Statistics South Africa
Private Bag X44
Pretoria 0001
South Africa

274 Schoeman Street
Pretoria

Users enquiries: (012) 310-8600
Fax: (012) 310-8500
Main switchboard: (012) 310-8911
Fax: (012) 322-3374

E-mail: info@statssa.pwv.gov.za
Website: http://www.statssa.gov.za

The people of South Africa

Population census, 1996

Calculating the undercount in Census '96

Statistics South Africa
1998

Dr FM Orkin
Head

Published by Statistics South Africa
Private Bag X44
Pretoria
0001

ISBN 0-621-28080-1
Report no. 03-01-18 (1996)

Notes: In September 1998, the Central Statistical Service (CSS) changed its name to Statistics South Africa (Stats SA). It is referred to by the new name throughout this report.

The data contained in this report were gathered in October 1996. Since then, there have been demographic changes in South Africa associated, *inter alia*, with internal and external migration, and population growth. This means that population profiles may have changed at differing geographic levels. Stats SA is not responsible for any damages or losses, arising directly or consequently, which might result from the application or use of the data gathered as part of the 1996 population census.

The data have been evaluated and reviewed by a task team of the Interim Statistics Council. The Council's statement and the task team's report are available from the Statistics Council through Stats SA.

Census '96 was statistically adjusted for undercount on the basis of a nationwide post-enumeration survey, instead of being brought into conformity with a population-projection model and demographic assumptions reaching back 20 years, as was the case with the 1991 census. Consequently, Census '96 results differ in some respects from those based on the projections of the model previously used, including:

- the population size;
- the age distribution;
- the implied extent of international migration; and
- the relative size of population groups and provinces.

Statistics South Africa Library Cataloguing-in-Publication (CIP) Data
The people of South Africa : population census 1996 , calculating the undercount in Census '96 / Statistics South Africa. - Pretoria : Statistics South Africa, 1998
p. cm. [Report No 03-01-18 (1996)]
ISBN 0-621-28080-1
1. Population - Statistics. 2. Methodology. 3. Statistics South Africa. 4. Census - South Africa. 5. Census - undercount.
I. Statistics South Africa. II. Title. III. Series
LCSH-16

Contents

Page

Tables and figures

Page

Introduction

During October 1996, the first census of the new South Africa was conducted. Every person present in South Africa on census night (9-10 October) should, in principle, have been enumerated on a census questionnaire between 10 and 31 October. However, it is inevitable that some people will, in practice, not be reached by any census.

In a regionally diverse, unevenly skilled, ethnically heterogeneous country like South Africa, and with a degree of political suspicion towards the new government in different areas or among different groups, the proportions of people not enumerated will vary in intricate ways. The challenges were heightened in Census '96 by the short time (slightly more than a year) available for planning and implementation of the fieldwork.

The post-enumeration survey (PES) was conducted immediately after the census on a countrywide basis to measure the proportion of the population not reached in Census '96. This was the first time that a post-enumeration survey of this scope and nature was conducted in South Africa. This report details the methodology of the PES, notes some of the problems encountered and how they were addressed, and presents the estimates of undercount.

1. Undercount in the census

There are a number of reasons why a person or dwelling may have been not reached in Census '96. Some of these would be inevitable in any exercise of the magnitude of the population census. Other reasons reflect the fact that this was the first census of the new South Africa, covering the whole of the country, conducted with little time to prepare. People may have been missed because:

- They moved around during the period of the census and were difficult to contact.
- They mistakenly thought that they were included by the informant in another household.
- They were not included by the householder completing the questionnaire who may have thought that, for example, young babies need not be included in the census.
- They were concerned about the confidentiality of their data and declined to be interviewed or to fill in the questionnaire.
- They were concerned about security and denied access to enumerators (particularly in some more affluent urban areas).
- They were on a farm where the enumerator encountered difficulty gaining access, particularly in remote areas or where farmers were not co-operative.
- Their dwelling was missed by the enumerator.
- The area they lived in was not demarcated.

Information on the undercount in Census '96 has two main purposes. First, the data is used to adjust the census count to produce estimates of the population of South Africa at the time of the census. Second, it is used to evaluate the census and identify improvements for the next census, particularly with respect to segments of the population subject to higher levels of undercount.

The final calculation of the undercount of *persons*, based on a detailed analysis of the PES, including a sophisticated process of matching and imputation, is described below. This indicated that 10,7% of the people in South Africa on the night of 9-10 October 1996 were not enumerated. The preliminary calculation of the undercount using an elementary analysis of the PES, was 6,8% (see Section 3.2).

The table overleaf presents the final undercount rates for each province. The 'undercount rate' is defined as the difference between the final estimate and the raw census count expressed as a percentage of the *final* count. Note that the raw census count used in that obtained after the completion of processing and differs from that estimated for the preliminary population estimates (35 296 000). The change in the raw census count was highest in Northern Province, where administrative problems affected the accuracy of the original estimate.

Table 1.1: Undercount of persons by province

Province	Raw census count	Final estimate	Undercount rate (%)
Western Cape	3 612 835	3 956 875	8,69
Eastern Cape	5 636 408	6 302 525	10,57
Northern Cape	709 348	840 321	15,59
Free State	2 403 009	2 633 505	8,75
KwaZulu-Natal	7 338 554	8 417 021	12,81
North West	3 040 607	3 354 825	9,37
Gauteng	6 614 205	7 348 423	9,99
Mpumalanga	2 518 065	2 800 711	10,09
Northern Province	4 373 560	4 929 368	11,28
South Africa	36 246 591	40 583 574	10,69

The table shows that the rate of undercount of persons varied between provinces, being lowest in Western Cape and highest in Northern Cape. In part, undercount was related to the proportion of the population living in urban areas: some provinces with a higher proportion of people living in urban areas had lower undercount rates (Gauteng, Western Cape), and *vice versa* (Eastern Cape, Northern Province). More information on undercount is presented in later sections. Undercount rates by type of area are presented in Section 3.4. Undercount rates by demographic characteristics are included in Section 4.

Undercount rates were also calculated for *households*. The census data on households are also important, because the questions asked of households (such as access to electricity, water and telephones) are vital for planning. The undercount rate for households in South Africa was 6,6%. The method and details of calculating undercount rates for households are presented in Section 3.5.

2. The post-enumeration survey

2.1 Overview

The post-enumeration survey (PES) was conducted as soon as possible after the completion of the census enumeration, from 15 to 24 November 1996. A sample of around 800 enumerator areas (EAs) was selected, as described in Section 2.3. Within these selected areas, an attempt was made to visit every household, using better quality and senior census fieldwork staff under direct supervision of head office professionals. The interviewer posed a series of questions about the household to the householder regarding all persons present on the previous night as well as all persons present on the night of 9-10 October, census night. In addition to obtaining basic demographic information about all household members, the PES questionnaire included the question 'Was the person counted in the census?'. The design of the questionnaire is covered in Section 2.4.

The data on the questionnaires was captured and used in elementary fashion in the calculation of a *preliminary* undercount rate. This was based on the yes/no responses to the question on whether each person was counted in the census. The method used for establishing the preliminary population estimates is explained further in Section 3.2.

Subsequently, in a much more protracted undertaking, a match of the PES questionnaires with the corresponding census questionnaires was sought to determine whether the people included on the PES questionnaire were enumerated in the census at the same address, as described in Section 2.7. The final calculation of undercount incorporated the results of the matching exercise, plus some intricate imputations for different kinds of non-match, providing a more reliable *final* undercount. This is explained in detail in Section 3.3.

The PES was designed to provide an independent check of census coverage. As such, it was important that the survey should be conducted as independently of the census as possible. A different section of Statistics South Africa (Stats SA) was assigned responsibility for the PES to that with responsibility for the census. Although it was necessary for some aspects of the PES to utilise the infrastructure developed for the census, measures were taken to ensure that the processes were as independent as possible. Thus, while the enumeration area boundaries defined were those prepared for the census, the listing of dwellings within these areas was redone for the PES. In addition, while most of the interviewers for the PES had previously worked on the census, they were allocated to areas different from those where they had worked on in the census.

2.2 Scope

Census '96 was intended to cover every person present in South Africa on census night (except foreign diplomats and their families). However, for practical reasons, the coverage of the subsequent PES was limited to persons present in households in residential dwellings.

Difficulties in enumeration and matching meant that it was not possible to include prisons, hospitals and other institutions in the PES. An attempt was made to include a sample of hostels in the PES but matching proved to be impossible and the quality of the data obtained was not adequate to establish undercount. Thus the adjustments given to hostels had to be determined from the remainder of the sample for persons in households in residential dwellings. Homeless people were also beyond the scope of the PES although, as the process was finally applied, they would have received the same adjustment for undercount as people present in residential dwellings in urban formal areas.

2.3 Sample design

A sample of around 800 enumerator areas (EAs) was drawn for the PES. This represented around 1% of EAs for most provinces with the exception of Northern Cape. As this province has a very small population, the sample size was doubled to ensure a sufficient sample size for provincial estimates to be calculated.

The sample was stratified by province and EA type. EAs were classified as formal urban, informal urban, tribal, commercial farms or other non-urban (see Appendix E for precise definitions). EAs corresponding to an EA type of hospital and prison institutions were beyond the scope of the PES. EAs in the EA type of hostels were initially included in the scope of the PES and were sampled and enumerated. However, they were later excluded during the estimation process.

Within each province and EA type stratum, the EAs were sorted by magisterial district and an independent systematic sample was drawn. Empty EAs and nearly empty EAs (based on the estimates from census demarcation) were excluded from the sampling frame. This was done as empty EAs would not contribute to the sample estimate, and it was felt that the expense of enumerating nearly empty EAs could not be justified given their minimal contribution to the overall results.

No sampling of dwellings was undertaken within EAs. Instead, PES enumerators should have visited every dwelling in selected EAs. As a result, the sample was self-weighting within each stratum.

The sample was drawn from a list compiled from administrative records created prior to census enumeration. The number and boundaries of EAs changed to some extent during census enumeration, in that some areas were still being demarcated as enumeration began, while other EAs were split or combined during enumeration. The areas demarcated during the census were not included in the original list and were therefore not covered in the PES. The other changes made during the census had implications for matching, as will be detailed in Section 2.6.

It is possible that some people may have been missed in areas not demarcated by the census. As the list of EAs for the PES sample selection was based on census records, any areas not demarcated for the census would not be in scope for the PES and would not be adjusted for in

the undercount calculations. However, after the completion of Census '96 fieldwork, Statistics South Africa was involved in a project which involved capturing the 1996 census EAs on a geographical information system (GIS). This enabled a comprehensive review of the demarcation of EAs used for Census '96. While it is possible that some areas were completely missed by the census, the project showed that most areas that were not demarcated were either unpopulated or were enumerated as a part of another EA

2.4 Questionnaire design

The PES questionnaire was very brief. Every person who spent the previous night in the household should have been included on the PES questionnaire. In addition, persons who were not present on the previous night but who spent census night in the household should also have been included. The questions covered the following issues:

- Basic demographic information for each person present (age, sex, marital status, language, education).
- Whether the household was visited and whether each person was counted in the census.
- The opinions of the householder towards the census.
- Whether the questionnaire could be matched back to a census questionnaire and, if so, whether each person was found on the questionnaire (these questions were marked 'For office use only').

A reproduction of the questionnaire is included in Appendix A. The results of the opinion questions were summarised in the preliminary report. The other questions were used in the matching process and in the calculation of the undercount estimate.

