DATA QUALITY

Report 03-01-04

COMMUNITY SURVEY 2016

www.statssa.gov.za Stats SA @statssa

THE SOUTH AFRICA I KNOW, THE HOME I UNDERSTAND

STATS SA
STATISTICS SOUTH AFRICA

COMMUNITY SURVEY 2016 DATA QUALITY REPORT

Report number: 03-01-04

ISBN: 978-0-621-44998-3

Community Survey 2016: Data Quality Report 03-01-04/ Statistics South Africa

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

Stats SA Library Cataloguing-in-Publication (CIP) Data
Community Survey 2016: Data Quality Report 03-01-04/ Statistics South Africa. Pretoria: Statistics South Africa, 2016

Report no. 03-01-04
59 pp

ISBN 978-0-621-44998-3

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, Mbombela
Central Reference Collection, Kimberley
Central Reference Library, Mmabatho

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

For technical enquiries please contact:
Christine Khoza
Contact: (012) 310 2916
Email: christinek@statssa.gov.za

Angela Ngyende
Email: angelan@statssa.gov.za
Contact: (012) 310 4699

Table of contents

LIST OF TABLES

LIST OF FIGURES

List of abbreviations and acronyms

CAPI	Computer Assisted Personal Interviewing
CS	Community Survey
DU	Dwelling Unit
DURF	Dwelling Unit Record Form
EA	Enumeration Area
FET	Further Education and Training
HH	Household
PAPI	Paper and Pencil Interviewing
SCM&E	Survey Coordination, Monitoring and Evaluation
STATSSA	Statistics South Africa
TVET	Technical and Vocational Education and Training
UN	United Nations
WB	World Bank

PREFACE

Community Survey 2016 (CS) is the second largest survey undertaken by Statistics South Africa following the one conducted in 2007. The survey remains one of the main data sources that provide indicators at national, provincial and municipal levels for planning and monitoring the performance of specific development programmes such as education, health, sanitation, water supply, housing and transport. In addition, the survey provides demographic information critical in understanding population-development nexus. The objective of the community survey was thus to provide population estimates as well as household characteristics. The information will be used to inform Integrated Development Plans and infrastructure investment budgeting.

The purpose of this report is to increase awareness on data/information gaps inherent in Community Survey 2016 data as a result of processes that may have affected output data. Direct and indirect data assessments were done, where imputation rates are used to highlight editing effects. Comparability of selected indicators between Community Survey 2007 and 2016 provides the extent to which both surveys can be useful for informed decision making.

P J Lehohla
Statistician-General

SECTION 1: INTRODUCTION

1.1 Introduction

Community Survey 2016 is the second intercensal survey in a democratic South Africa. This household based survey is one of the few available data sources providing data at municipal level. Provision of data at this level supports evidence-based decision making that has become increasingly a best practice which many countries embrace, including South Africa. CS 2016 results are thus critical in promoting optimal resource allocation and utilisation in all spheres of government in order to reduce poverty and vulnerability among South Africa's most marginalised. Furthermore, the development and implementation of policy, implementation of legislature deems it necessary to have reliable statistics that inform social, demographic and economic standing of the country.

The CS 2016 is the second large sample survey Statistics South Africa has undertaken after CS 2007, but this time around data was collected electronically using Computer Assisted Personal Interviewing (CAPI) system as opposed to the paper collection method used in CS 2007. The new initiative in the organisation is expected to reduce financial and time costs in data processing as well as data quality enhancement. Eligible persons for enumeration were all persons present in the household(s) of sampled dwelling units on the reference night (midnight 6th March 2016 to 7th March 2016) including visitors. Members of the household who were absent overnight, for example, working, travelling, at entertainment or religious gatherings but returned the next day should also be counted. For purposes of Stats SA a household is a group of persons who live together, and provide for themselves jointly with food and other essentials for living, or a person who lives alone. Babies born before the reference night were also included in the count, the reason being that they were already born by the midnight 6th March to 7th March 2016. Members of the household who died after the reference night were counted in as they were alive during the midnight of the reference period. In contrast, those born after the reference night were excluded.

The use of the CAPI system was not the only new process used, others include the utilisation of the present updated dwelling frame data captured from the Census 2011 listing process. Newly incorporated data items in the CS questionnaire include:

- main religious affiliation
- main reason for leaving previous municipality
- mode of transport used to reach the person's destination for going to school or work and time taken to reach the destination
- main challenge/problem/difficulty facing a municipality
- quality and satisfaction with basic services

- opinion on improving standards of living in households
- year of death of mother and father if reported deceased
- perceptions of safety and crime experienced in households
- food security measures
- extended additional questions on agricultural activities in households
- characteristics of emigrants (persons who left the country to live elsewhere)

This section highlights the planning processes for CS 2016. During the planning phase, the focus in the early stages was primarily on setting strategic directions and ensuring that all dependencies between the different phases and role players were identified, potential risks highlighted and control measures put in place to minimise their adverse effects. This facilitated effective integration and implementation of various activities by ensuring that each phase was properly resourced. During the planning phase, all work streams and focus areas prepared operational plans which provided detailed lists of activities that were to be undertaken to achieve specific objectives and outputs as profiled in the CS 2016 Project Charter.

1.1.1 Geography frame

The Dwelling Frame (DF) is a structures frame, and dwelling units (DUs) form part of the feature classification of structures. Datasets in the integrated DF base layer include: Dwelling Frame 2011 (formal and informal), Listing Census 2011, Spot Building Count 2012, Address Assignments and Municipal data. Worth mentioning is the fact that the use of existing updated dwelling frame for a large sample survey such as CS 2016 is the first of its kind in line with other first-time data collection processes in the case of Stats SA. It is also accepted that ongoing improvement is expected as it is not that perfect at this stage.

1.1.2 Community Survey 2016 sampling methodology

The sample design for CS 2016 was a stratified single-stage sample design. At EA level, all in-scope EAs were included in the sample and a sample of dwelling units was taken within each EA (i.e. there was no sub-sampling of EAs). The EA frame was based on the Census 2011 information. The updated dwelling unit (DU) frame was constructed by the Geography team using geo-referenced spatial systems.

1.1.3 Questionnaire development processes for CS 2016

The development and design processes of the CS 2016 questionnaire was informed by national priorities, global and continental emerging population issues embedded in the SDGs, data needs of both existing and prospective users and comparability with the previous community survey (i.e. CS 2007) and previous censuses. The development of the CS Questionnaire involved a number of phases as mentioned below:

- Stakeholder needs assessment is an international best practice in survey and census planning aimed at producing products that meet user needs. Stakeholders play a fundamental role in providing information on questions to be asked in a survey. During this phase, processes, including review of previous censuses' data items and questionnaires were undertaken, and user consultations were held with key internal stakeholders on what needed to be measured in CS 2016.
- Through the consultation process, it became clear that there is increased demand for data at municipal level. Following the consultation process, Community Survey 2016 data items were finalised and categorised into broader themes of demographics, migration, general health and functioning, parental survival, education, income and social grants, employment, fertility, mortality, housing conditions, access and quality of basic service provision, agricultural activities and emigration.

The Community survey 2016 questionnaire was designed using the World Bank Survey Solutions designer, which is an on-line based questionnaire design application. During the design, skipping patterns and validation rules were predetermined and embedded in the electronic questionnaire. Data collection instruments, questionnaires in particular were developed and subjected to thorough testing and review processes to ensure that the final product (questionnaire) solicits accurate information. The Community Survey 2016 questionnaire consists of new questions while some other questions have been adopted from existing household based surveys and Census 2011. Two-stage testing was adopted for CS 2016; viz. Behind-the-glass test, and integrated field testing. The results of each test were used to improve the quality of the draft questions and CS 2016 indicators. The draft CS questionnaire was presented at different fora for approval. These included CS Technical Committee, Questionnaire Clearance Committee, CS Management, Population & Social Statistics Cluster, Statistics Council Population Sub-committee and the Statistics Council.

1.1.4 The Survey Solution CAPI System

Survey Solutions is just one of several CAPI platforms available in the market currently (including CSPro, ODK, Blaise, SurveyToGo, etc.). It was first developed in 2012 after a comprehensive assessment of CAPI software products found no existing software provided exactly the right mix of features necessary for the sort of surveys conducted by the World Bank and its clients. It mainly assists governments, national statistical offices and non-governmental organisations (NGOs) in conducting complex surveys with dynamic structures using tablet devices, as was the case in the CS 2016. The software can be tailored to the needs of specific clients, allowing them to successfully complete simple and more sophisticated projects: from basic evaluation questionnaires to complicated multi-stage panel surveys. The software is offered free of charge and surveys can be conducted on low-cost Android tablets.

