjoining areas are provided.
Final matching	Assigning the final match status based on reconciliation visits outcomes.
Final matching for estimation	Assigning the final match status for estimation for all RV and non-RV cases for persons.
Household	A household is a group of people who live together and provide themselves jointly with food or other essentials for living, or a single person who lives alone. For PES and census purposes, only people present in the household on the reference nights (census and PES) are included as part of the household.
Head/Acting head of the household	In the first instance, the head/acting head of household is the person that the household regards as such. If necessary, the head can be defined as the main decision-maker, or alternatively, the person who owns or rents the dwelling, or the person who is the main breadwinner. The head can be either male or female. If two people are equal decision-makers, the older of the two should be named as head of the household. In a household of totally unrelated persons, the oldest should be named as the household head.
Initial matching	Use information from the PES and census questionnaires to assign the initial match status.

CONCEPT	DEFINITION
In-movers	Persons who were present in the household on the PES reference night (6–7 November), but who were absent on the census reference night (9–10 October).
Matching	The process of matching entails the comparison of PES and census household and person information in order to check whether everyone was enumerated during census.
Moving status	An indication of a household or a person's presence based on the PES and census reference nights.
Non-movers	Persons who were present in the household on the PES reference night (6–7 November), and who were also present on the census reference night (9–10 October).
Out-movers	Persons who were absent from the household on the PES reference night (6–7 November), but who were present on the census reference night (9–10 October).
Out of scope	Persons who are not non-movers, in-movers or out-movers.
PES reference night	PES reference night is the midnight between **06 and 07 November**. It is in the PES questionnaire.
Primary matching variables	The main variables used to determine a match status for persons; i.e. date of birth, age, sex and population group.
Secondary matching variables	These variables are not reliable enough to determine a match status for persons (see primary matching variables). This includes relationship to head of household and marital status.
Splitting	The process of sub-dividing the large EAs into manageable portions that are equivalent to normal EAs (150<DUs).
Sub-EA	This is a smaller manageable area which was split from a large EA for easier fieldwork management.

ANNEXURE B: PES QUESTIONNAIRE

Statistics South Africa
your leading partner in quality statistics

HOUSEHOLD QUESTIONNAIRE
FOR STATISTICAL USE ONLY
Post Enumeration Survey

PES

CENSUS 2011

STATISTICS ACT NO. 6 OF 1999 (CONFIDENTIALITY)

17(1) Despite any other law, no return or other information collected by Statistics South Africa for the purposes of official or other statistics that relates to an individual or a household may be disclosed to any person.

17(3b) Any person who is involved in the collection of, or who may use, that information or data, must first take an oath of confidentiality.

18(1a) Any officer of Statistics South Africa who willfully discloses any data or

18(1g) information obtained in the course of such employment to a person not authorised to receive that information is guilty of an offence and liable on conviction to a fine not exceeding R10 000, or to imprisonment for a period not exceeding 6 months or to both.

FOR OFFICE USE ONLY

ENUMERATION AREA NUMBER

Province ..... Local municipality .....

Main place ..... Sub-place .....

Physical identification of the dwelling unit .....

.....

Postal code Landline/Cell phone of enumerated household .....

PARTICULARS OF THE HOUSEHOLD

Dwelling unit number (DU)

Household number (HH)

Total number of households at this dwelling

Map reference number

Listing record number

Total number of persons in the household — Males Females Total

Questionnaire of completed for this household

If more than one questionnaire is used in the household, write the barcode of the 1st questionnaire below:

METHOD OF QUESTIONNAIRE COMPLETION - *Mark the appropriate circle with an X*

A fieldworker through an interview

A household member through self-completion

FIELD STAFF

Fieldworker ID No.

Supervisor ID No.

Interview Date D D M M Y Y Y Y

Date Checked D D M M Y Y Y Y

Signature .....

Signature .....

Fieldwork Coordinator ID No.

Quality Assurance Monitor ID No.

Date Checked D D M M Y Y Y Y

Date Checked D D M M Y Y Y Y

RESPONSE DETAILS

		Interview			Next Visit (Planned)	
Visit No.	Date (actual)	Start Time	End Time	Result Code	Date	Time
1						
2						
3						
4						

RESULT CODE	RESPONSE DETAILS
11	Completed
12	Partly completed
21	Non-contact
22	Refusal
31	Unoccupied
32	Vacant
33	Demolished
34	New dwelling under construction

Comments and full details of all non-response/unusual circumstances

.....

.....

FINAL RESULT CODE

H-01 Was this household living in this dwelling on the night of 9-10 October?

1 = Yes → *Go to P-01*

2 = No

Write the appropriate code in the box.

H-02 When did the household move to this dwelling?

D D M M Y Y Y Y

SHOULD YOU ENCOUNTER ANY DIFFICULTIES IN THE COMPLETION OF THE QUESTIONNAIRE, PLEASE CONTACT: .....

ON ..... **OR PHONE THE CENSUS HOTLINE, TOLL FREE, ON 0800 110 248**

Census 2011 - PES© Statistics South Africa, February 2011

P 00000001 00

P0C

FLAP: PARTICULARS OF ALL INDIVIDUALS

(a) Please tell me the name and surname of the head/acting head of the household and of each person who was present in the household on midnight between 6-7 November including members who were absent on 6-7 November but were working on night shift or travelling back home.

(b) In addition, tell me the names and surnames of any persons who did not spend the night between 6-7 November, but who spent the night between 9-10 October in this household.

Always start with the head or acting head of the household i.e. **person number 1.** The **head or acting head** is the person who is the main decision-maker in the household. If two or more people are equal decision-makers, take the oldest person first.