There was no opportunity to test the questionnaire or the methodology. Subsequent discussions have indicated that more thorough information on people who had moved since census night might have been useful in the estimation process. There was no question specifically addressing this issue. Related information was obtained from the questions on where the person was on census night and whether this dwelling was their usual residence, but this was not used in the estimation process.

In order to identify possible *overcount*, the last questions on the PES questionnaire attempted to obtain a complete list of all people who should have been included on the census questionnaire. This was done by asking about people who were absent at the time of the PES but present on census night. People should have been enumerated on a census questionnaire at the address where they spent census night, 9-10 October 1996. Thus, for households which had not moved since the census, anyone present on the census questionnaire but not on the PES questionnaire should, in theory, not have been entered onto the census questionnaire at that household as they were not present on census night. These people may have been overcounted if they were also entered on the census questionnaire at the address where they actually spent census night. However, examination of the PES showed that it was not possible to draw this conclusion as there were various causes of discrepancies between the census and PES listings. For example, as the census was conducted over a month, many people were enumerated at a location other than their census night address.

While this information could not be used in the manner originally intended, it had some consequences for the final PES data although, as the number of people involved is small, these are minor. It was not possible to distinguish people absent at the time of the PES but present on census night from those present at the time of the PES, due to the manner in which the data was captured. This reflects the fact that the way in which the data would be used was not considered at the time data capture (see Section 2.6). The impact on final data was on the applicable population of the PES, which is now wider than the scope of the census, potentially distorting comparisons of counts of EAs and average household size between the PES and census. However, the slight increase in the PES sample should not have had an effect on the final undercount estimates (see Section 3.3.2 for the alternative method used to adjust for overcount for the final calculations).

2.5 Enumeration

The PES was designed to provide an independent check of the census count and, accordingly, the team responsible for the PES was not directly involved in the census. For practical reasons, PES enumeration procedures were largely the same as those used in the census. However, measures were taken to ensure that the PES was conducted as independently of the census as possible and was thorough in identifying persons missed by the census.

PES fieldworkers were recruited from the chief enumerators and controllers who had been employed on the census. Although this compromised the independence of the PES to some extent, this had a number of advantages for the PES. The recruitment procedures were simplified and the staff recruited were drawn from those known to have performed well in the census. Using experienced interviewers familiar with the processes helped to ensure that PES enumeration was as complete as possible. At the same time, independence between the census and PES was obtained by assigning fieldworkers to areas other than where they had worked in the census.

Around 1 850 temporary staff were involved in the PES enumeration including approximately 1 600 interviewers (two per EA), 200 fieldwork supervisors and 50 regional managers. All temporary staff received two days of training. As the staff were already familiar with the census questionnaire and the PES questionnaire was comparatively short and simple, the emphasis of the training was on the concepts and procedures specifically related to the PES.

Census enumeration had consisted of two main phases. The first was demarcation, when the country was divided into enumerator areas consisting of a sufficient number of dwellings to form a workload for a census enumerator. For each EA, an enumerator's summary book was prepared which usually contained a description of the boundaries of the area, a map or aerial photograph, and a listing of all visiting points within the area. The second phase was the actual census enumeration. During this phase, enumerators attempted to obtain completed questionnaires from all households in each visiting point listed in the summary book. They should also have added in any visiting points missed during the original demarcation.

The PES consisted of two similar phases. Firstly, the selected EAs were identified and copies of the maps and boundary descriptions were obtained from enumerators' summary books

prepared for the census. Although it was a feature of the design to use the same boundaries as prepared for the census, the visiting points were re-listed so any points not listed during the census should have been included in the PES. However, the PES could obviously not identify any households missed in the census when they were not included in the boundaries of an enumerator area originally demarcated for the census.

Secondly, each visiting point in the selected EAs was enumerated. Having two PES field workers per EA reduced the time taken to complete enumeration, which is important in the PES as it is essential to interview respondents as soon as possible after the census to ensure that they recall the details of census enumeration as accurately as possible. This required the two field workers to work closely together to ensure that every dwelling in the EA was enumerated and none were double-counted. PES enumeration took place from 15 to 24 November 1996.

Where it was impossible to obtain a completed questionnaire from a household, the enumerators were instructed to note this in their summary books. A completed questionnaire was obtained for almost 95% of households in the sample. The most common reason for non-completion of a questionnaire was non-contact (3,1%) with less than 1% of households refusing to participate (see Appendix C, Table C.1 for details by province).

2.6 First phase data capture

Once the PES questionnaires were returned to head office, the data from the household and personal questions were captured by an external organisation for use in the calculation of the preliminary estimates.

Due to time pressures, the PES questionnaires were not checked before data entry and no editing of the data was performed before the preliminary estimates were produced. While this subsequently required investigations and corrections, as will be explained in Section 2.8, it had little impact on the very simple method used for the preliminary calculation of undercount, which was based on the yes/no responses to the question on whether each person was enumerated in the census. The establishing of the preliminary estimates is discussed further in Section 3.2.

2.7 Matching process

After establishing the preliminary estimates, the next stage required PES questionnaires to be matched against the corresponding census questionnaires for that address to check the completeness of the census enumeration.

Matching proved to be a challenging exercise. There were around 80 000 households and 342 500 persons in the PES and a match status had to be determined for all of them. It took a team of usually 30, but up to 60, people around nine months to complete the process.

In the run-up to Census '96, Stats SA had planned that questionnaires would be anonymous, to try and minimize the extent to which enumerators would encounter refusals in politically tense parts of the country. Hence, the census questionnaire asked for only 'first name or initials to make it easy to complete the questionnaire'. The lack of detailed name information made matching to the same address difficult at times. Other obstacles faced included the lack of detailed addresses in many parts of the country and the logistical difficulties encountered in locating census questionnaires at the provincial processing centers.

As the methodology was developed during the process, matching was actually done twice. In the first phase, the 'For office use only' sections on the PES questionnaires, simply recording whether each household and person was found, were completed. However, further investigations showed that this information was insufficient for the calculation of undercount and matching was repeated with more detailed information being recorded on a 'matching sheet', a reproduction of which is included in Appendix B. This matching sheet had two main purposes:

- To record whether the household was not matched and, if not, whether it was missed or unresolved and the reason why.
- To record the identifying information for the census questionnaire against which the PES questionnaire was matched.

In both phases, matching involved several stages: matching EAs, matching households and matching individuals.

2.7.1 Matching EAs

The first stage of matching an EA involved locating the census questionnaires corresponding to the PES EA. This was often not as straightforward as may have been expected. The process should have just involved obtaining the box of questionnaires for the census EA which had the same number as the PES EA. However, the PES EAs were based on the areas as determined prior to the census and the changes were made during census enumeration meant that EA numbers did not always correspond. For example, during the census some EAs were renumbered, split, combined or had boundaries altered. Thus, the box of census questionnaires for an EA with a different number, or for a combination of EAs, may have been required for matching to a PES EA.

Attempts were made in head office to resolve the problems but sometimes it was necessary for staff to visit the provincial offices to look at maps or other boxes of census questionnaires to find the corresponding EA or EAs. In some cases it was even necessary to visit the actual EA to try to determine what had happened. Where it was impossible to locate a corresponding EA, the PES EA was usually excluded from the PES sample used for calculating the undercount rate. This affected 23 of the PES EAs, and will be discussed further in Section 3.3 on the final estimation.

The problems with correspondence of boundaries between census and PES EAs, while making matching more difficult, should not have had an impact on the calculated undercount. The methodology was based on the match status of each household and person. If a household and occupants were enumerated as part of another EA in the census, the thorough searching

procedures would have ensured that they were not erroneously treated as undercount. If there was any doubt about whether they were enumerated, codes of unresolved would have been assigned.

Sometimes, even when it seemed a corresponding EA had been found, other problems meant that a visit to the province was necessary. For example, there were cases when a different numbering system appeared to have been used, with stand numbers used in the census and street numbers used in the PES, and the relationship between the numbers was determined by visiting the EA.

2.7.2 Matching households

Once the corresponding census questionnaires were located, the next task was to match at household level. This was done by comparing the address listings in the enumerators' summary books for the census and PES to try to identify the corresponding households. Where this was inconclusive, the questionnaires were compared to see if a match could be found based on names and household structure.

Each household was classified as *matched*, *missed* or *unresolved* on the matching sheet. Where a household was classed as missed or unresolved, the reason for this was recorded if known. For example, the corresponding entry in the census enumerator's summary book may have recorded the visiting point as a refusal or non-contact. In this case, it is clear that the dwelling was missed and the cause was identified.

In areas with formal addresses, matching was a relatively straightforward process and it was not difficult to identify a corresponding household or to confirm whether a household was missed in the census. However, in areas without formal addresses, matching was more complex. In these areas, visiting points were usually listed in the summary book using the names of householders rather than a street address and more reliance was placed on the comparison of questionnaires. Difficulties arose when the householders' names did not match or were not unique or when the composition of households had changed. Sometimes it was impossible to confirm whether a particular household and its members were enumerated or not. In these cases, the household was classed as unresolved.

Sometimes a corresponding questionnaire was found but the household present at the time of the census was completely different. This is possible, as the original household may have left and a new household moved in between the census and the PES. A code reflecting this situation was allocated on the matching sheet and the household was treated as unresolved.

2.7.3 Matching persons

Once a corresponding household was identified, a match status was allocated for each person on the PES questionnaire. In most cases it was possible to identify a match on the basis of the name. However, this was more difficult in some cases where, for example, a different name or initials appeared to have been used. In these cases, a judgment on whether or not a person was

matched was made based on a number of variables including age, marital status, population group and gender. These variables did not have to be exactly the same for a match to be made as often, particularly for age, the responses differed slightly.

Usually when a corresponding census questionnaire was identified, all members of the PES household were classified as enumerated or missed. However, there were still some situations where it was necessary to allocate a code of unresolved for a person on the PES questionnaire. For example, where some characteristics of a person on the census questionnaire were similar but others were very different, and the structure of the households did not assist in indicating whether the persons matched, the unresolved category was used.

No action was taken if a person was on the census questionnaire but not on the PES questionnaire. As mentioned in Section 2.5, initially the PES intended to identify these people as potential overcount if the household was the same in the census and PES. However, it was not possible to draw this conclusion and the estimation of undercount was based solely on the people selected in the PES, and not on the people who were included in the census but not in the PES.

2.8 Second phase data capture

There were a number of steps involved in capturing the data. These reflected both the need for preliminary population estimates, as mentioned in Section 2.6, and the development of the methodology for the PES even after processing and estimation had commenced.

The first phase of data entry involved capturing the information recorded in the PES interviews which was used for the calculation of the preliminary estimates. After the first stage of matching was completed, the information from the 'For office use only' questions was captured. Next, the data from the matching sheet from the second stage of matching was captured. These three data sets were then merged together.

As mentioned earlier, the initial phase of data entry did not involve any checking of the questionnaires or editing of the captured results. Later stages of data entry did involve some checks and edits. However, as final estimation commenced, a number of errors were discovered on the file, relating particularly to the numbering of questionnaires which resulted in distorted household counts, and to the manner in which the data was captured for some questions. As a result, a final stage of data checks and corrections took place before an adequate dataset was available for the calculation of the final undercount rate and the establishing of final population estimates.

3. Calculation of undercount

3.1 Introduction

The same basic methodology was used to calculate both the preliminary and final undercount rates. The difference arose in the data used to calculate the components. The basic steps involved in calculating undercount are as follows.