Some of the key features of Survey Solutions include automated data validations at entry, real-time availability of results and automated survey management. This led to improved timeliness of data collection and minimised editing and imputation post-data collection; and imputation rates emanating from CS 2016 are discussed in detail in section 2. Consequently, CS 2016 results were released only two months following data collection – a record-breaking time line for the organisation.

Nevertheless, CAPI usage in CS 2016 was not without its own associated flaws. A number of errors during fieldwork were reported, mostly in connection with the Go Survey navigation tool, a completely independent system from Survey Solutions, but which was used in conjunction with the system. After the sample is uploaded Go Survey is used to navigate to a sampled point, then a household is enumerated using the Survey Solutions interviewer application. A number of fieldworkers reported cases where Go Survey did not provide routes to specific points; was not able to locate points, points mismatched between Go Survey and interviewer. Other reported errors mainly revolved around the synchronisation process, whereby devices struggled or could not send completed questionnaires to the headquarters and timely receive new assignments leading to unprecedented delays in the enumeration process.

1.1.5 Survey Coordination, Monitoring and Evaluation (SCM&E)

The Survey Coordination, Monitoring and Evaluation (SCM&E) Division in the Survey Operations Cluster is responsible for the monitoring and evaluation of the quality aspects of all population and household survey processes in the organisation. As part of the enhancement of quality, SCM&E Division monitored CS 2016 field operations activities in all provinces. The objectives of the CS 2016 monitoring included conducting quality checks on the collected data, conducting verification

on in- and out-of-scope cases and monitoring any other issues that can have an impact on data quality with the purpose of compiling lessons learnt. For the purpose of the study, monitors also used the tablets to conduct the monitoring activities. Their tablets were loaded with some questionnaires with observation rights only, questionnaire quality checks, out-of-scope verification and control visit forms.

A total of 7 184 questionnaires were checked in the entire data collection period and 5 376 (74,8%) errors were identified. During the interview observations, 1 852 observations and 3 501 control visits were conducted while a total of 7 870 points were verified and 3 787 DUs were in agreement while 4 083 were not in agreement.

Intensive training is required during the use of technology, from the Questionnaire, Navigation system and CAPI. In addition, the administration of the DURF at a point needs close Monitoring because in instances where there is growth or shrinkage, the FWs might decide to leave the DUs unlisted to avoid more work load. Intensive and continuous training on the use of DURF should be encouraged. The FWSs and DSCs should take a responsibility of observing the FWs especially during the first two weeks of data collection. DSCs should conduct immediate and continuous quality checks and identify the non-response cases which should immediately be verified by the FWSs.

1.1.6 Data editing strategy

Quality assurance in CS 2016 was largely automated and handled in two phases. The first phase of quality assurance involved the electronic questionnaire being subjected to conditions and validation rules. This process eliminated unnecessary inconsistencies in the data during data collection. An additional automated quality assurance process was used during data collection where completed questionnaires were flagged as REJECTED or ACCEPTED based on minimum processability rules. Any questionnaire submitted to database that did not meet the set minimum rules were marked as REJECTED, and sent back to the fieldworker for verification and correction. The fieldwork supervisors were involved in taking note of the flagged questionnaires and assist the fieldworker in correcting the mistakes accordingly. For any record marked as REJECTED once, the running of the rejection was done at least for four different times and at different dates. This was necessary for the fieldworker to try and correct mistakes before a particular questionnaire could be declared “Complete”. This process contributed tremendously in reducing missing values on a number of questions.

1.2 Objectives of the report

- To increase awareness to users on data/information gaps inherent in CS 2016 data
- To provide imputation rates in CS 2016 data
- To provide confidence intervals of key selected variables for CS 2007 and 2016
- To highlight consistency for selected indicators between CS 2007 and CS 2016

1.3 Data evaluation procedure

Data interrogations commenced at around the third week of data collection using incoming data. Direct (consistency checks within the same data) and indirect analysis (comparability with other data) techniques were utilised. Towards the end of data collection international consultants were recruited for the purpose of providing expert advice to both the Statistics Council and the Statistician-General. Areas of assessment included evaluation of pre-enumeration processes, data collection systems used during different phases, sample size and weighting strategies used as well as plausibility of demographic estimation outcomes.

Broadly, CS 2016 data show expected trends for a substantial number of variables when compared to Census time series data over time. That notwithstanding, there are some indicators that reflect unexpected results. The total number of persons who reported that they were born outside South Africa is much lower (1,6 million) compared to 2,1 million collected during Census 2011. In addition, there seems to be unexpectedly fewer persons reporting internal migration even to the two known receiving provinces namely, Gauteng and Western Cape relative to previous censuses. Interestingly, the proportion of the employed is much lower than that provided by the Quarterly Labour Force Survey (QLFS) collected in the first quarter of 2016. As a result, such data would need to be first subjected to coding as it was done for Census 2011 prior to further evaluation. Another unexpected finding relates to fewer children reporting deceased parents compared to Census 2011 data.

1.4 Overview of sections and method of analysis

This report/statement mainly focuses on comparing data from CS 2007 and CS 2016 in order to gauge and comprehend data quality of CS 2016. This section sets the scene while describing numerous processes prior and after field operations. Admittedly, output data are affected during all phases prior to editing and even at such a stage and later at the time of tabulation and dissemination. Different methods of analysis are applied to apprehend the differences and similarities of the two surveys.

Section two of the report presents imputation rates in tabular formats for all sections of the CS 2016 questionnaire. These imputation rates give the percentage at which values of each variable in the survey were changed in order to account for missing and inconsistent data. In the analysis, an imputation rate will be considered statistically significant if its value is less than or equals to 5,0%. Section three presents confidence intervals which are calculated to give the estimated values of key variables that are comparable between CS 2007 and CS 2016 and the interval where these estimates have a certain probability to lie within it. Indirect data assessment involving comparison of selected indicators between Community Survey 2007 and 2016 are presented and discussed in Section four. Lastly, conclusion and recommendations for future household surveys are elaborated on in Section five.

SECTION 2: IMPUTATION RATES

2.1 Introduction

In the following section, data quality of variables is expressed by looking at imputation rates for each variable. As a standard, it is generally considered by the UN that any variable imputed for less than 10% of its cases is acceptable, and for less than 5% it is very good. However, for the purpose of this report if a variable is imputed for less than 2% of its cases it is considered very good and acceptable if imputation rates range between 2% and 4,99%. In CS 2016, limited time was allocated for the editing process during the planning phase relative to previous surveys. This was based on the impression that output data would need minimal editing, given the built-in validation and enabling conditions in the electronic questionnaire.

Data editing and imputation is mainly done to account for missing, invalid or inconsistent data. Missing data occur primarily as a result of item non-responses i.e. no information was provided for one or more data items. Missing values can reduce the representativeness of the sample and can also falsify the inference about the population. Hence the necessity of applying different imputation methods to account for missing values. Thus, an imputation rate is the number of imputed values per variable divided by the total population (persons or households) of the same variable, multiplied by 100. See formula below:

$$Imputation\ Rate = \frac{\sum_{i=0}^{n} Xi}{\bar{X}}\ x\ 100$$

Where $\sum_{i=0}^{n} Xi$ *the sum of imputed values and* $\bar{X}$ *is the total population (person or households) of the variable.*

Given the rigorous built-in checks embedded into the CAPI questionnaire, imputation on CS 2016 data was limited to the following methods:

- *Logical imputation*, which rely solely on deducing the replacement value from data available from other data items within the same survey (i.e. using logical relationships between variables).
- *Hot-deck imputation*, which result from cell mean imputations. This method uses information of the nearest neighbour, information of a randomly selected respondent and information from the previous nearest respondent that had information on the same variable.

SAS Enterprise Guide 7.1 software and Microsoft Excel were used to calculate imputation rates. Below is a list of tables with imputation rates for each variable released in CS 2016 data. Derived variables and/or indicators were not included as they were not imputed but derived from other variables.

2.2 The person file

The person file gives information on characteristics of household members. Any information regarded as essential in profiling the population (persons) of South Africa is included in the file. The information covers demographical information of all persons, movements (migration), general health and functioning for persons aged 5 years and older, parental survival, education and fertility of women aged between 12 years and 50 years.