For babies with no name, write BABY.

P-00 PERSON NUMBER	**P-01** PERSON NAME
Write 0 or 1 in the first box for all persons listed on the flap. Example: Row 1 0 1 Row 10 1 0	**Example** E R I C M O K O E N A
1	
2	
3	
4	
5	
6	
7	
8	
9	
0	

Census 2011 • PES© Statistics South Africa, February 2011

P0F

PROCEDURES OF ENUMERATION

Who should be the respondent?

- *The head/acting head of the household.*
- *In the absence of head/acting head, any responsible adult member left in charge of the household.*
  Note:
  *A **household** is a group of persons who live together, and provide themselves jointly with food or other essentials for living, or a single person who lives alone.*
- *Domestic workers are counted as a separate household even if they live in the same dwelling as the employer.*

Who should be counted in this questionnaire?

- *All persons present in the household on the PES reference night (midnight 6-7 November 2011) and/or census reference night (midnight 9-10 October 2011).*
- *Include babies born before the reference night as well as visitors.*
- *Members who died after the PES or census reference night must be counted as alive.*
- *Include small children, old people and the bed-ridden who were present in the household on either the PES or census reference night.*
- *Members of the household who were absent overnight, for example working, travelling, or at an entertainment venue, religious gathering, if they returned to the household the next day.*
- *Individuals in converted and unconverted workers hostels, residential hotels and old age homes (depending on arrangement).*

P 00000001 00

How to complete the questionnaire

- ***Read** every question carefully.*
- *Make sure that all the codes are written inside the boxes.*
  For example:

  [2] *Correct* [2] *Incorrect*

- *For numeric values, such as age, person number, etc. the enumerator/respondent should write the correct answer in the box and include leading zeros. For example:*

  [0][0][9]

- *For open-ended questions, the enumerator/respondent should write legibly in CAPITAL LETTERS in the boxes provided with no spaces between the words. For example East London should be written as:*

  [E][A][S][T][L][O][N][D][O][N]

- *Do not write zeros in boxes where questions are not applicable.*

What to use when completing this questionnaire?

Use only a pencil. If you make a mistake, use a soft rubber to erase the mistake and write the correct answer.

PARTICULARS OF ALL INDIVIDUALS IN THE HOUSEHOLD - ASK OF EVERYONE LISTED ON THE FLAP

P-02 PRESENCE ON PES NIGHT	**P-03** PRESENCE ON CENSUS NIGHT	**P-04** COUNTED	**P-05** WHERE COUNTED	**P-06** DATE OF BIRTH
Where did *(name)* **spend the night between 6-7 November?** 1 = In this dwelling 2 = Elsewhere 3 = Deceased *Write the appropriate code in the box.*	**Where did** *(name)* **spend the night between 9-10 October?** 1 = In this dwelling 2 = Elsewhere 3 = Unborn → ***Go to P-06*** *Write the appropriate code in the box.*	**Was** *(name)* **counted during Census?** 1 = Yes 2 = No 3 = Do not know → ***Go to P-06*** *Write the appropriate code in the box.*	**Where was** *(name)* **counted?** 1 = In this dwelling 2 = Elsewhere *Write the appropriate code in the box.*	**What is** *(name's)* **date of birth?** **Example** 1 7 0 4 2 0 0 4 *For any information that is not known leave the box blank.*
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y
				D D M M Y Y Y Y

Census 2011 • PES© Statistics South Africa, February 2011

P01

PARTICULARS OF ALL INDIVIDUALS IN THE HOUSEHOLD - ASK OF EVERYONE LISTED ON THE FLAP

P-07 AGE IN COMPLETED YEARS	**P-08** SEX	**P-09** RELATIONSHIP	**P-10** MARITAL STATUS	**P-11** POPULATION GROUP
What is *(name's)* **age in completed years?** *If age is not known, ask for an estimate of age. If no one is able to estimate, write 998.* *For babies less than 1 year write 000 for age. For a person 7 years and 6 months write 007 for age.* **Example** 0 0 7	**Is** *(name)* **male or female?** *Mark the appropriate circle with an X.* (X) 1 = Male ( ) 2 = Female	**What is** *(name's)* **relationship to the head or acting head of the household?** *The head or acting head is the person listed in row 1 of the first questionnaire, if more than one questionnaire has been completed for this household.* 01 = Head/Acting Head 02 = Husband/Wife/Partner 03 = Son/Daughter 04 = Adopted Son/Daughter 05 = Stepchild 06 = Brother/Sister 07 = Parent (Mother/Father) 08 = Parent-in-law 09 = Grand/Great Grandchild 10 = Son/Daughter-in-law 11 = Brother/Sister-in-law 12 = Grandmother/Father 13 = Other relative 14 = Non-related person *Write the appropriate code in the boxes.*	**What is** *(name's)* **PRESENT marital status?** 1 = Married 2 = Living together like married partners 3 = Never married 4 = Widower/widow 5 = Separated 6 = Divorced 7 = Not applicable *Write the appropriate code in the box.*	**How would** *(name)* **describe him/herself in terms of population group?** 1 = Black African 2 = Coloured 3 = Indian or Asian 4 = White 5 = Other *Write the appropriate code in the box.*
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			
	1 = Male 2 = Female			

Census 2011 - PES® Statistics South Africa, February 2011

P 00000001 00

P02

PLACE STICKER HERE

If the loose sticker is not available, copy the number of the pasted sticker here:

END OF INTERVIEW

THANK YOU FOR YOUR CO-OPERATION

Additional comments and full details of all non-response/unusual circumstances

Census 2011 - PES[R] Statistics South Africa, February 2011

P03