Firstly, the data from the PES is used to calculate an adjustment factor:

$$\text{Adjustment factor} = \frac{\text{No. of people in PES in scope of census}}{\text{No. of people in PES counted in census}}$$

Secondly, this adjustment factor is then applied to the census counts:

$$\text{Population estimate} = \text{Adjustment factor} * \text{census count}$$

Lastly, the undercount rate can be calculated:

$$\text{Undercount rate} = \frac{\text{Population estimate - census count}}{\text{Population estimate}}$$

3.2 Preliminary calculations of undercount

The preliminary undercount rates were based solely on the answers to the question in the PES: 'Was the person counted in the census?'. People who responded 'Yes, here' or 'Yes, elsewhere' were treated as enumerated in the census, while people who responded 'No' were treated as missed. People who did not respond, or responded 'Don't know', were excluded from the estimation.

Calculation of the preliminary undercount rate was performed within the strata of province and EA type. These adjustment factors were applied to preliminary census counts determined from a 25% sample of census boxes. No adjustment for undercount was made for persons in hostels and institutions (see Appendix E for definitions of EA types including hostels and institutions). For more information on the preliminary calculations of undercount, see the publication on Census '96 preliminary estimates.[1]

This simple method was used as the problems with discrepancies in boundaries between census and PES EAs, as described in Section 2.7 on matching, meant that alternatives such as comparing counts in the census and PES were not feasible.

[1] Central Statistical Service (1997). *Census '96: Preliminary estimates of the size of the population of South Africa.* Pretoria: Central Statistical Service.

3.3 Final calculations of undercount

The final calculations of undercount were, in effect, a combination of responses to the PES question and the matching results. The matching results were, in themselves, insufficient for the calculations because:

- There was an unresolved component – EAs, dwellings or persons – for which it was not possible to determine whether they were enumerated in the census or not.
- Matching was only done to the dwelling corresponding to the PES enumeration and not to all dwellings where a person may have been located.

EAs corresponding to hostels were excluded from the matching process, since the hostel population tended to be very mobile and the quality of the resulting PES data was often poor with many missing responses. As mentioned earlier, a different method had to be used to calculate weights for persons in hostels and institutions. This is described below in section 3.3.3 on weighting.

Matching had also been impossible for a few PES EAs where, even after visiting provinces and using all other available information, a corresponding census EA could not be located. This was largely the result of administrative problems – most of the people in these EAs reported that they had been enumerated. These 23 EAs were therefore not used in the calculation of the final estimates. In effect, it was assumed that the magnitude and distribution in the remaining EAs was similar. After the exclusion of the 18 EAs for hostels and the 23 EAs for which matching was impossible, the sample used for final calculation of the undercount was 765 EAs.

Matching was limited to comparing the PES questionnaire with the census questionnaire at the same address. In some other countries, an attempt is made to confirm whether people who said they were counted somewhere other than the PES address were in fact enumerated. This was not attempted in South Africa as the circumstances, such as a lack of precise address information and practicalities accessing individual questionnaires elsewhere, made this too difficult. As a result, the questionnaire did not include questions on alternative addresses. The main impacts of this were the increased role of imputation and the incomplete identification of overcount, although attempts were made to address this, as explained below.

3.3.1 Imputation

After matching, all persons were classed as either *matched*, *missed* or *unresolved*. Those coded as matched or missed were considered resolved and they were assigned a probability of having being counted of 1 or 0, respectively, for use in subsequent calculations. For those coded as unresolved, in order to produce estimates of undercount, a probability of being enumerated was imputed. The probability was calculated from the resolved cases by making use of a technique known as CHAID (Chi-square Automatic Interaction Detection).

The CHAID model was run on the resolved cases, excluding people who reported being enumerated elsewhere (see section 3.3.2 below), with a probability of 1 allocated to a matched

case and a probability of 0 to a missed, as noted above. The following variables were used as predictors in the CHAID analysis: EA type, the response to the question whether the person was counted, age group, gender, population group and household size.

The CHAID technique determines the predictors in order of the strength of the predictive power of the variables and so identifies the statistically significant predictors and interactions between them. This analysis was done separately for each province. In every instance, the response to the question on whether a person was counted was indicated as the most significant predictor. Further predictors varied between the different provinces.Thus, for each province, the CHAID model created a number of 'branches' from combinations of categories of the predictive variables.

The proportion of persons enumerated among the resolved cases was calculated within each CHAID branch. This was interpreted as the estimated probability of being enumerated for cases with the characteristics defined by each branch. For example, in Mpumalanga, for persons aged between 19 and 48 who were in households of 4 to 8 persons and who said they were counted, the probability of enumeration was 0,9113. Accordingly, each unresolved record in each province was allocated the appropriate probability of enumeration.

3.3.2 Overcount

A number of people with responses of 'Yes, elsewhere' for the PES question on where counted in the census were, in fact, found on the census questionnaire at the PES address. This raises the possibility that these people were counted more than once (overcounted) if they were actually counted elsewhere, as reported. The fact that census enumeration took place over an extended period, during which time many people moved around, meant that there was some scope for people to be overcounted in Census '96.

It was possible to use the information gathered in the PES to make allowance some aspects of overcount. Accordingly, the imputation process was extended to impute the probability that people who responded 'Yes, elsewhere' were correct and were enumerated elsewhere, based on the accuracy of responses of 'Yes, here'. However, it was not possible, given the various constraints relating to logistics, time and funding, to conduct an exercise that would give a comprehensive estimate of overcount. While the allowance for overcount is not complete (for example, there is no allowance for people being counted more than twice), it still goes some way towards addressing the issue.

3.3.3 Weighting

Once a probability of being enumerated had been allocated to every PES record, another model was used to apply the results to the final census data. This modeling technique, XAID, was used to determine the appropriate weighting classes and the associated weights to be applied to each person record in the census. XAID is version of the methodology used for CHAID but using a continuous dependent variable rather than a dichotomous (that is, 0 or 1) dependent variable.

In the XAID analysis, the allocated probability of being enumerated was taken as dependent variable. The set of predictors was the same as in the CHAID analysis above, with the exception of the response to the question on whether a person was counted since this variable is not applicable to census records. The XAID analysis was run on all PES records in each province. The significant variables and their order of appearance in the XAID branches varied between the provinces. Age group and household size, however, figured prominently.

The XAID model determined combinations of the predictive variables that were significant in modeling the probability of being enumerated. The characteristics defined by the XAID branches were then taken as the weighting classes, and the average values of the probability of being enumerated were interpreted as the estimated counted rates. The reciprocal values of these counted rates were taken as the weights associated with all census records falling in the identified weighting classes. For example, in Mpumalanga, for African persons aged 19 to 48 in a households of 1 to 5 persons and in informal urban areas, the counted rate was 0,8339, yielding a reciprocal of 1,1992 which was the respective weight.

Once the weights had been calculated, they were applied to final census data and checked for anomalies. Where anomalies were identified, adjustments were made, although this occurred in very few cases and the adjustments were minor. For example, some situations were identified where the age distribution within a population group was unrealistically distorted by the weights which had adjusted adjoining age groups by very different amounts. Other anomalies were also found with respect to EA type. Accordingly, the XAID results were re-examined and the weighting classes used, or the weights within the classes were recalculated to smooth the distortions. In all, adjustments were made in five provinces, usually with negligible impact on the overall undercount rate. In some cases, the undercount decreased very slightly, in others it increased.

For the weighting of persons in hostels and institutions (as determined by EA type as detailed in Appendix E), a different approach was taken. It was not possible to use the results from the XAID as information on household size and EA type was not applicable. Instead, categories were developed manually based on examination of the data and using combinations of population group, province, urban/non-urban, age and sex. Within these categories, the appropriate adjustment factor was calculated based on all the records in the PES and applied to each person enumerated in hostels and institutions in the census.

The final weighting matrices used are presented in Appendix C, Table C.3 (persons not in hostels and institutions) and Table C.4 (persons in hostels and institutions).

3.4 Comparison of preliminary and final calculations

The undercount rate in the preliminary estimates was 6,8% while the final adjustment rate was 10,7%. This change is not surprising given the differences in the methodology. The preliminary estimates relied solely on the accuracy of the responses of the household informant as to who in the household had or had not been enumerated. There are a number of reasons why these responses may not be accurate:

- The person completing the PES may not be aware of whether the other people in the household were enumerated and may make assumptions about whether they were or not.
- People themselves may not be sure whether they were enumerated in the census. For example, some people may assume that they were included at their usual residence even though they were staying elsewhere during the census.
- Although the PES is conducted as soon as possible after census night, the householder may have forgotten the details of who was enumerated in the census (usually around a month would have elapsed between census and PES enumeration but, in extreme cases, this may have been up to two months).

The difference between relying on people's responses and actually looking at census questionnaire was the main reason for the difference between preliminary and final undercount rates. This can be seen in the detailed table comparing the responses and the matching results (before and after imputation), which is included in Appendix C, Table C.2.

However, there were some other factors that also affected the comparison. The various corrections to records carried out during the matching process and the finalisation of the dataset would also have had an impact. Thus, if estimates were calculated using the same methodology as for the preliminary estimates but based on the final dataset, they would obviously differ slightly from those published as preliminary estimates.

In addition, the method used for applying the calculated adjustment factors differed. The adjustment factors for the preliminary estimates were calculated in and applied to each EA type within each province. For the final estimates, adjustment factors took into account a wider range of factors, not only EA type but also age, gender, population group and household size. Adjustment factors by EA type for the preliminary estimates are thus not immediately comparable with those for the final estimates. Nevertheless, the table below compares preliminary and final estimates by EA type for each province.

Table 3.1: Preliminary* and final undercount by province and EA type

	Formal %		Informal %		Tribal %		Farms %		Total % inc. Other	
	Prelim	Final	Prelim	Final	Prelim	Final	Prelim	Final	Prelim	Final
Western Cape	8,3	8,8	4,5	9,4	0,0	..	1,8	6,9	7,3	8,7
Eastern Cape	9,0	10,1	12,2	16,4	1,8	10,2	2,9	8,6	4,8	10,6
Northern Cape	4,0	10,7	16,7	19,2	..	..	24,3	28,1	9,0	15,6
Free State	4,2	8,4	5,7	12,0	2,1	3,7	17,7	11,7	6,4	8,8
KwaZulu-Natal	8,2	11,9	14,4	16,7	6,8	12,4	18,4	15,6	8,5	12,8
North West	3,9	7,9	..	9,2	3,1	9,8	27,5	13,2	6,0	9,4
Gauteng	8,8	9,6	4,4	11,4	..	..	33,4	9,9	8,7	10,0
Mpumalanga	5,9	11,3	13,8	13,7	2,3	8,7	4,3	10,4	5,3	10,1
Northern Province	5,8	13,0	..	11,4	1,6	9,3	41,6	40,4	3,9	11,3
South Africa	7,6	9,9	8,0	13,1	3,3	10,2	17,2	15,0	6,8	10,7

* As published in Table 3, *Census '96: Preliminary estimates*, (CSS, 1997).

The undercount rate increased between the preliminary and final estimates in all provinces and most EA types. Some provinces and EA types were affected more than others. In particular, tribal areas, and Northern Province and Eastern Cape – which have a large proportion of the population in tribal areas – showed the greatest differences between the preliminary and final estimates. This may reflect persons believing that they were counted at a tribal level, even though a census enumerator had not visited them. Thus, the preliminary undercount rate in tribal areas was surprisingly low (3,3%) while the final estimate, incorporating the matching results, was similar to other EA types (10,2%, compared to 10,7% for South Africa).