2.2.1 Demographics

Table 2.1: Demographics imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Demographics			
Sex	4	3 328 867	0,000
Year of birth	1 048	3 328 867	0,031
Month of birth	4	3 328 867	0,000
Age	158	3 328 867	0,005
Relationship to household head	485	3 328 867	0,015
Marital Status	4 447	2 526 309	0,176
Population group	33	3 328 867	0,001
Language most spoken in the household	5 449	3 267 574	0,167
Religious belief	0	3 328 867	0,000
Christianity	0	2 614 167	0,000

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

None of the variables in the demographics sections were imputed for above 2%. Furthermore, all variables in this section were statistically significant at 0,5%.

2.2.2 Lifetime immigration and internal migration

Table 2.2: Migration variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Province of birth	6 445	3 328 867	0,194
Country of birth	1 555	81 244	1,914
Year in which the person moved to SA	1 037	81 244	1,276
Country of citizenship	2 141	3 328 867	0,064
Usual residence	96	3 328 867	0,003
Province of usual residence	0	23 482	0,000
Local/metropolitan municipality of usual residence	0	23 482	0,000
Since Oct 2011	8 542	3 327 640	0,257
Year moved to current place	31	199 018	0,016
Month moved to current place	1 651	199 018	0,830
Province of previous residence	0	199 018	0,000
Local/metropolitan municipality of previous residence	0	199 018	0,000
Main reason for moving current place	0	199 018	0,000

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

The majority of variables in the migration section were statistically significant at 1%. Nonetheless, Country of birth and Year in which person moved to South Africa were significant at 2% with imputation rates of 1,914% and 1,276% respectively.

2.2.3 General health and functioning

Table 2.3: General health and functioning variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Difficulty in seeing	4 796	3 003 210	0,160
Difficulty in hearing	4 796	3 003 210	0,160
Difficulty in communicating	4 796	3 003 210	0,160
Difficulty in walking	4 796	3 003 210	0,160
Difficulty remembering	4 796	3 003 210	0,160
Difficulty with self-care	4 796	3 003 210	0,160
Use eyeglasses/spectacles/contact lenses	4 796	3 003 210	0,160
Use a hearing aid	4 795	3 003 210	0,160
Use a walking stick, walking frame or crutches	4 795	3 003 210	0,160
Use a wheelchair	4 796	3 003 210	0,160
Use any other assistive device/aid	4 796	3 003 210	0,160

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

All variables in the general health and functioning section were imputed and are all significant at 0,5%. The imputation rates are similar for all the variables since imputation was done for the same cases.

2.2.4 Parental survival

Table 2.4: Parental survival variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Biological mother alive	0	3 328 867	0,000
Year in which biological mother passed away	1 102	863 543	0,128
Biological mother part of the household	0	2 372 111	0,000
Person number of biological mother	33 477	1 325 304	2,526*
Biological father alive	0	3 328 867	0,000
Year in which biological father passed away	37 887	1 168 236	3,243*
Biological father part of the household	0	1 885 172	0,000
Person number of biological father	21 117	572 442	3,689*

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

Three variables in the parental survival section were found to be statistically insignificant at 2%. These variables collected information on the number of the biological mother and biological father still alive or no longer alive. The variables “Person number of biological mother” and “Person number of biological father” which aimed to identify the respondent’s parent(s) in cases they were still alive and stayed in the same household were imputed because no checks were added to validate responses given to these specific questions; i.e. male household members being identified as mothers and vice versa.

2.2.5 Education

Table 2.5: Education variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Attendance	0	3 328 867	0,000
Educational institution type	73 080	1 152 877	6,339**
Type of educational institution	0	1 152 877	0,000
Mode of transport education-walking	0	1 152 877	0,000
Mode of transport education-bicycle	0	1 152 877	0,000
Mode of transport education- Motorcycle/scooter	0	1 152 877	0,000
Mode of transport education -Minibus taxi/sedan taxi	0	1 152 877	0,000
Mode of transport education-Bakkie taxi	0	1 152 877	0,000
Mode of transport education - Metered taxi	0	1 152 877	0,000
Mode of transport education- Public bus	0	1 152 877	0,000

Variable name	Total cases imputed	N	Imputation rates
Mode of transport education - Train	0	1 152 877	0,000
Mode of transport education-Vehicle provided by the institution	0	1 152 877	0,000
Mode of transport education- Vehicle provided by government and not paid for	0	1 152 877	0,000
Mode of transport education-Vehicle hired by a group of parents/students	0	1 152 877	0,000
Mode of transport education-Own car/private vehicle	0	1 152 877	0,000
Mode of transport education- Animal-drawn transport/use of animals	0	1 152 877	0,000
Mode of transport education- Other	0	1 152 877	0,000
Highest level of education	162 881	3 328 867	4,893*
Higher educational institution type	758	210 446	0,360
Field of higher educational institution	20	125 839	0,016
Field of TVET	81	74 410	0,109

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

Table 2.5 indicates the highest imputation rate for Educational institution type variable in the education section, which is insignificant at 5%. The variable was mainly imputed due to inconsistencies between the age of the respondent and the educational institution that they had indicated to be attending. Highest level of education, on the other hand is statistically insignificant at 2% with the imputation rate of 4,893%, it was also imputed on the basis of age being inconsistent with the highest level of education.

2.2.6 Fertility

Table 2.6: Fertility variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Ever given birth	2 081	1 012 535	0,206
Total number of children still alive	3 875	565 600	0,685
Total number of boys still alive	406	562 026	0,072
Total number of girls still alive	509	562 026	0,091
Total number of children no longer alive	1 475	565 600	0,261
Total number of boys no longer alive	678	45 473	1,491
Total number of girls no longer alive	680	45 473	1,495
Total number of children born alive	1 703	565 600	0,301
Total number of boys born alive	11 823	565 600	2,090*
Total number of girls born alive	16 371	565 600	2,894*
Year of birth of last child born	2 281	565 600	0,403
Sex of last child born	26	565 600	0,005
Breastfeeding last child	27	565 600	0,005
Length of breastfeeding last child	14 406	445 712	3,232*
Last child born still alive	23	565 600	0,004
Year of death of last child born	556	10 027	5,545**

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

A total of three variables were statistically insignificant at a percentage between 2% and 4,99 %, i.e. total number of boys born alive, total number of girls born alive and length of breastfeeding of last child born. Year of death of last child born variable was imputed because of inconsistencies between date of birth and date of death.

2.3 The household file

The household file provides information on the characteristics of the household. It covers information on perceptions of service delivery (problem facing the municipality), satisfaction with basic services, type of dwelling, household goods, services that the household have access to, agricultural activities and food security, crime experienced and household mortality.

2.3.1 Household types and goods

Table 2.7: Household type and household goods variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Dwelling Unit			
Main dwelling that the household currently lives in	0	984 627	0,000
Tenure status	0	984 627	0,000
Possession of title deed	0	829 725	0,000
RDP or government subsidised dwelling	0	984 627	0,000
Rating overall quality of RDP or government subsidised dwelling	0	228 413	0,000
Households goods			
Refrigerator/Freezer	0	984 627	0,000
Electric/Gas stove	0	984 627	0,000
Vacuum cleaner/ Floor polisher	0	984 627	0,000
Washing machine	0	984 627	0,000
Tablet/Phablet	0	984 627	0,000
Personal computer/Desktop/Laptop	0	984 627	0,000
Satellite decoder	0	984 627	0,000
Motor vehicle	0	984 627	0,000
Television	0	984 627	0,000
Radio	0	984 627	0,000
DVD Player/Blu-ray player	0	984 627	0,000
Home theatre system	0	984 627	0,000
Landline	0	984 627	0,000
Cell phone	0	984 627	0,000
Microwave oven	0	984 627	0,000
Geyser (providing hot water)	0	984 627	0,000
Air conditioner (excluding fans)	0	984 627	0,000
Internet services	0	984 627	0,000

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

No imputations were done on dwelling unit and household goods variables.