The only areas where a decrease occurred were in farming areas in some provinces. Particularly in provinces with a small proportion of farmers, such as Gauteng, this may reflect the difference in the method of calculating and applying the adjustment factors for the final estimates. However, it is also possible, for example, that some people on farms were enumerated indirectly, i.e., without their knowledge, by farmers or other farmworkers, and this was picked up in the matching process.

3.5 Household calculations

Undercount of households was not addressed for the preliminary estimates when the focus was on producing an estimate of the number of persons in South Africa. However, it was important for the final estimates to ensure that the person and household estimates from the census corresponded. It was also important to provide accurate data for planning purposes, as mentioned in Section 1, since the census obtained information on households' access to services including electricity, water and telephones.

The method used to calculate the undercount of households for the final release of census data was similar to that used for persons. During matching, whether or not a household had been enumerated in the census was recorded on the matching sheet and this was the basis for establishing the undercount, along with whether the household reported being enumerated in the census. In the PES, each householder was asked whether the household was visited by the census. However, this question was not used in the final undercount calculation. Instead, a variable was derived from the responses of the persons in the household to the question 'Was this person counted in the census?'. This was used in preference to the question asked of households because the captured responses for households sometimes contradicted those for persons and the household responses appeared to be less accurate.

As with individuals, during matching households were coded as resolved (either matched or missed) or unresolved. Where the match status for a household was unresolved, a probability of enumeration was imputed using a CHAID model. The most important factor in the model was the derived variable concerning whether a household had been visited in the census. Other factors included in the model were household size, EA type and population group of the first person in the household.

Once the imputation was completed, weighting classes and weights were calculated for the census data using an XAID model based on a similar set of characteristics – household size, EA type and population group. Again, as with individuals, the weights varied within different

combinations of categories in each province. For example, in Mpumalanga, for a household of five or more persons with the first person African and in an informal urban area, the counted rate was 0,9199, yielding a reciprocal of 1,0871 which was the respective weight. The weighting matrices used are presented in Appendix C, Table C.5. The final undercount rates for households are shown in Table 3.2.

Table 3.2: Undercount of households* by province

Province	Raw census count	Final estimate	Undercount rate (%)
Western Cape	928 831	983 047	5,52
Eastern Cape	1 262 236	1 332 392	5,27
Northern Cape	167 531	186 984	10,40
Free State	586 176	625 013	6,21
KwaZulu-Natal	1 531 399	1 660 935	7,80
North West	687 181	720 644	4,64
Gauteng	1 830 562	1 964 710	6,83
Mpumalanga	567 652	604 015	6,02
Northern Province	903 827	982 459	8,00
South Africa	8 465 395	9 060 200	6,57

* excluding institutions and hostels

People could be missed in the census either as a result of being missed within an enumerated household, or because the entire household which they were in was missed. Calculations based on the final undercount rates for persons and households indicate that just under two-thirds of the people missed in the census were in missed households.

3.6 Limitations of PES calculations and resulting estimates

The 1996 PES was an advance on that conducted in the counted portion of South Africa in 1991, notably in undertaking matching and being conducted countrywide soon after enumeration. Even so, it was subject to a number of factors that affected the quality of the data. It is important to note these limitations in order to use the results effectively and to improve procedures for the next PES. However, given the limitations, analysis of the final calculations indicates that they appear to yield a fairly accurate representation of the undercount in the 1996 population census.

Given the limited planning time of a year before the full-scale enumeration, it was difficult to give full attention and planning to the PES. This meant that the methodologies and procedures had to be revised through the stages of the PES. Stats SA intends to conduct a thorough study to evaluate the 1996 PES and to improve methodology and implementation of the PES for the 2001 census exercise, in order to ensure that all the required information is collected efficiently and the accuracy of the undercount calculations is increased.

Some of the problems already mentioned involved questionnaire design, data entry and matching. It is this last problem that probably had the greatest impact on the final estimates of undercount. The matching process is inherently difficult, even in countries where most of the

country has a formal address system and the expected undercount is low. For instance, there is always a tendency for the PES and census alike to miss the same people who are difficult to contact or do not want to be identified, which may lead to the estimate of undercount being slightly understated.

In South Africa, additional problems were encountered in matching, as has been explained Section 2.7. A conservative approach was taken towards matching and, if there was any doubt about whether or not a household or person was enumerated, it was set to unresolved. This could happen if, for example, the addresses in a particular area were vague and there were a number of households in the PES and census which did not appear to match, but which could not conclusively be said to have been missed. Differences in names and household structure between the census and PES mean that there could be some matches of households and persons that are not at first apparent. It is impossible to resolve such situations without revisiting the EA so these persons and households were set to unresolved.

As a result of this approach, the proportion of persons and dwellings coded as 'Unresolved' tended to be fairly high, with 22% of all persons in the PES coded as unresolved. While it is usually easy to confirm a match – the address and most of the occupants are the same – it is often difficult to confirm a miss, particularly where addresses are vague as in the example above. Thus, a number of missed households and persons would have been coded as 'unresolved' and, as a result, this would have lowered the estimated rate of missed households and persons.

It is not possible to predict any overall bias of the undercount calculations as they were also subject to bias in other directions as well, resulting from the matching exercise and other factors. However, it is possible that the final indications of undercount are slight underestimates given the high unresolved rate.

The issues concerning matching for the 1996 PES will be addressed in the development of the post-enumeration survey for the next census. The methodology behind matching and estimation will be further developed which will lead to improvements in many areas. For example, clarifications in procedures for matching coders should enable them to resolve a greater proportion of cases. In addition, improved mapping and administrative procedures should reduce the difficulty of locating EAs where households may have been enumerated. In particular, the GIS should remove the problems encountered with boundaries differing between the census and PES, simplifying the matching process and increasing the accuracy of estimates.

Despite these problems, the ambitious and arduous matching exercise for the 1996 PES provided valuable information for the calculation of final undercount rates and the experience indicates that some form of matching can be performed successfully in South Africa.

4. Characteristics of undercount

As noted earlier, there are a range of reasons people may be missed in the census. A consequence of this is that some segments of the population are more likely to be missed by a census than others. This section presents undercount results by various demographic characteristics to give an indication of the factors affecting undercount.

The undercount rates in this section are calculated directly from the raw PES dataset. An alternative would have been to compare unweighted and weighted counts from the census but this would have revealed less detail about the separate characteristics as the XAID model selected and combined characteristics to create weighting classes. Thus, these rates do not directly reflect the adjustment factors as applied to the raw census counts. As the rates are included here to indicate broad trends rather than provide precise data, they are presented in graphs rather than tables.

4.1 Age and sex

Figure 1, below, shows the undercount rate from the PES by age and sex. As has been observed in other countries, undercount is closely related to age and sex. A survey of undercount rates in Australia, Britain, Canada and the USA[2] noted the following trends:

- Young children are less reliably captured than children in their early teens, for both sexes.
- Young adult men are the hardest group to enumerate.
- Among adults, older adults are more easily enumerated than younger adults.

The authors of the report also noted that in Britain and the USA, women above 75 years of age were more difficult to enumerate than middle-aged women.

Although Census '96 had many unique problems and a higher undercount rate than the highly developed countries covered in the study, very similar trends were observed, as is clear in the following graph.

[2] Simpson, S and Middleton, E, (1997). *Who is missed by a national census? A review of empirical results from Australia, Britain, Canada and the USA.* CCSR Working Paper No. 2. Manchester: The Cathie Marsh Centre for Census and Survey Research, University of Manchester.

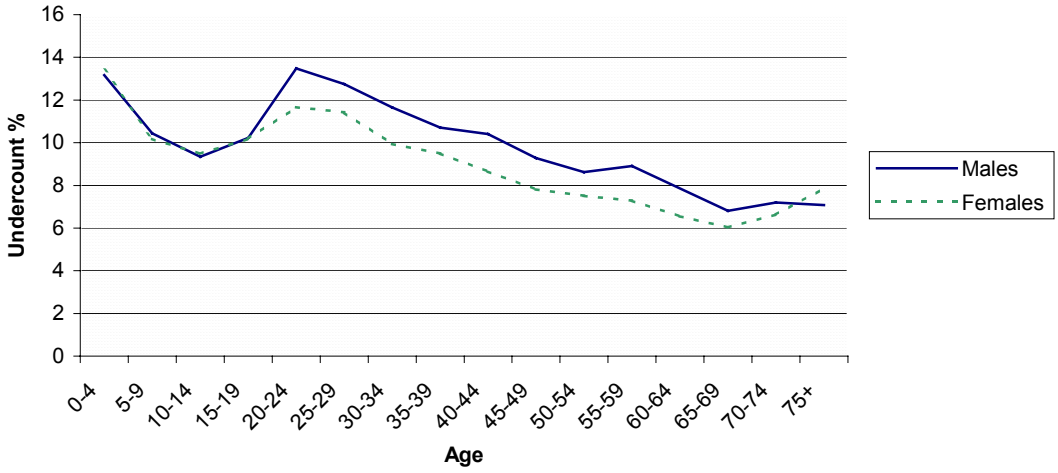

Figure 1: Undercount by age and sex, raw PES data

The groups most likely to be missed by the census are babies and young adult males. The most likely reasons for missing these groups are very different. Babies may be missed because the person completing the census questionnaire may not yet consider them as household members. Young adult males tend to lead more mobile lifestyles and be less attached to a particular household and thus are more difficult to enumerate. The decrease in the undercount rate for older age groups probably reflects the transition to more settled lifestyles. However, it is also interesting to note that, for most adult age groups, the undercount rate for males is higher than that for females.

4.2 Population group

Figure 2 presents the undercount rate by population group and sex.

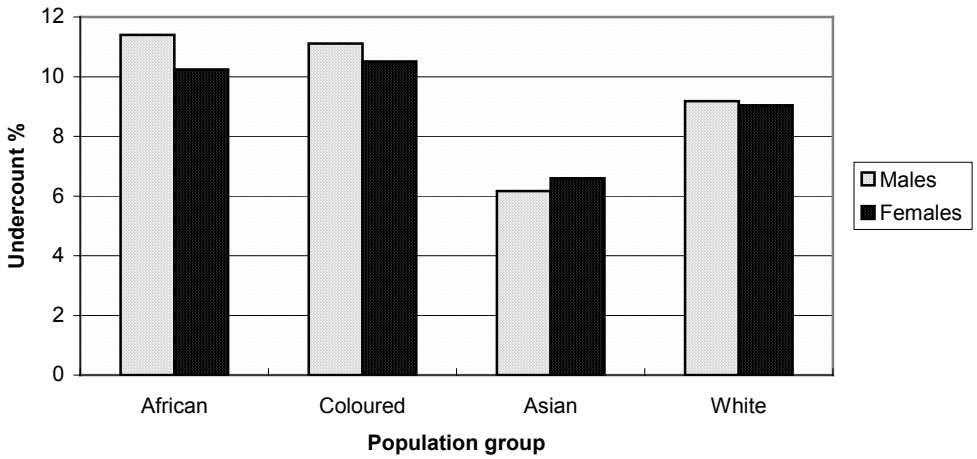

Figure 2: Undercount by population group and sex, raw PES data

Africans and coloureds have higher undercount rates than whites and Indians. This is also reflected in the undercount rates by EA type as presented in Table 3.1. A large proportion of white and Indian people live in formal urban areas which are easier to enumerate, while a larger proportion of black and coloured people live in informal and non-urban areas where enumeration is more difficult. Regarding the sex breakdown, only among Indians is the undercount among women higher than among men – which, when one uncovers the finding, makes sense for likely cultural reasons.

4.3 Marital status

The link between lifestyle and likelihood of enumeration in the census is also shown in the undercount rates by marital status (Figure 3).

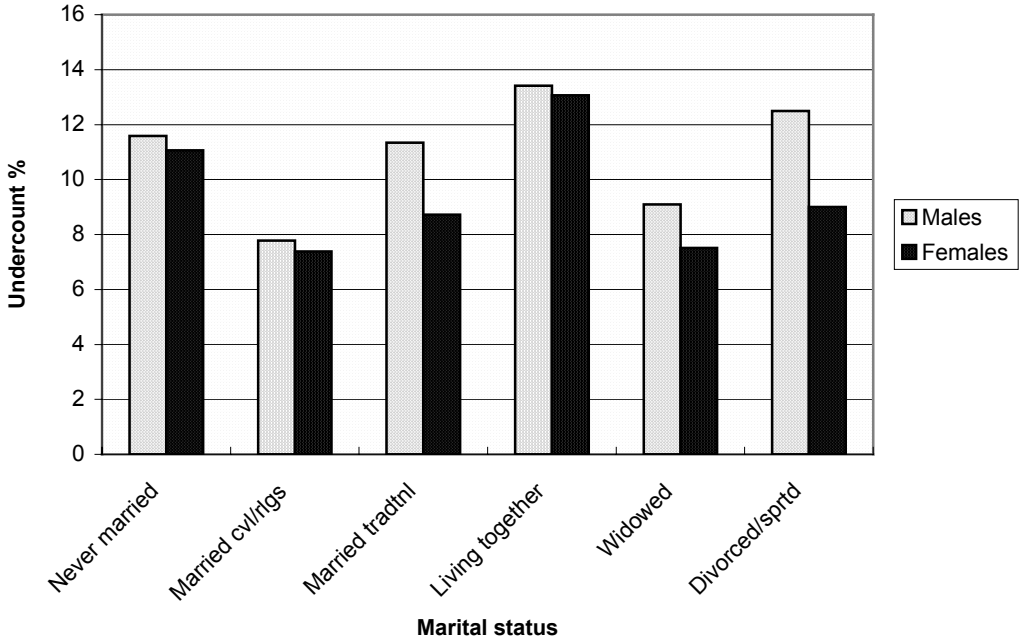

Figure 3: Undercount by marital status and sex, raw PES data

Undercount is lowest for people who are married (civil or religious) and highest for people who are never married, living together or divorced/separated. These results are, in part, correlated to age and reflect the high undercount rates for young adults and the lower undercount rates for older adults. People who are divorced/separated are likely to be less attached to a household and this is reflected in the higher undercount rates observed. There are also appreciable variations by marital status in the differences between the undercount rates by sex.

Appendix A: PES questionnaire

Did anyone from Census '96 visit this household some time during the period 10 to 31 October 1996 to count all the people living here and to ask questions about the household, or else to leave a questionnaire for you to complete?	Yes	1	No	2
FOR OFFICE USE ONLY: Was a census household questionnaire found?	Yes	1	No	2

ENUMERATOR: ENTER THE DETAILS OF EACH PERSON IN THE HOUSEHOLD IN THE COLUMNS BELOW, ONE PERSON PER COLUMN, BY GOING THROUGH ALL THE QUESTIONS FOR EACH PERSON

	Person 1	(Person 2 to Person 9 similarly)
Please indicate the first NAME or initials of each person who was present in the household last night.		
Is this person male or female? 1 = Male 2 = Female	1 2	
What is this person's date of birth? Or age in years OR nearest estimate?	D D M M Y Y ……..YEARS	
How would this person describe him/herself? 1 = African 2 = Coloured 3 = Indian 4 = White	1 2 3 4	
Which language does the person speak most often at home?		
What is the person's present marital status? 1 = Never married 2 = Married: Civil/religious 3 = Married: Traditional/customary 4 = Living together (with partner) 5 = Widow/widower 6 = Divorced/separated	1 2 3 4 5 6	
What was the highest school class/standard that the person has completed?		
Is this dwelling (house, room, shack) the place where the person usually lives, i.e. where the person spends at least four nights a week? 1 = Yes 2 = No	1 2	
Where was the person on Census night (the night of 9-10 October 1996)? 1 = In this household 2 = Elsewhere 3 = Do not know	1 2 3	
Was the person counted in the Census? 1 = Yes: in this household 2 = Yes: somewhere else 3 = No 4 = Do not know	1 2 3 4	
FOR OFFICE USE ONLY: Was the person included on the census questionnaire? 1 = Yes 2 = No	1 2	

Are there any other persons who **were here** (in this household) **on Census night** (the night of 9-10 October 1996), but who were not here last night?	Yes	1	No	2
If yes: How many such people are we talking about?	Number of people:			
Please give some details about this person (these persons) – Go to the next column in the above table. Enter the details of each such person.				

Appendix B: PES matching sheet

Matching sheet for PES questionnaires:

PES Questionnaire

Magisterial District Number: ☐☐☐

Enumerator area Number: ☐☐☐☐ Return Number: ☐☐☐

Household Number: ☐☐☐ Visiting point Number: ☐☐☐

House many questionnaires for this household?: ☐

Match

Corresponding visiting point found and completed questionnaire ☐ 1

No-Match

- Corresponding visiting point found but:

a) Refusal ☐ 2

b) Non-contact ☐ 3

c) Unoccupied dwelling ☐ 4

d) Vacant stand ☐ 5

e) No reason given ☐ 6

f) Other (explain below) ☐ 7

g) Household missed in census ☐ 8

- Visiting point missed in Census ☐ 9

Unresolved

Corresponding visiting point found but household different ☐ 10

Visiting point not found but might have been included in another EA (explain below) ☐ 11

Matching not possible (explain below) ☐ 12

Comments/Explanations: .....

.....

Census Questionnaire

Enumerator area No.: ☐☐☐☐ Visiting point No:: ☐☐☐

Household No: ☐☐ Return Number: ☐☐☐

If more than one questionnaire, what is the number of this questionnaire? .....

PES Person No.	1	2	3	4	5	6	7	8	9
Census Person No.									

Appendix C: Technical details

Table C.1: PES response rates for households

	Count				
	Refusal	Non-contact	Unspecified	Responding occupied households	Total occupied households
Western Cape	54	578	9	8 393	9 034
Eastern Cape	96	527	133	12 951	13 707
Northern Cape	3	16	94	2 984	3 097
Free State	36	134	20	4 782	4 972
KwaZulu-Natal	120	512	67	12 628	13 327
North West	10	49	102	5 302	5 463
Gauteng	318	597	251	17 717	18 883
Mpumalanga	15	105	170	5 702	5 992
Northern Province	4	66	96	8 058	8 224
South Africa	656	2 584	942	78 517	82 699

	Percent				
	Refusal	Non-contact	Unspecified	Responding	Total
Western Cape	0,60	6,40	0,10	92,90	100,00
Eastern Cape	0,70	3,84	0,97	94,48	100,00
Northern Cape	0,10	0,52	3,04	96,35	100,00
Free State	0,72	2,70	0,40	96,18	100,00
KwaZulu-Natal	0,90	3,84	0,50	94,76	100,00
North West	0,18	0,90	1,87	97,05	100,00
Gauteng	1,68	3,16	1,33	93,83	100,00
Mpumalanga	0,25	1,75	2,84	95,16	100,00
Northern Province	0,05	0,80	1,17	97,98	100,00
South Africa	0,79	3,12	1,14	94,94	100,00

Table C.2: Comparison of responses and matching results

a) PES results after matching

	Response to whether counted in the census				
Results of matching	Counted, here	Counted, elsewhere	Not counted	Don't know/ Not stated	Total
Enumerated	227 742	2 268	3 998	2 518	236 526
Missed	19 315	3 912	5 305	1 261	29 793
Unresolved	58 248	3 488	12 434	2 035	76 205
Total	305 305	9 668	21 737	5 814	342 524

b) Results of imputation

	Response to whether counted in the census				
Results of imputation	Counted, here	Counted, elsewhere*	Not counted	Don't know/ Not stated	Total
Enumerated	53 597	8 884	6 206	1 453	70 140
Missed	4 651	-5 396	6 228	582	6 065
% Enumerated	92,02	..	49,94	71,40	92,04
% Unresolved	19,08	36,08	57,20	35,00	22,25

* including allowance for double counting

c) PES results after matching and imputation

	Response to whether counted in the census				
Final results	Counted, here	Counted, elsewhere*	Not counted	Don't know/ Not stated	Total
Enumerated	281 339	11 152	10 204	3 971	306 666
Missed	23 966	-1 484	11 533	1 843	35 858
Total	305 305	9 668	21 737	5 814	342 524
% missed	7,85	-15,35	53,06	31,70	**10, 47 Final undercount rate**
			6,46 Preliminary undercount rate		

* including allowance for double counting

Table C.3: Weighting matrices for persons not in hostels and institutions

Western Cape

Age group	Household size	EA type	Undercount estimate	Adjustment factor (weight)
0 - 1	1 - 9	-	11,95	1,1357
0 - 1	10 +	-	31,58	1,4616
2 - 18	1 - 7	-	8,03	1,0873
2 - 8	8 - 13	Fml, Infml, Frm, Oth	15,38	1,1818
9 - 18	8 - 13	Fml, Infml, Frm, Oth	10,59	1,1184
2 - 18	14 +	-	22,60	1,2920
19 - 23	1 - 8	-	11,36	1,1282
19 - 23	9	-	19,87	1,2480
19 - 23	10 +	-	11,13	1,1252
24 - 33	1	-	13,34	1,1539
24 - 33	2 - 7	Other	18,75	1,2308
24 - 33	2 - 7	Fml, Infml	9,13	1,1005
24 - 33	2 - 7	Trbl, Frm	2,63	1,0270
24 - 33	8 - 9	-	14,93	1,1755
24 - 33	10 +	-	9,52	1,1052
34 - 48	1 - 2	Fml, Infml, Trbl, Oth	10,27	1,1145
34 - 48	3 - 8	Fml, Infml, Trbl, Oth	6,76	10,725
34 - 48	9 - 14	Fml, Infml, Trbl, Oth	10,56	1,1181
34 - 48	15 +	Fml, Infml, Trbl, Oth	12,22	1,1392
34 - 48	-	Frm	1,76	1,0179
49 +	-	-	4,76	1,0500
All other cases			8,76	1,0960

Note: the following abbreviations are used in these weighting matrices

EA type

Fml = Urban formal
Infml = Urban informal
Trbl = Tribal
Frm = Commercial farms
Oth = Other non-urban