2.3.2 Basic service delivery

Table 2.8: Basic services variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Water			
Main source of water for drinking	72	984627	0,007
Distance to get main source of water for drinking	0	266 156	0,000
Access to safe water supply drinking service	0	984 627	0,000
Supplier of the main source of drinking water	0	984 627	0,000
Municipal water interruption in the past 3 months	0	781 477	0,000
How long the water interruption lasted	0	200 840	0,000
Water interruption longer than 2 days	0	200 840	0,000
Alternative water source during interruptions	0	200 840	0,000
Sanitation			
Main type of toilet facility used	212	984 627	0,022
Main toilet facility in the dwelling/yard/outside the yard	0	955 972	0,000
Is the toilet facility shared	0	955 987	0,000
Maintenance of the toilet facility	170 252	955 994	17,809**
Electricity			
Household access to electricity	136	984 627	0,014
Household electricity supplier	0	860 234	0,000
Interruption in electricity in the past 3 months	0	860 234	0,000
Did electricity interruptions last longer than 12 hours?	0	150 177	0,000
Energy source			
Main source of energy for cooking	0	984 627	0,000
Main source of energy for lighting	0	984 627	0,000
Main source of energy for water heating	0	984 627	0,000
Main source of energy for space heating	0	984 627	0,000
Expenditure on energy sources			
Electricity	0	984 627	0,000
Paraffin	0	984 627	0,000
Gas	0	984 627	0,000
Candles	0	984 627	0,000
Coal	0	984 627	0,000
Firewood	0	984 627	0,000
Solar system	0	984 627	0,000
Car batteries	0	984 627	0,000
Other batteries	0	984 627	0,000
Generator	0	984 627	0,000
Other	0	984 627	0,000
Saving energy in the Household			

Variable name	Total cases imputed	N	Imputation rates
Switching off your lights when leaving home	0	898 170	0,000
Using energy-saving light bulbs	0	898 170	0,000
Switching off all your lights, except security lights in the home when not in use	0	898 170	0,000
Switching off appliances at the wall when not in use	0	898 170	0,000
Switching off geyser at certain times	0	898 170	0,000
Boiling only as much water with a pot or kettle that is needed	0	898 170	0,000
Using stove plates and oven as little as possible	0	898 170	0,000
Using warm clothing or blankets instead of an electric heater	0	898 170	0,000
Closing windows and doors when heater is on	0	898 170	0,000
Allowing clothes to drip-dry instead of ironing	0	898 170	0,000
Installing a solar water heater instead of an electric geyser	0	898 170	0,000
Insulating household's geyser and hot pipes	0	898 170	0,000
Refuse removal	902	984 627	0,092
Internet services			
Internet services	0	984 627	0,000
Internet-Connection in the dwelling	0	984 627	0,000
Internet-Connection from a library/community hall/Thusong centre	0	984 627	0,000
Internet-at a school/university/college	0	984 627	0,000
Internet-Connection at a place of work	0	984 627	0,000
Internet-Cafe 2km or less from the dwelling	0	984 627	0,000
Internet-Cafe more than 2km from the dwelling	0	984 627	0,000
Internet-Any place via a cell phone	0	984 627	0,000
Internet-Any place via other mobile access service	0	984 627	0,000
Internet-Other	0	984 627	0,000
Household mode for receiving of mail/post	0	984 627	0,000

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

Very little imputation was done on service delivery variables, only five variables were imputed. Four of the five variables were statistically significant at 1%. The variable "Maintenance of the toilet facility" had the highest imputation rate of 17,809%. This was largely as a result of inconsistencies emanating from the misinterpretation of the question, only households sharing a toilet facility were meant to respond. However, all households responded regardless of the location of their facility, leading to the unusually high imputation rate.

2.3.3 Perception and satisfaction on living conditions and crime

Table 2.9: Perceptions and satisfaction on living conditions variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Problem/Difficulty in the municipality			
Difficulties facing the municipality presently	0	984 627	0,000
The extent to which the municipality is trying to resolve the problem	0	905 527	0,000
Satisfaction with basic services			
Rating of the overall quality of the water services	0	984 627	0,000
Rating of the overall quality of refuse removal services	0	984 627	0,000
Rating of the overall quality of the electricity supply services	0	984 627	0,000
Rating of the overall quality of toilet/sanitation services	0	984 627	0,000
Rating of the overall quality of the local public hospital	0	984 627	0,000
Rating of the overall quality of the local public clinic	0	984 627	0,000
Rating of the overall quality of the local police services	0	984 627	0,000
Rating of the overall quality of the local public school	0	984 627	0,000
Improving the standard of living of the household			
Importance of education to improve the standard of living of the household	0	984 627	0,000
Importance of health to improve the standard of living for the household	0	984 627	0,000
Importance of living conditions to improve the standard of living for the household	0	984 627	0,000
Importance of ownership of household assets to improve the standard of living for the household	0	984 627	0,000
Importance of employment to improve the standard of living for the household	0	984 627	0,000
Importance of safety and security to improve the standard of living for the household	0	984 627	0,000
Perceptions of safety			
Safety during the day	0	984 627	0,000
Safety when it is dark	0	984 627	0,000
Crime experienced by the household			
Victim of crime in the past 12 months	0	984 627	0,000
Experience of crime – Murder	0	67 726	0,000
Experience of crime – Home robbery	0	67 726	0,000
Experience of crime – House breaking	0	67 726	0,000
Experience of crime – Robbery	0	67 726	0,000
Experience of crime – Theft pf livestock, poultry and other animals	0	67 726	0,000
Experience of crime – Theft of motor vehicle and/or motorcycle	0	67 726	0,000
Experience of crime – Other crime	71	67 726	0,105

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

All perception and rating of services variables were not imputed. Only “other crimes experienced by the household” (experience of other crime) was imputed with the imputation of 0,105% which is statistically significant at 1%.

2.3.4 Household involvement in agricultural activities and food security

Table 2.10: Household agricultural activities variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Agricultural activities			
Household involved in agriculture	0	984 627	0,000
Type of agricultural activity – Livestock production	0	124 481	0,000
Type of agricultural activity - Poultry production	0	124 481	0,000
Type of agricultural activity - Grains and food crops	0	124 481	0,000
Type of agricultural activity - Industrial crops	0	124 481	0,000
Type of agricultural activity - Fruit production	0	124 481	0,000
Type of agricultural activity - Vegetable production	0	124 481	0,000
Type of agricultural activity – Other	0	124 481	0,000
Type of farm practice for crop production	0	91 853	0,000
Owning livestock or poultry	0	984 627	0,000
How many cattle owned	0	112 058	0,000
How many sheep owned	0	112 058	0,000
How many goats owned	0	112 058	0,000
How many pigs owned	0	112 058	0,000
How many chickens	0	112 058	0,000
How many other poultry owned	0	112 058	0,000
Food security			
Run out of money to buy food in past 12 months	0	984 627	0,000
Running out of money to buy food for 5 or more days in past 30 days	0	212 645	0,000
Skipped meal in past 12 months	0	984 627	0,000
Skipping meal for 5 or more days in the past 30 days	0	141 889	0,000

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

Table 2.10 indicates that variables that collected information on agricultural activities and food security of households were not imputed at all.

2.4 Emigration file

The emigration file contains information on emigrants i.e. persons who have left South Africa in the past ten years to reside in another country and were still residing there at the time of the CS 2016.

Table 2.11: Emigration variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Emigrants			
Sex of emigrant	7	5 205	0,134
Age of emigrant	8	5 205	0,154
Country where emigrant resides	0	5 205	0,000
Year in which emigrant left South Africa	62	5 205	1,191

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

All the imputed variables were statistically significant at 2%. However, the “Year in which emigrant left South Africa” variable is not statistically significant at 1%.

2.5 Mortality file

The mortality file contains information on deaths that occurred in each household in the 12 months preceding CS 2016, it profiles the sex of the deceased, age at the time of death and date of death as well as maternal deaths.

Table 2.12: Mortality variables imputation rates, CS 2016

Variable name	Total cases imputed	N	Imputation rates
Sex of the deceased	0	30 053	0,000
Age of the deceased	7	30 058	0,023
Year of death	131	30 058	0,436
Month of death	72	30 053	0,240
Maternal deaths	0	5 329	0,000

*Imputation Rate>=2,00 and Imputation Rate<=4,99
**Imputation Rate>5,00

All variables are statistically significant at 1%. Minimal imputations were done as a result of inconsistencies in the date of death, respondents tended to report deaths that occurred outside the 12-month stipulated period.

2.6 Conclusion

Overall, imputation rates for CS 2016 were statistically significant at 5,0%. A majority of the variables had imputation rates less than 1%. However, few variables were statistically insignificant at 5,0%; these included Year of death of last child born; Educational institutions type and maintenance of the toilet facility, which had imputation rates standing at 5,5%, 6,3% and 17,8%, respectively.