Population group

Blck = African/Black
Clrd = Coloured
Indn = Indian/Asian
Whte = White

Eastern Cape

Household size	EA type	Age group	Population group	Undercount estimate	Adjustment factor (weight)
1	-	-	-	6,60	1,0707
2 - 5	Informal	-	-	17,13	1,2067
2 - 5	Fml, Trbl, Oth	-	Clrd, Indn	11,80	1,1338
2 - 5	Fml, Trbl, Oth	0 - 1	Black	13,41	1,1549
2 - 5	Fml, Trbl, Oth	2 - 18	Black	7,94	1,0862
2 - 5	Fml, Trbl, Oth	19 - 38	Black	9,27	1,1022
2 – 5	Fml, Trbl, Oth	39 +	Black	6,15	1,0655
2 – 5	Fml, Trbl, Oth	-	White	5,52	1,0584
2 – 5	Farms	-	Blck, Clrd	7,10	1,0764
2 - 5	Farms	-	Whte	2,23	1,0228
6 - 8	Informal	-	-	17,89	1,2179
6 - 8	Fml, Trbl, Frm	0 - 1	-	15,40	1,1820
6 - 8	Fml, Trbl, Frm	2 - 8	Coloured	14,08	1,1639
6 – 8	Fml, Trbl, Frm	2 - 8	Blck, Whte	10,02	1,1114
6 – 8	Fml, Trbl, Frm	9 - 13	-	7,55	1,0817
6 – 8	Fml, Trbl, Frm	14 - 38	Blck, Clrd	11,15	1,1255
6 – 8	Fml, Trbl, Frm	14 - 38	White	3,12	1,0322
6 – 8	Fml, Trbl, Frm	39 +	-	7,07	1,0761
6 - 8	Other	-	-	5,39	1,0570
9 - 10	-	0 - 3	-	19,70	1,2453
9 - 10	Informal	9 - 43	Coloured	33,99	1,5149
9 - 10	Informal	9 - 43	Black	9,79	1,1085
9 - 10	Tribal	9 - 23	-	10,52	1,1176
9 - 10	Tribal	24 - 43	-	16,08	1,1916
9 - 10	Fml, Frm, Oth	9 - 43	-	8,67	1,0949
9 - 10	-	44 +	-	7,53	1,0814
11 - 12	-	-	Coloured	25,31	1,3389
11 - 12	Trbl, Oth	-	Blck, Whte	16,14	1,1925
11 - 12	Fml, Infml, Frm	-	Blck, Whte	8,89	1,0976
13 - 14	-	-	-	18,01	1,2197
15 +	-	-	Clrd, Indn	44,10	1,7889
15 +	-	-	Blck, Whte	20,70	1,2610
All other cases				10,43	1,1164

Northern Cape

EA type	Household size	Population group	Age group	Undercount estimate	Adjustment factor (weight)
Farms	1 - 3	-	0 - 8	16,66	1,1999
Farms	1 - 3	-	9 +	17,01	1,2050
Farms	4 - 6	Blck, Clrd	0 - 28	35,92	1,5605
Farms	4 - 6	Blck, Clrd	29 +	23,31	1,3040
Farms	4 - 6	White	-	14,66	1,1718
Farms	7 - 8	Coloured	-	56,54	2,1524
Farms	7 - 8	Blck, Whte	-	13,70	1,1587
Farms	9 - 10	-	-	18,43	1,2259
Farms	11 +	-	-	35,52	1,5509
Informal	1 - 6	-	0 - 13	17,02	1,2051
Informal	7 +	-	0 - 13	39,66	1,6573
Informal	-	-	14 +	16,10	1,1919
Formal	-	-	0 - 1	20,43	1,2568
Formal	1 - 4	-	2 - 48	12,25	1,1396
Formal	5 - 9	-	2 - 48	9,48	1,1047
Formal	10 +	-	2 - 48	13,39	1,1546
Formal	-	-	49 - 68	6,88	1,0739
Formal	-	-	69+	6,24	1,0666
All other cases				14,95	1,1758

Free State

EA type	Population group	Household size	Age group	Undercount estimate	Adjustment factor (weight)
Infml, Frm	Blck, Clrd	1 - 7	-	11,01	1,1237
Infml, Frm	Blck, Clrd	8 +	-	1,05	1,2055
Infml, Frm	White	-	-	1,63	1,0166
Formal	Clrd, Whte	1 - 4	-	14,42	1,1685
Formal	Coloured	5 +	-	14,24	1,1660
Formal	White	5 +	-	5,90	1,0627
Formal	Blck, Indn	1 - 3	-	8,22	1,0896
Formal	Blck, Indn	4 - 11	0 - 3	8,94	1,0982
Formal	Blck, Indn	4 - 11	4 - 23	6,04	1,0643
Formal	Blck, Indn	4 - 11	24 - 53	7,61	1,0824
Formal	Blck, Indn	4 - 11	54 +	2,64	1,0271
Formal	Blck, Indn	12 - 14	-	11,31	1,1275
Formal	Blck, Indn	15 +	-	26,46	1,3598
Trbl, Oth	-	-	-	3,65	1,0379
All other cases				7,86	1,0853

KwaZulu-Natal

Population group	Household size	Age group	EA type	Gender	Undercount estimate	Adjustment factor (weight)
White	-	-	-	-	16,30	1,1947
Blck, Clrd	1 - 4	0 - 33	Informal	-	18,10	1,2210
Blck, Clrd	1 - 4	34 - 43	Informal	-	25,39	1,3403
Blck, Clrd	1 - 4	44 +	Informal	-	15,67	1,1858
Blck, Clrd	1 - 4	-	Fml, Frm	-	14,39	1,1681
Blck, Clrd	1 - 4	0 - 43	Trbl, Oth	-	11,80	1,1338
Blck, Clrd	1 - 4	44 +	Trbl, Oth	-	7,05	1,0758
Blck, Clrd	5 - 9	0 - 8	Infml, Frm	-	15,75	1,1869
Blck, Clrd	5 - 9	0 - 8	Fml, Trbl, Oth	-	10,86	1,1218
Blck, Clrd	5 - 9	9 - 13	Infml, Frm	-	14,14	1,1647
Blck, Clrd	5 - 9	9 - 13	Fml, Trbl, Oth	-	9,18	1,1011
Blck, Clrd	5 - 9	14 - 18	-	-	12,22	1,1392
Blck, Clrd	5 - 9	19 - 28	-	Male	16,86	1,2028
Blck, Clrd	5 - 9	19 - 28	-	Female	12,98	1,1492
Blck, Clrd	5 - 9	29 - 43	Infml, Frm	-	16,69	1,2003
Blck, Clrd	5 - 9	29 - 43	Fml, Trbl, Oth	-	12,11	1,1378
Blck, Clrd	5 - 9	44 +	Infml, Frm	-	12,15	1,1383
Blck, Clrd	5 - 9	44 +	Fml, Trbl, Oth	-	7,49	1,0810
Blck, Clrd	10 - 12	0 - 1	-	-	24,11	1,3177
Blck, Clrd	10 - 12	2 - 38	-	-	16,23	1,1937
Blck, Clrd	10 - 12	39 +	-	-	9,49	1,1049
Blck, Clrd	13 +	0 - 3	-	-	29,02	1,4088
Blck, Clrd	13 +	4 - 8	-	-	20,43	1,2568
Blck, Clrd	13 +	9 - 18	Formal	-	23,72	1,3110
Blck, Clrd	13 +	9 - 18	Infml, Trbl, Frm, Oth	-	13,92	1,1617
Blck, Clrd	13 +	19 - 48	-	-	20,59	1,2593
Blck, Clrd	13 +	49 +	-	-	12,06	1,1371
Indian	1 - 6	-	Trbl, Frm	-	20,97	1,2653
Indian	1 - 6	0 - 1	Fml, Infml	-	12,93	1,1485
Indian	1 - 6	2 - 8	Fml, Infml	-	6,85	1,0735
Indian	1 - 6	9 - 23	Fml, Infml	-	4,75	1,0499
Indian	1 - 6	24 - 28	Fml, Infml	-	7,31	1,0789
Indian	1 - 6	29 +	Fml, Infml	-	3,61	1,0375
Indian	7 - 9	0 - 3	-	-	13,58	1,1571
Indian	7 - 9	4 - 18	-	-	4,64	1,0487
Indian	7 - 9	19 - 48	-	-	9,07	1,0997
Indian	7 - 9	49 +	-	-	5,42	1,0573
Indian	10 +	-	-	-	20,88	1,2639
All other cases					12,22	1,1392

North West

Age group	Household size	EA type	Undercount estimate	Adjustment factor (weight)
0 - 1	1 - 9	-	15,12	1,1781
0 - 1	10 +	-	30,46	1,4380
2 - 8	-	Farms	21,97	1,2816
2 - 8	-	Infml, Trbl, Oth	10,48	1,1171
2 - 8	-	Formal	7,75	1,0840
9 - 13	-	Farms	10,43	1,1164
14 - 18	-	Farms	20,74	1,2617
9 - 18	-	Tribal	8,55	1,0935
9 - 18	-	Fml, Infml, Oth	5,96	1,0634
19 - 23	-	-	10,23	1,1140
24 - 28	-	-	12,47	1,1425
29 - 38	1 - 2	-	13,05	1,1501
29 - 38	3 - 7	-	6,99	1,0752
29 - 38	8 +	-	11,36	1,1282
39 - 58	-	Infml, Trbl, Frm	8,62	1,0943
39 - 58	-	Fml, Oth	5,60	1,0593
59 +	-	Trbl, Frm	6,58	1,0704
59 +	-	Fml, Infml, Oth	1,71	1,0174
All other cases			8,91	1,0978

Gauteng

Age group	Household size	Population group	EA type	Gender	Undercount estimate	Adjustment factor (weight)
0 - 1	1 - 4	Blck, Clrd	-	-	14,59	1,1708
0 - 1	5 - 11	Blck, Clrd	-	-	18,33	1,2244
0 - 1	12 +	Blck, Clrd	-	-	14,94	1,1756
0 - 1	-	Trbl, Frm	-	-	10,77	1,1207
2 - 18	1 - 3	Blck, Clrd	-	-	13,54	1,1566
2 - 18	4 - 5	Blck, Clrd	-	-	10,97	1,1232
2 - 18	6	Blck, Clrd	-	-	8,56	1,0936
2 - 18	7 - 10	Blck, Clrd	Informal	-	14,85	1,1744
2 - 18	7 - 10	Blck, Clrd	Fml, Frm, Oth	-	9,70	1,1074
2 - 18	11 +	Blck, Clrd	-	-	8,18	1,0891
2 - 18	-	Indn, Whte	-	-	7,83	1,0850
19 - 28	-	Blck, Clrd	-	Male	14,10	1,1641
19 - 28	-	Blck, Clrd	-	Female	11,74	1,1330
19 - 28	1 - 4	Indn, Whte	-	-	7,16	1,0771
19 - 28	5 +	Indn, Whte	-	-	10,04	1,1116
29 - 43	1 - 2	-	-	-	10,71	1,1199
29 - 43	3 - 5	Blck, Clrd	-	-	8,74	1,0958
29 - 43	3 - 5	Indn, Whte	-	-	6,15	1,0655
29 - 43	6 +	-	-	-	10,89	1,1222
44 - 58	1	Black	-	-	15,82	1,1879
44 - 58	1	Clrd, Indn, Whte	-	-	5,20	1,0549
44 - 58	2 +	-	-	-	7,45	1,0805
59 +	-	-	Infml, Frm	-	8,74	1,0958
59 +	1 -11	-	Formal	-	6,00	1,0638
59 +	12 +	-	Formal	-	5,05	1,0532
All other cases					9,90	1,1099