Results show a high concentration of missing and inconsistent data in the fertility, education, migration, emigration sections, and also on one of the newly-introduced question, the “maintenance of the toilet facility used by the household”.

Approximately seven imputed variables in the person file were statistically insignificant at between 2% and 4,99 %, as compared to the household file, emigration file and mortality file which had no variable insignificant in the same range. Only two variables are statistically insignificant at 5% for the person file as compared to one in the household file.

SECTION 3: CONFIDENCE INTERVALS OF SELECTED VARIABLES

3.1 Introduction

In order to measure the precision of key variables comparable between CS 2007 and CS 2016, confidence intervals were calculated. Statisticians use a confidence interval to express the degree of uncertainty associated with a sample statistic. A confidence interval is an interval estimate combined with a probability statement. It gives the most likely range of the unknown population or estimate.

For the purpose of this report, confidence intervals for CS2007 and CS2016 were calculated using SAS Enterprise Guide 7.1. Due to repeated municipal demarcation changes in South Africa, CS 2007 and CS 2016 confidence intervals are calculated based on geographical boundaries of their respective preceding censuses (2001 and 2011). CS 2007 confidence intervals were calculated using records from version 9 of the person and household files.

3.2 Confidence Intervals for Person file variables

This sub-section looks at person-level variables' confidence intervals. There were 270 municipalities in the CS 2007 as per the on 2001 geographical boundaries, whereas 234 municipalities were on the 2011 Census data set.

Table 3.1: Confidence interval for sex variable, CS2007 and CS 2016

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Sex	Estimate	Lower Limit	Upper Limit	Sex	Estimate	Lower Limit	Upper Limit
Male	23 412 064	23 356 118	23 468 010	Male	27 247 226	27 208 337	27 286 115
Female	25 089 999	25 035 238	25 144 760	Female	28 406 428	28 369 317	28 443 539
Total	48 502 063	48 464 849	48 539 277	**Total**	55 653 654	55 621 573	55 685 736

Table 3.2: Confidence interval for age variable, CS 2007 and CS 2016

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Age	Estimate	Lower Limit	Upper Limit	Age	Estimate	Lower Limit	Upper Limit
0	1 003 597	988 780	1 018 414	0	1 126 133	1 116 238	1 136 028
1	1 059 554	1 044 374	1 074 734	1	1 209 787	1 199 466	1 220 109
2	1 011 748	996 903	1 026 593	2	1 225 654	1 215 270	1 236 039
3	976 525	961 850	991 199	3	1 207 625	1 197 390	1 217 860
4	933 753	919 409	948 097	4	1 207 319	1 196 964	1 217 675
5	975 908	960 792	991 025	5	1 134 067	1 124 566	1 143 569
6	1 048 804	1 033 198	1 064 411	6	1 115 073	1 105 462	1 124 684

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Age	Estimate	Lower Limit	Upper Limit	Age	Estimate	Lower Limit	Upper Limit
7	1 049 459	1 033 920	1 064 998	7	1 141 129	1 131 661	1 150 596
8	1 005 905	990 625	1 021 184	8	1 115 792	1 106 498	1 125 086
9	1 036 737	1 021 198	1 052 275	9	1 113 735	1 104 389	1 123 081
10	992 157	977 889	1 006 425	10	1 140 862	1 131 147	1 150 576
11	975 104	960 943	989 265	11	1 093 216	1 083 664	1 102 767
12	995 001	980 784	1 009 218	12	1 006 860	997 688	1 016 033
13	983 492	969 357	997 628	13	982 834	973 868	991 800
14	1 001 248	986 958	1 015 537	14	966 031	957 114	974 948
15	1 018 396	1 003 979	1 032 812	15	1 033 154	1 024 113	1 042 195
16	1 060 590	1 045 776	1 075 405	16	1 041 938	1 032 808	1 051 068
17	1 047 549	1 032 892	1 062 207	17	1 001 924	992 937	1 010 911
18	1 002 363	987 855	1 016 871	18	1 011 525	1 002 492	1 020 559
19	964 241	949 804	978 677	19	1 015 941	1 007 057	1 024 825
20	981 625	967 046	996 204	20	1 057 160	1 047 611	1 066 710
21	990 984	976 220	1 005 748	21	1 091 182	1 081 533	1 100 831
22	961 272	946 658	975 887	22	1 049 152	1 039 651	1 058 654
23	914 732	900 567	928 897	23	1 062 403	1 052 859	1 071 948
24	943 195	928 742	957 647	24	1 042 437	1 033 080	1 051 794
25	852 936	838 904	866 968	25	1 125 371	1 115 554	1 135 188
26	850 045	836 006	864 083	26	1 080 318	1 070 700	1 089 935
27	847 916	833 876	861 957	27	1 035 171	1 025 720	1 044 622
28	765 956	752 517	779 396	28	1 013 063	1 003 719	1 022 406
29	747 939	734 668	761 211	29	1 026 582	1 017 150	1 036 014
30	771 367	757 878	784 856	30	975 147	966 289	984 004
31	779 016	765 491	792 541	31	915 913	907 203	924 623
32	756 710	743 394	770 026	32	885 452	876 863	894 041
33	697 829	685 043	710 616	33	897 635	888 985	906 286
34	733 566	720 544	746 587	34	780 541	772 584	788 498
35	635 440	623 287	647 592	35	859 477	850 576	868 378
36	682 251	669 648	694 854	36	827 193	818 541	835 846
37	670 958	658 593	683 322	37	721 378	713 318	729 438
38	650 199	638 091	662 307	38	714 270	706 115	722 425
39	578 783	567 393	590 174	39	725 642	717 518	733 766
40	562 123	550 999	573 248	40	725 239	717 288	733 191
41	580 473	569 198	591 748	41	676 504	668 969	684 040
42	571 678	560 503	582 852	42	623 521	616 317	630 725
43	567 682	556 561	578 803	43	667 073	659 410	674 736
44	553 342	542 390	564 293	44	568 247	561 181	575 312
45	480 398	470 172	490 624	45	620 193	612 969	627 417
46	511 193	500 669	521 718	46	591 521	584 329	598 714
47	491 280	480 899	501 661	47	576 612	569 651	583 573
48	472 457	462 180	482 734	48	498 947	492 212	505 683
49	452 954	442 993	462 916	49	495 657	489 229	502 086

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Age	Estimate	Lower Limit	Upper Limit	Age	Estimate	Lower Limit	Upper Limit
50	423 815	414 188	433 443	50	501 744	495 238	508 249
51	393 004	383 754	402 253	51	485 418	479 017	491 819
52	390 938	381 710	400 166	52	477 720	471 442	483 997
53	369 181	360 155	378 207	53	468 890	462 717	475 063
54	390 447	381 119	399 774	54	400 738	394 835	406 641
55	316 996	308 547	325 446	55	439 421	433 207	445 636
56	338 873	330 256	347 490	56	407 092	401 348	412 837
57	328 902	320 340	337 464	57	396 202	390 432	401 972
58	298 490	290 342	306 638	58	371 460	365 953	376 967
59	282 891	274 879	290 904	59	360 021	354 393	365 650
60	261 681	253 965	269 397	60	348 000	342 762	353 239
61	239 963	232 642	247 283	61	327 054	321 894	332 213
62	222 992	215 891	230 092	62	312 170	307 056	317 284
63	198 746	192 061	205 430	63	325 057	319 961	330 153
64	234 304	227 140	241 468	64	260 636	256 002	265 271
65	195 496	188 884	202 108	65	281 613	276 905	286 320
66	227 484	220 422	234 547	66	260 827	256 283	265 370
67	209 692	202 869	216 514	67	232 888	228 420	237 355
68	174 290	168 047	180 533	68	212 099	207 859	216 339
69	155 234	149 269	161 199	69	191 861	187 904	195 818
70	155 771	149 858	161 684	70	191 532	187 309	195 755
71	138 854	133 259	144 449	71	171 991	167 826	176 156
72	128 609	123 222	133 996	72	157 814	153 714	161 915
73	119 790	114 618	124 962	73	169 809	165 820	173 798
74	122 072	116 849	127 294	74	133 587	130 223	136 950
75	90 289	85 765	94 814	75	133 834	130 429	137 238
76	122 467	117 272	127 661	76	111 049	108 146	113 952
77	109 922	105 031	114 814	77	90 831	87 880	93 781
78	84 526	80 167	88 886	78	76 704	74 096	79 313
79	72 882	68 584	77 180	79	73 919	71 041	76 797
80	65 081	61 237	68 925	80	64 762	62 435	67 089
81	62 642	58 646	66 637	81	54 583	52 517	56 650
82	55 758	52 128	59 388	82	48 596	46 701	50 491
83	41 965	38 632	45 298	83	47 423	45 367	49 479
84	38 351	35 065	41 637	84	35 643	33 909	37 377
85	36 066	33 203	38 928	85	44 565	42 803	46 328
86	42 838	39 762	45 914	86	36 036	34 526	37 546
87	30 921	28 312	33 529	87	25 081	23 623	26 539
88	32 397	29 766	35 028	88	19 768	18 621	20 915
89	17 996	16 020	19 972	89	15 342	14 266	16 417
90	12 444	10 818	14 069	90	13 800	12 757	14 843
91	8 896	7 560	10 232	91	10 957	10 022	11 893
92	10 768	9 292	12 244	92	8 314	7 510	9 118