Mpumalanga

Population group	Age group	EA type	Household size	Gender	Undercount estimate	Adjustment factor (weight)
Clrd, Indn, Whte	0 - 18	-	-	-	17,69	1,2149
Clrd, Indn, Whte	19 - 43	Fml, Trbl, Oth	-	-	24,46	1,3238
Clrd, Indn, Whte	19 - 43	Infml, Frm	-	-	8,09	1,0880
Clrd, Indn, Whte	44 +	-	-	-	10,01	1,1112
Black	0 - 1	-	-	-	14,49	1,1695
Black	2 - 3	-	-	-	10,75	1,1204
Black	4 - 18	Infml, Frm	1 - 8	-	11,44	1,1292
Black	4 - 18	Infml, Frm	9 - 10	-	4,36	1,0456
Black	4 - 18	Infml, Frm	11 +	-	20,71	1,2612
Black	4 - 18	Fml, Trbl, Oth	-	-	7,26	1,0783
Black	19 - 48	Informal	1 - 5	-	16,61	1,1992
Black	19 - 48	Fml, Trbl, Frm, Oth	1 - 5	-	7,38	1,0797
Black	19 - 48	-	6 - 10	Male	13,87	1,1610
Black	19 - 48	-	6 - 10	Female	8,74	1,0958
Black	19 - 48	-	11 +	-	15,15	1,1786
Black	49 +	Informal	-	-	17,64	1,2142
Black	49 +	Fml, Trbl, Frm, Oth	-	-	6,38	1,0681
All other cases					9,85	1,1093

Northern Province

EA type	Household size	Age group	Population group	Gender	Undercount estimate	Adjustment factor (weight)
Farms	-	-	-	-	40,40	1,6779
Other	-	-	-	-	11,37	1,1283
Formal	1 - 3	-	Black	-	17,17	1,2073
Formal	1 - 3	-	Clrd, Whte	-	10,40	1,1161
Formal	4 - 11	-	-	-	10,74	1,1203
Formal	12 +	-	-	-	24,54	1,3252
Tribal	1 - 2	-	-	-	11,37	1,1283
Tribal	3 - 8	0 - 1	-	-	10,18	1,1133
Tribal	3 - 8	2 - 23	-	-	7,57	1,0819
Tribal	3 - 8	24 - 48	-	Male	12,49	1,1427
Tribal	3 - 8	24 - 48	-	Female	7,78	1,0844
Tribal	3 - 8	49 +	-	-	7,73	1,0838
Tribal	9 - 10	-	-	-	12,61	1,1443
Tribal	11 +	0 - 1	-	-	19,33	1,2396
Tribal	11 +	2 - 13	-	-	12,65	1,1448
Tribal	11 +	14 - 23	-	-	7,03	1,0756
Tribal	11 +	24 - 38	-	Male	20,06	1,2509
Tribal	11 +	24 - 38	-	Female	8,23	1,0897
Tribal	11 +	39 +	-	-	7,25	1,0782
All other cases					11,37	1,1283

Table C.4: Weighting matrices for persons in hostels and institutions

Note that, in addition to the weights set out here, there were also additional weights calculated to accommodate records where population group and/or age were missing. In these cases, the weight was based on the weight for all persons without regard for the missing characteristic but taking into account all other available information.

Population group	Province	Age group	Males		Females	
			Undercount estimate	Adjustment factor (weight)	Undercount estimate	Adjustment factor (weight)
Black	WC	0 - 1	19,35	1,2400	17,76	1,2160
Black	WC	2 - 3	12,22	1,1392	9,70	1,1075
Black	WC	4 - 8	10,21	1,1137	14,71	1,1725
Black	WC	9 - 13	10,50	1,1174	9,91	1,1100
Black	WC	14 - 18	14,90	1,1751	12,47	1,1425
Black	WC	19 - 23	11,84	1,1343	9,58	1,1060
Black	WC	24 - 28	10,97	1,1232	11,78	1,1336
Black	WC	29 - 33	9,66	1,1070	10,88	1,1221
Black	WC	34 - 43	7,43	1,0803	9,90	1,1099
Black	WC	44 - 53	7,36	1,0795	4,13	1,0431
Black	WC	54 +	6,07	1,0646	5,67	1,0601
Black	EC	0 - 1	15,95	1,1898	17,35	1,2099
Black	EC	2 - 3	13,31	1,1536	12,62	1,1445
Black	EC	4 - 8	10,57	1,1182	11,36	1,1281
Black	EC	9 - 13	7,97	1,0866	9,28	1,1022
Black	EC	14 - 18	9,73	1,1078	10,38	1,1159
Black	EC	19 - 23	11,93	1,1355	11,36	1,1282
Black	EC	24 - 28	11,58	1,1309	11,36	1,1281
Black	EC	29 - 33	11,23	1,1266	9,97	1,1107
Black	EC	34 - 43	10,33	1,1152	9,62	1,1065
Black	EC	44 - 53	8,41	1,0919	7,46	1,0807
Black	EC	54 +	7,77	1,0843	6,45	1,0689
Black	NC	0 - 1	29,74	1,4232	26,08	1,3528
Black	NC	2 - 3	20,62	1,2598	21,00	1,2659
Black	NC	4 - 8	15,79	1,1875	16,37	1,1957
Black	NC	9 - 13	16,30	1,1948	14,86	1,1746
Black	NC	14 - 18	13,66	1,1582	14,24	1,1661
Black	NC	19 - 23	16,10	1,1919	13,35	1,1541
Black	NC	24 - 28	16,19	1,1932	14,99	1,1764
Black	NC	29 - 33	16,52	1,1979	14,50	1,1696
Black	NC	34 - 43	9,72	1,1076	11,40	1,1287
Black	NC	44 - 53	11,77	1,1334	9,44	1,1042
Black	NC	54 +	12,21	1,1391	9,99	1,1110

Population group	Province	Age group	Males		Females	
			Undercount estimate	Adjustment factor (weight)	Undercount estimate	Adjustment factor (weight)
Black	FS	0 - 1	5,84	1,0620	12,33	1,1406
Black	FS	2 - 3	7,93	1,0861	9,99	1,1110
Black	FS	4 - 8	8,65	1,0947	6,53	1,0699
Black	FS	9 - 13	7,65	1,0828	5,97	1,0635
Black	FS	14 - 18	8,17	1,0889	4,98	1,0524
Black	FS	19 - 23	7,13	1,0768	8,91	1,0978
Black	FS	24 - 28	9,91	1,1100	8,01	1,0870
Black	FS	29 - 33	10,21	1,1137	7,87	1,0855
Black	FS	34 - 43	8,48	1,0926	6,75	1,0724
Black	FS	44 - 53	8,54	1,0933	7,24	1,0781
Black	FS	54 +	4,62	1,0485	3,71	1,0385
Black	KZN	0 - 1	17,56	1,2130	16,88	1,2030
Black	KZN	2 - 3	13,67	1,1583	15,09	1,1778
Black	KZN	4 - 8	14,06	1,1636	12,76	1,1463
Black	KZN	9 - 13	11,66	1,1320	11,47	1,1296
Black	KZN	14 - 18	12,84	1,1473	13,30	1,1534
Black	KZN	19 - 23	16,32	1,1950	14,34	1,1674
Black	KZN	24 - 28	18,58	1,2282	14,46	1,1690
Black	KZN	29 - 33	15,10	1,1778	13,52	1,1563
Black	KZN	34 - 43	15,30	1,1806	12,92	1,1484
Black	KZN	44 - 53	11,66	1,1320	8,63	1,0945
Black	KZN	54 +	10,29	1,1147	7,84	1,0851
Black	NW	0 - 1	19,76	1,2463	17,41	1,2108
Black	NW	2 - 3	13,10	1,1507	10,57	1,1182
Black	NW	4 - 8	10,62	1,1188	9,52	1,1052
Black	NW	9 - 13	7,53	1,0815	7,86	1,0853
Black	NW	14 - 18	7,74	1,0839	7,89	1,0857
Black	NW	19 - 23	11,43	1,1290	9,95	1,1105
Black	NW	24 - 28	12,06	1,1372	13,44	1,1552
Black	NW	29 - 33	11,22	1,1264	7,11	1,0765
Black	NW	34 - 43	8,71	1,0955	7,96	1,0865
Black	NW	44 - 53	8,53	1,0933	6,32	1,0675
Black	NW	54 +	5,51	1,0583	5,17	1,0545

Population group	Province	Age group	Males		Females	
			Undercount estimate	Adjustment factor (weight)	Undercount estimate	Adjustment factor (weight)
Black	G	0 - 1	15,68	1,1859	17,69	1,2149
Black	G	2 - 3	9,84	1,1092	11,40	1,1286
Black	G	4 - 8	10,46	1,1168	10,29	1,1147
Black	G	9 - 13	11,13	1,1252	9,65	1,1068
Black	G	14 - 18	11,75	1,1332	10,89	1,1223
Black	G	19 - 23	13,96	1,1623	11,87	1,1347
Black	G	24 - 28	14,24	1,1661	11,74	1,1330
Black	G	29 - 33	9,80	1,1086	10,46	1,1169
Black	G	34 - 43	9,63	1,1066	9,98	1,1109
Black	G	44 - 53	9,00	1,0989	8,46	1,0924
Black	G	54 +	8,84	1,0970	6,10	1,0650
Black	MP	0 - 1	14,88	1,1748	15,21	1,1794
Black	MP	2 - 3	9,94	1,1104	11,44	1,1292
Black	MP	4 - 8	8,03	1,0873	8,25	1,0899
Black	MP	9 - 13	7,62	1,0825	6,87	1,0738
Black	MP	14 - 18	8,17	1,0889	7,74	1,0839
Black	MP	19 - 23	10,28	1,1146	8,75	1,0960
Black	MP	24 - 28	11,61	1,1313	8,02	1,0872
Black	MP	29 - 33	11,18	1,1259	9,26	1,1020
Black	MP	34 - 43	11,73	1,1328	7,55	1,0816
Black	MP	44 - 53	9,43	1,1042	6,65	1,0712
Black	MP	54 +	8,11	1,0883	6,03	1,0642
Black	NP	0 - 1	13,46	1,1555	13,58	1,1571
Black	NP	2 - 3	10,76	1,1205	11,96	1,1358
Black	NP	4 - 8	10,74	1,1203	10,39	1,1159
Black	NP	9 - 13	9,74	1,1079	9,25	1,1019
Black	NP	14 - 18	8,91	1,0978	9,48	1,1047
Black	NP	19 - 23	14,24	1,1661	8,55	1,0936
Black	NP	24 - 28	17,60	1,2136	12,83	1,1472
Black	NP	29 - 33	18,44	1,2261	11,98	1,1361
Black	NP	34 - 43	19,86	1,2478	8,99	1,0988
Black	NP	44 - 53	15,63	1,1852	10,33	1,1152
Black	NP	54 +	11,40	1,1286	8,52	1,0932