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Age	Estimate	Lower Limit	Upper Limit	Age	Estimate	Lower Limit	Upper Limit
93	7 598	6 257	8 940	93	6 684	6 066	7 301
94	5 461	4 418	6 504	94	6 302	5 653	6 951
95	4 255	3 320	5 190	95	6 999	6 386	7 611
96	5 829	4 744	6 915	96	4 056	3 645	4 468
97	4 860	3 852	5 868	97	4 355	3 899	4 810
98	5 190	4 157	6 223	98	2 116	1 799	2 434
99	2 108	1 434	2 781	99	983	779	1 187
100	2 166	1 514	2 819	100	986	772	1 201
101	1 598	1 032	2 164	101	1 151	939	1 363
102	1 604	1 037	2 172	102	637	480	794
103	2 045	1 382	2 709	103	613	451	775
104	1 294	763	1 824	104	358	237	480
105	1 956	1 306	2 607	105	559	406	712
106	1 221	703	1 739	106	340	233	448
107	495	177	812	107	190	107	272
108	195	2	388	108	79	25	134
109	54	0	159	109	85	32	139
110	3	0	6	110	91	30	152
111	713	319	1 106	111	75	26	124
112	6	0	17	112	64	15	113
113	62	0	179	113	69	19	118
114	36	0	106	114	88	26	150
115	1	0	3	115	219	101	338
116	3	0	6	116	100	8	193
117	5	1	9	117	0	0	0
118	60	0	174	118	0	0	0
119	37	0	106	119	0	0	0
120	56	0	160	120	0	0	0
Total	48 502 063	48 464 849	48 539 277	**Total**	55 653 654	55 621 573	55 685 736

Table 3.3: Confidence interval for broad age variable, CS 2007 and CS 2016

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Broad age groups	Estimate	Lower Limit	Upper Limit	Broad age groups	Estimate	Lower Limit	Upper Limit
0-14 years	15 048 992	14 999 782	15 098 201	0-14 years	16 786 118	16 753 168	16 819 069
15-34 years	17 688 227	17 636 140	17 740 314	15-34 years	20 142 009	20 107 091	20 176 927
35-59 years	11 994 749	11 949 211	12 040 287	35-59 years	14 200 181	14 168 976	14 231 387
60 + years	3 770 095	3 741 903	3 798 288	60 + years	4 525 346	4 506 435	4 544 256
Total	48 502 063	48 464 849	48 539 277	**Total**	55 653 654	55 621 573	55 685 736

Table 3.4: Confidence interval for population group variable, CS 2007 and CS 2016

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Population Group	Estimate	Lower Limit	Upper Limit	Population Group	Estimate	Lower Limit	Upper Limit
Black African	38 255 165	38 210 763	38 299 567	Black African	44 891 603	44 864 007	44 919 199
Coloured	4 375 527	4 350 605	4 400 449	Coloured	4 869 526	4 851 062	4 887 991
Indian/Asian	1 244 634	1 228 136	1 261 132	Indian/Asian	1 375 834	1 362 743	1 388 924
White	4 626 738	4 592 130	4 661 345	White	4 516 691	4 488 006	4 545 376
Total	48 502 063	48 464 849	48 539 277	**Total**	55 653 654	55 621 573	55 685 736

Table 3.5: Confidence interval for attendance variable, CS 2007 and CS 2016

CS 2007				CS 2016			
		95% Confidence Limits				95% Confidence Limits	
Attendance	Estimate	Lower Limit	Upper Limit	Attendance	Estimate	Lower Limit	Upper Limit
Yes	16 157 141	16 108 043	16 206 239	Yes	19 411 189	19 376 911	19 445 466
No	31 693 153	31 641 595	31 744 710	No	36 157 294	36 118 662	36 195 927
Total	47 850 294	47 824 159	47 876 429	Total	55 653 654	55 621 573	55 685 736

Table 3.6: Confidence interval for highest level of education variable, CS 2007

CS 2007			
		95% Confidence Limits	
Highest level of education	Estimate	Lower Limit	Upper Limit
No schooling	4 966 251	4 934 607	4 997 896
Grade R	918 637	904 242	933 033
Grade 1	1 150 635	1 134 592	1 166 678
Grade 2	1 844 792	1 824 740	1 864 844
Grade 3/Std 1/ABET 1	1 826 866	1 807 106	1 846 627
Grade 4/Std 2	1 755 020	1 736 046	1 773 994
Grade 5/Std 3/ABET 2	2 174 373	2 153 598	2 195 147
Grade 6/Std 4	2 225 820	2 204 779	2 246 861
Grade 7/Std 5/ABET 3	2 779 951	2 756 463	2 803 438
Grade 8/Std 6	3 377 762	3 351 881	3 403 642
Grade 9/Std 7/ABET 4	3 062 376	3 037 549	3 087 203
Grade 10/Std 8/NTC I	3 657 668	3 630 207	3 685 128
Grade 11/Std 9/NTC II	2 888 862	2 864 289	2 913 434
Attended Grade 12 but not completed Grade 12	1 764 247	1 744 487	1 784 008
Grade 12/Std 10/ NTC III (without University Exemption)	4 359 353	4 328 726	4 389 981
Grade 12/Std 10 (with University Exemption)	1 152 766	1 135 828	1 169 704
Certificate with <Std 10/Gr. 12	666 782	654 247	679 317
Diploma with <Std 10/Gr. 12	482 375	471 440	493 311
Certificate with Std 10/Gr. 12	428 744	418 657	438 832
Diploma with Std 10/Gr. 12	805 341	791 144	819 538
Bachelor's Degree	624 466	611 089	637 843
BTech	82 403	77 672	87 134

CS 2007			
		95% Confidence Limits	
Highest level of education	Estimate	Lower Limit	Upper Limit
Post graduate diploma	193 957	186 583	201 331
Honours Degree	214 737	206 871	222 603
Higher Degree (Masters/PhD)	190 859	183 194	198 523
Total	47 850 294	47 824 159	47 876 429

Table 3.7: Confidence interval for highest level of education variable, CS 2016

CS 2016			
		95% Confidence Limits	
Highest level of education	Estimate	Lower Limit	Upper Limit
No schooling	8 183 201	8 159 022	8 207 379
Grade R/0	2 100 147	2 087 214	2 113 080
Grade 1/Sub A/ Class 1	1 471 399	1 460 826	1 481 972
Grade 2/Sub B/ Class 2	1 287 925	1 278 161	1 297 688
Grade 3/Standard 1/ ABET 1	1 948 274	1 936 222	1 960 325
Grade 4/Standard 2	1 896 637	1 884 893	1 908 381
Grade 5/Standard 3/ ABET 2	1 962 510	1 950 422	1 974 599
Grade 6/Standard 4	2 288 259	2 275 028	2 301 490
Grade 7/Standard 5/ ABET 3	2 219 524	2 206 760	2 232 287
Grade 8/Standard 6/ FORM 1	3 062 898	3 047 884	3 077 913
Grade 9/Standard 7/FORM 2/ABET /Occupational Certificate NQF level 1	3 265 838	3 250 375	3 281 301
Grade 10/Standard 8/FORM 3/NCV level 2 /Occupational Certificate NQF level 2	4 577 157	4 558 369	4 595 945
Grade 11/Standard 9/FORM 4/NCV level 3 /Occupational Certificate NQF level 3	4 956 123	4 937 221	4 975 024
Grade 12/Standard 10/FORM 5/NCV level 4 /Occupational Certificate NQF level 4	11 630 801	11 601 087	11 660 516
NTC I/N1	42 761	40 811	44 711
NTC II/N2	54 531	52 205	56 857
NTC III/N3	119 715	116 214	123 216
N4/ NTC 4/ Occupational Certificate NQF Level 5	162 297	158 316	166 278
N5/ NTC 5/ Occupational Certificate NQF Level 5	82 927	80 187	85 667
N6/ NTC 6/ Occupational Certificate NQF Level 5	149 397	145 514	153 279
Certificate with less than Grade 12/STD10	38 198	35 985	40 411
Diploma with less than Grade 12/STD10	85 020	81 891	88 149
Higher/National/Advanced Certificate with Grade 12/STD 10/Occupational Certificate NQF Level5	363 352	356 990	369 713
Diploma with Grade 12/STD 10/Occupational Certificate NQF Level 6	823 563	813 997	833 128
Higher Diploma/Occupational Certificate NQF Level 7	386 818	379 717	393 920
Post-Higher Diploma (University of Technology Masters/Doctoral Diploma)	293 785	287 500	300 071
Bachelor's Degree/Occupational Certificate NQF Level 7	799 772	789 163	810 380
Honours Degree/ Post Graduate Diploma/Occupational Certificate NQF Level 8	363 077	355 612	370 542
Masters/Professional Masters NQF Level 9	146 660	141 480	151 840
PhD(Doctoral Degree)/Professional Doctoral Degree at	74 509	71 113	77 906

CS 2016			
		95% Confidence Limits	
Highest level of education	Estimate	Lower Limit	Upper Limit
NQF Level 10			
Other	186 756	181 904	191 608
Total	55 653 654	55 621 573	55 685 736

Table 3.8: Confidence interval for internal migration variable, CS 2007

CS 2007			
		95% Confidence Limits	
Internal movement	Estimate	Lower Limit	Upper Limit
Yes	34 831 118	34 782 376	34 879 860
No	7 651 124	7 612 458	7 689 791
Born after October 2001	5 368 052	5 335 198	5 400 906
Total	47 850 294	47 824 159	47 876 429

Table 3.9: Confidence interval for internal migration variable, CS 2016

CS 2016			
		95% Confidence Limits	
Internal movement	Estimate	Lower Limit	Upper Limit
Yes	45 976 405	45 941 148	46 011 662
No	3 752 655	3 732 006	3 773 305
Born after October 2011, but never moved	5 828 950	5 807 131	5 850 769
Born after October 2011 and moved	41 450	39 171	43 729
Total	55 653 654	55 621 573	55 685 736

3.3 Confidence Intervals for household file variables

The following sub-section presents household-level confidence intervals for both CS 2007 and CS 2016.

Table 3.10: Confidence interval for main dwelling type variable, CS 2007

CS 2007			
		95% Confidence Limits	
Main dwelling type	Estimate	Lower Limit	Upper Limit
Formal dwelling	9 179 667	9 154 761	9 204 573
Traditional dwelling	1 459 377	1 446 598	1 472 155
Informal dwelling	1 804 430	1 787 089	1 821 772
Other	57 135	53 682	60 588
Total	12 500 609	12 486 754	12 514 464

Table 3.11: Confidence interval for main dwelling type variable, CS 2016

CS 2016			
		95% Confidence Limits	
Main dwelling type	Estimate	Lower Limit	Upper Limit
Formal dwelling	13 404 199	13 381 787	13 426 612
Traditional dwelling	1 180 745	1 173 629	1 187 861
Informal dwelling	2 193 968	2 180 397	2 207 539
Other	142 271	138 753	145 788
Total	16 921 183	16 901 456	16 940 909

Table 3.12: Confidence interval for piped water variable, CS 2007

CS 2007			
		95% Confidence Limits	
Access to piped water	Estimate	Lower Limit	Upper Limit
Piped water inside dwelling/house	5 894 171	5 868 773	5 919 569
Piped water inside yard	2 785 632	2 764 980	2 806 284
Piped water outside yard	2 402 421	2 383 420	2 421 423
No access to piped water	1 418 384	1 405 002	1 431 767
Total	12 500 609	12 486 754	12 514 464

Table 3.13: Confidence interval for piped water variable, CS 2016

CS 2016			
		95% Confidence Limits	
Access to piped water	Estimate	Lower Limit	Upper Limit
Piped water inside dwelling/house	7 511 853	7 489 555	7 534 151
Piped water inside yard	5 081 255	5 064 973	5 097 538
Piped water outside yard	2 625 645	2 612 982	2 638 307
No access to piped water	1 704 556	1 693 961	1 715 151
Total	16 923 309	16 903 582	16 943 036

Table 3.14: Confidence interval for refuse removal variable, CS 2007

CS 2007			
		95% Confidence Limits	
Refuse removal	Estimate	Lower Limit	Upper Limit
Removed by local authority/ private company at least once a week	7 485 569	7 463 846	7 507 292
Removed by local authority/private company less often	210 565	203 922	217 209
Communal refuse dump	269 485	261 916	277 054
Own refuse dump	3 602 713	3 584 391	3 621 036
No rubbish disposal	892 614	880 170	905 058
Other	39 663	36 834	42 492
Total	12 500 609	12 486 754	12 514 464

Table 3.15: Confidence interval for refuse removal variable, CS 2016

CS 2016			
		95% Confidence Limits	
Refuse removal	Estimate	Lower Limit	Upper Limit
Removed by the local authority/private company/community members at least once a week	10 322 257	10 301 225	10 343 288
Removed by the local authority/private company/community members less often than once a week	488 193	481 096	495 290
Communal refuse dump	535 474	528 497	542 451
Communal container/central collection point	314 907	309 476	320 338
Own refuse dump	4 416 606	4 403 287	4 429 924
Dump or leave rubbish anywhere (no rubbish disposal)	669 485	662 142	676 827
Other	176 388	172 375	180 402
Total	16 923 309	16 903 582	16 943 036

Table 3.16: Confidence interval for toilet facility variable, CS 2007

CS 2007			
		95% Confidence Limits	
Toilet facility	Estimate	Lower Limit	Upper Limit
Flush toilet (connected to sewerage system)	6 874 220	6 851 180	6 897 261
Flush toilet (with septic tank)	347 105	338 680	355 530
Dry toilet facility	519 350	509 464	529 236
Pit toilet with ventilation (VIP)	823 634	811 512	835 755
Pit toilet without ventilation	2 587 200	2 569 080	2 605 321
Chemical toilet	43 884	40 935	46 834
Bucket toilet system	273 254	266 311	280 197
None	1 031 961	1 019 414	1 044 509
Total	12 500 609	12 486 754	12 514 464

Table 3.17: Confidence interval for toilet facility variable, CS 2016

CS 2016			
		95% Confidence Limits	
Toilet facility	Estimate	Lower Limit	Upper Limit
Flush toilet connected to a public sewerage system	10 260 829	10 239 593	10 282 064
Flush toilet connected to a septic tank or conservatory tank	461 934	455 177	468 690
Chemical toilet	713 856	707 163	720 549
Pit latrine/toilet with ventilation pipe	2 063 128	2 053 287	2 072 970
Pit latrine/toilet without ventilation pipe	2 315 279	2 303 781	2 326 776
Ecological toilet (e.g urine diversion, enviroloo, etc.)	49 277	47 436	51 118
Bucket toilet (collected by municipality)	244 411	239 708	249 113
Bucket toilet (emptied by household)	132 820	129 341	136 300
Other	271 895	267 187	276 603
None	409 881	404 190	415 571
Total	16 923 309	16 903 582	16 943 036

Table 3.18: Confidence interval for source of energy for lighting variable, CS 2007

CS 2007			
		95% Confidence Limits	
Source of lighting	Estimate	Lower Limit	Upper Limit
Electricity	10 010 273	9 987 156	10 033 390
Gas	20 764	18 674	22 854
Paraffin	658 576	647 595	669 558
Candles	1 713 613	1 697 617	1 729 608
Solar	30 404	28 009	32 799
Other	66 979	63 179	70 780
Total	12 500 609	12 486 754	12 514 464

Table 3.19: Confidence interval for source of energy for lighting variable, CS 2016

CS 2016			
		95% Confidence Limits	
Source of lighting	Estimate	Lower Limit	Upper Limit
Electricity from main	15 231 038	15 210 126	15 251 950
Other source of electricity (e.g. generator, etc.)	31 197	29 370	33 025
Gas	25 700	23 960	27 440
Paraffin	451 602	444 880	458 324
Candles	997 571	988 580	1 006 562
Solar	96 532	93 661	99 403
Other	23 784	22 241	25 327
None	35 498	33 473	37 524
Unspecified	30 387	28 589	32 185
Total	16 923 309	16 903 582	16 43 036

3.4 Conclusion

Overall, as presented above we can be 95% confident that the correct estimates lie within the interval for each variable for both CS 2007 and CS 2016. However, the range for CS 2007 confidence intervals seems wider than those of CS 2016, which presume that estimates for CS 2007 were less accurate as compared to CS 2016. Based on the confidence intervals of selected key variables presented on this section, we can conclude that CS 2016 data is of higher accuracy as compared to that of CS 2007.