Population group	Province	Age group	Males		Females	
			Undercount estimate	Adjustment factor (weight)	Undercount estimate	Adjustment factor (weight)
Coloured	Any	0 - 1	14,68	1,1720	16,57	1,1985
Coloured	Any	2 - 3	13,20	1,1521	11,28	1,1272
Coloured	Any	4 - 8	12,38	1,1413	10,21	1,1137
Coloured	Any	9 - 13	10,36	1,1155	11,88	1,1348
Coloured	Any	14 - 18	10,82	1,1214	11,06	1,1243
Coloured	Any	19 - 23	14,08	1,1638	13,22	1,1523
Coloured	Any	24 - 28	12,09	1,1375	11,04	1,1241
Coloured	Any	29 - 33	10,97	1,1232	10,59	1,1185
Coloured	Any	34 - 43	9,68	1,1072	8,75	1,0959
Coloured	Any	44 - 53	9,36	1,1033	7,97	1,0866
Coloured	Any	54 +	7,03	1,0756	6,80	1,0730
Indian	Any	0 - 1	14,63	1,1713	15,32	1,1809
Indian	Any	2 - 3	8,08	1,0879	13,21	1,1522
Indian	Any	4 - 8	6,42	1,0686	7,25	1,0782
Indian	Any	9 - 13	3,96	1,0413	5,55	1,0587
Indian	Any	14 - 18	5,56	1,0589	4,99	1,0525
Indian	Any	19 - 23	6,97	1,0749	8,00	1,0870
Indian	Any	24 - 28	9,39	1,1036	7,75	1,0840
Indian	Any	29 - 33	7,14	1,0769	8,21	1,0894
Indian	Any	34 - 43	6,00	1,0639	5,13	1,0541
Indian	Any	44 - 53	4,32	1,0452	4,17	1,0435
Indian	Any	54 +	3,25	1,0336	5,12	1,0540
White	Any	0 - 1	9,22	1,1016	13,42	1,1550
White	Any	2 - 3	9,80	1,1087	10,40	1,1161
White	Any	4 - 8	8,78	1,0963	9,68	1,1072
White	Any	9 - 13	8,58	1,0939	9,30	1,1026
White	Any	14 - 18	8,64	1,0946	8,35	1,0911
White	Any	19 - 23	12,29	1,1402	12,69	1,1453
White	Any	24 - 28	11,63	1,1316	10,14	1,1128
White	Any	29 - 33	10,88	1,1221	8,74	1,0958
White	Any	34 - 43	8,87	1,0973	8,11	1,0882
White	Any	44 - 53	7,84	1,0850	7,38	1,0797
White	Any	54 +	7,20	1,0776	8,23	1,0896

Table C.5: Weighting matrices for households

Western Cape

Household size	EA type	Undercount estimate	Adjustment factor (weight)
1	-	12,82	1,1471
2	-	7,09	1,0763
3 - 4	Fml, Infml	4,76	1,0500
3 - 4	Trbl, Frm, Oth	0,50	1,0050
5 +	-	2,28	1,0233
All other cases		4,84	1,0509

Eastern Cape

Household size	EA type	Population group	Undercount estimate	Adjustment factor (weight)
1	Trbl, Frm, Oth	-	17,81	1,2167
1	Fml, Infml	-	9,04	1,0994
2	-	-	8,47	1,0925
3 - 5	-	Clrd, Indn	5,49	1,0581
3 - 5	-	Blck, Whte	3,14	1,0324
6 +	-	Clrd, Whte	4,44	1,0465
6 +	Fml, Infml, Frm	Black	2,62	1,0269
6 +	Trbl, Oth	Black	1,04	1,0105
All other cases			4,93	1,0519

Northern Cape

EA type	Population group	Household size	Undercount estimate	Adjustment factor (weight)
Infml, Frm	Clrd	-	21,77	1,2783
Infml, Frm	Blck, Whte	-	12,48	1,1426
Formal	-	1	17,09	1,2061
Formal	-	2 - 5	6,49	1,0694
Formal	-	6 +	2,95	1,0304
All other cases			10,16	1,1131

Free State

Household size	Population group	EA type	Undercount estimate	Adjustment factor (weight)
1 - 2	-	-	10,54	1,1178
3 - 4	White	-	12,83	1,1472
3 - 4	Blck, Clrd	Infml, Frm	7,02	1,0755
3 - 4	Blck, Clrd	Fml, Trbl, Oth	2,46	1,0252
5 +	-	Farms	6,26	1,0668
5 +	-	Fml, Infml, Trbl, Oth	1,63	1,0166
All other cases			5,43	1,0574

KwaZulu-Natal

Household size	EA type	Population group	Undercount estimate	Adjustment factor (weight)
1	Infml, Frm, Oth	-	25,75	1,3468
1	Formal	-	19,98	1,2497
1	Tribal	-	10,16	1,1131
2	-	Clrd, Whte	17,78	1,2162
2	Informal	Black	18,86	1,2324
2	Fml, Frm	Black	12,89	1,1480
2	Trbl, Oth	Black	5,68	1,0602
2	-	Indian	3,81	1,0396
3 - 4	-	White	15,42	1,1823
3 - 4	Infml, Frm	Blck, Clrd	7,59	1,0821
3 - 4	Formal	Blck, Clrd	5,23	1,0552
3 - 4	Trbl, Oth	Blck, Clrd	3,13	1,0323
3 - 4	-	Indian	2,61	1,0268
5 +	-	White	13,45	1,1554
5 +	Infml, Frm	Blck, Clrd, Indn	5,89	1,0626
5 +	Formal	Blck, Clrd, Indn	2,68	1,0275
5 +	Trbl, Oth	Blck, Clrd, Indn	1,70	1,0173
All other cases			6,90	1,0741

North West

Household size	EA type	Undercount estimate	Adjustment factor (weight)
1	-	10,23	1,1140
2	-	7,40	1,0799
3 - 4	-	4,00	1,0417
5 +	Farms	5,42	1,0573
5 +	Tribal	1,92	1,0196
5 +	Fml, Infml, Oth	0,62	1,0062
All other cases		4,15	1,0433

Gauteng

Household size	EA type	Population group	Undercount estimate	Adjustment factor (weight)
1	Formal	-	15,07	1,1774
1	Infml, Frm, Oth	-	7,92	1,0860
2	-	-	7,93	1,0861
3	-	-	5,88	1,0625
4 - 5	-	Clrd, Indn, Whte	5,33	1,0563
4 - 5	-	Black	3,12	1,0322
6 - 7	-	Clrd, Indn, Whte	4,44	1,0465
6 - 7	-	Black	1,90	1,0194
8 +	-	Clrd, Indn	3,87	1,0403
8 +	-	White	4,84	1,0509
8 +	-	Black	0,50	1,0050
All other cases			6,59	1,0705

Mpumalanga

Population group	Household size	EA type	Undercount estimate	Adjustment factor (weight)
Clrd, Whte	-	-	17,60	1,2136
Black	1	-	10,29	1,1147
Black	2 - 4	Informal	12,12	1,1379
Black	2 - 4	Fml, Trbl, Frm, Oth	3,75	1,0390
Black	5 +	Informal	8,01	1,0871
Black	5 +	Fml, Frm, Oth	2,62	1,0269
Black	5 +	Tribal	0,69	1,0069
All other cases			6,05	1,0644

Northern Province

EA type	Household size	Population group	Undercount estimate	Adjustment factor (weight)
Farms	-	-	40,39	1,6777
Formal	1	-	23,42	1,3058
Formal	2 - 3	Black	14,92	1,1754
Formal	2 - 3	Clrd, Whte	7,82	1,0848
Formal	4 +	-	4,98	1,0524
Tribal	1 - 2	-	9,96	1,1106
Tribal	3	-	5,27	1,0556
Tribal	4 +	-	2,91	1,0300
All other cases			8,47	1,0925

Appendix D: Sampling error

The undercount rates presented in this report are subject to a number of sources of error. While any survey is subject to errors at various phases, there were some additional problems relating specifically to the 1996 PES in South Africa, as have already been discussed. Some of the possible sources of error in this PES include:

- Respondent errors in answering questions.
- Enumerator errors in identifying households that should be included.
- Errors made in matching PES and census questionnaires.
- Errors made in entering the data.
- Errors arising from the imputation of missing data.
- Other errors introduced through other assumptions made in the calculations.
- Sampling error, that is, the error arising from collecting data for the PES from a sample of the population rather than the whole population.

Estimates of sampling error have not been included in the body of this publication as sampling error constitutes only one source of error and providing sampling errors may have given a misleading impression of the accuracy of the data. However, the sampling errors are included here for information. These estimates were calculated based on the methodology used in selecting the PES sample.

Table D.1: Sampling errors for undercount by province

			95% confidence limits	
Province	Undercount rate (%)	Sampling error	Lower limit	Upper limit
Western Cape	8,69	0,73	7,23	10,15
Eastern Cape	10,57	0,64	9,29	11,85
Northern Cape	15,59	2,52	10,55	20,63
Free State	8,75	1,00	6,75	10,75
KwaZulu-Natal	12,81	0,68	11,45	14,17
North West	9,37	0,59	8,20	10,54
Gauteng	9,99	0,49	9,01	10,97
Mpumalanga	10,09	1,04	8,01	12,17
Northern Province	11,28	0,83	9,62	12,94
South Africa	10,69	0,27	10,15	11,23

The confidence intervals on the undercount rate mean that there is a 95% chance that the population of South Africa (estimated to be 40,58 million) is between 40,34 million and 40,83 million. However, as mentioned above, this does not take into account the sources of error other than sampling error.

Appendix E: Enumerator area type definitions

There were 15 different EA types, in three broad categories: urban, semi-urban and rural. The definitions below are based on the Census '96 manual for census officers:

Urban: situated within the boundaries of municipalities/local authorities

Formal	Ordinary town or city area as well as vacant areas within municipal boundaries. Various structures, eg., houses, flats, hotels, boarding houses, old age homes, caravan parks and school and university hostels may be found.
Informal	Area with informal dwellings (the so-called squatter areas).
Hostels	Area with mainly hostels where housing for employees is provided by employers (such as mines, factories and power stations).
Institutions	Area with mainly hospital and prison institutions.

Semi-urban: population concentrations adjacent to a municipality (must have one common boundary)

Semi-town	Semi-towns (ie. a town without a local authority) with predominantly formal dwellings.
Informal	Area with informal dwellings (the so-called squatter areas).
Hostels	Area with mainly hostels.
Institutions	Area with mainly hospital and prison institutions.

Rural: situated in rural areas

Semi-town	Semi-town (ie. a town without a local authority) with predominantly formal dwellings.
Town	Town without a local authority and which is not situated within a tribal area and with formal and semi-formal dwellings such as houses, huts and rondavels.
Tribal	Tribal area.
Informal	Area with informal dwellings (the so-called squatter areas).
Hostels	Area with mainly hostels.
Institutions	Area with mainly hospital and prison institutions.
Commercial farms	Area with farms, agricultural holdings, holiday resorts, agricultural schools and colleges and other rural areas.

These categories are referred to in the PES (and this report) in the following combinations:

Urban formal	Urban formal
Urban informal	Urban informal
Tribal	Rural tribal
Commercial farms	Rural commercial farms
Other non-urban	Semi-urban (all categories), Rural (all except Tribal and Commercial farms)
Hostels	Urban, Semi-urban and Rural Hostels
Institutions	Urban, Semi-urban and Rural Institutions

How to get more information about Census '96 products

Telephone: By contacting the Stats SA users enquiries department:
Tel: **(012) 310-8600**
Fax: **(012) 310-8500**
or by contacting your nearest provincial office (details below).

Internet and e-mail: Information and most free products will be available on the Stats SA home page situated at **http:\\www.statssa.gov.za**; or you can contact the Census Marketing and Dissemination Unit at **Patrickke@statssa.pwv.gov.za**.

Province	Office	Area code	Telephone	Fax
Eastern Cape	Bisho	(040)	635-0433	635-0449
Free State	Bloemfontein	(051)	447-7766	447-8402
Gauteng	Johannesburg	(011)	331-0122	331-0260
KwaZulu-Natal	Durban	(031)	305-3904	304-5508
Mpumalanga	Nelspruit	(013)	752-3561	755-2898
Northern Cape	Kimberley	(0531)	33-965	82-5407
Northern Province	Pietersburg	(015)	295-3300	295-3579
North West	Mmabatho	(0140)	84-2877	84-2832
Western Cape	Cape Town	(021)	423-1040	22-1741