SECTION 4: SIMILARITIES AND DIFFERENCES BETWEEN COMMUNITY SURVEY 2007 and 2016

4.1 Introduction

Comparing the CS 2016 to the 2011 Census in terms of data quality would be misleading mainly because the former is a sample survey while the latter is a population Census. Therefore, it is necessary to compare the CS 2016 to another large-scale household sample survey in order to effectively assess its data quality. For this reason, the first part of this section discusses similarities, especially the trend analysis of key indicators between CS 2016 and CS 2007. For data to be considered of good quality, indicators must be comparable well with other surveys conducted by Stats SA and other surveys. It is, however, worth noting that although data from the two surveys are similar in more ways than one, different methodologies were applied, thus, the second part of this section focuses on the differences between the surveys. Evidently, in CS 2007 a paper-based (PAPI) questionnaire was used as the main collection tool, whilst CS 2016, owing to advancement in technology, a CAPI questionnaire was used as the collection tool.

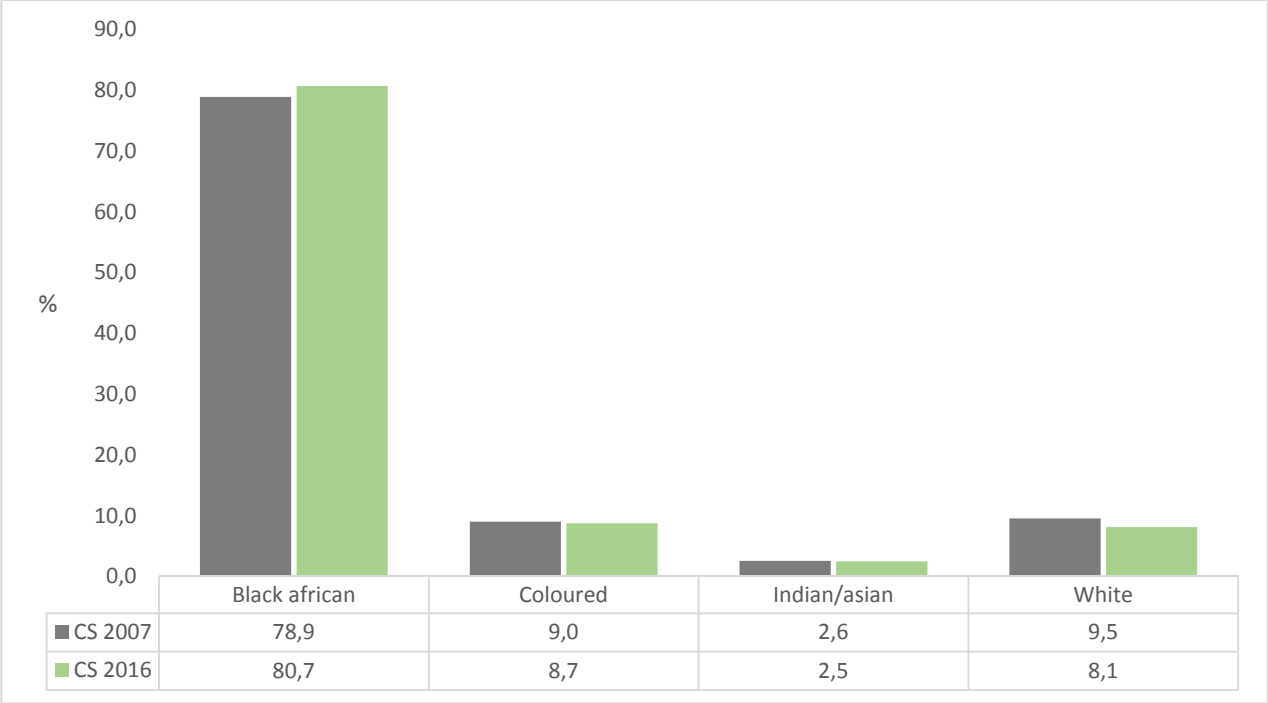
4.2 A comparison of output data for selected indicators between Community Survey 2007 and 2016

4.2.1 PERSON FILE

4.2.1.1 Population structure by population group, CS 2007 versus CS 2016

According to Figure 4.1, the distribution of the South African population by population group has remained fairly comparable between CS 2007 and CS 2016. Minor changes, however, are noticeable between black Africans and white people. Owing to a fairly representative sample in CS 2016 it is safe to deduce that all population groups were adequately enumerated relative to CS 2007 sample design. Another basic difference between the two aforementioned surveys is the enumeration of population with the aim of collecting both de jure and de facto population or not; while CS 2007 collected both, CS 2016 reiterated the Census 2011 question that aimed at collecting the de facto population only.

Figure 4.1: Percentage distribution of total population by population group, CS 2007 and CS 2016

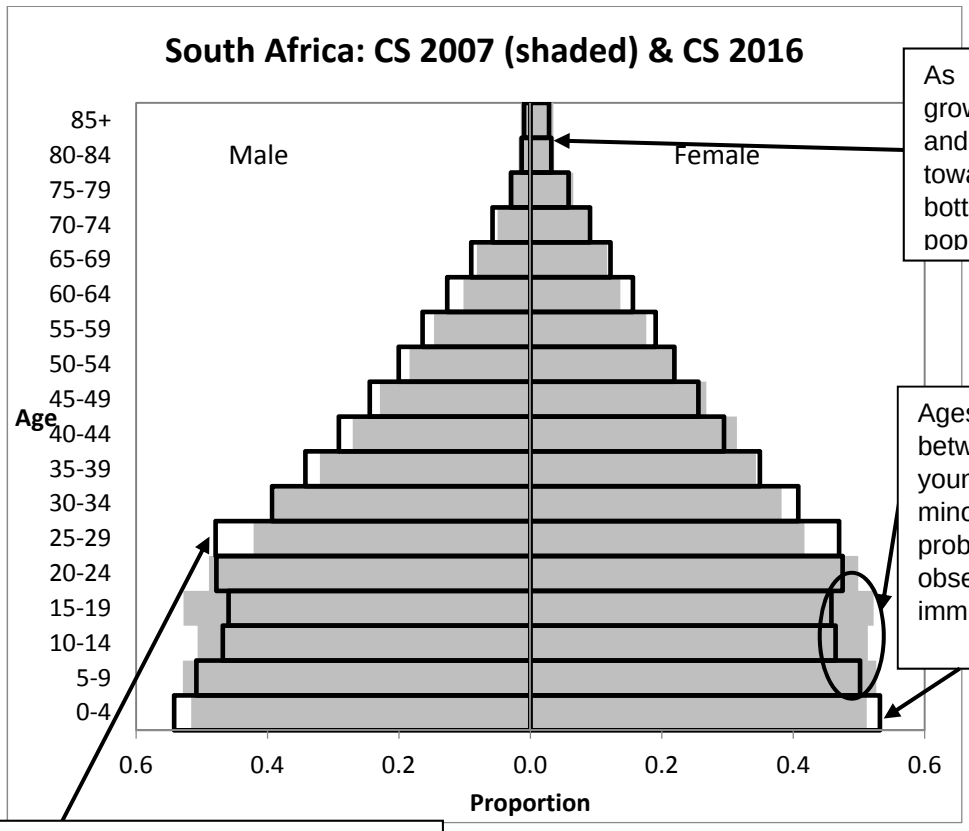


	Black african	Coloured	Indian/asian	White
CS 2007	78,9	9,0	2,6	9,5
CS 2016	80,7	8,7	2,5	8,1

Source: Statistics South Africa

A population structure varies by socio-economic characteristics, especially by population group in the case of South Africa, given the history of racial segregation. It is therefore necessary to depict the population structure by population group. A comparison of population structure by sex and population group is presented in Figure 4.3 to Figure 4.6.

Figure 4.2: South African population structure, CS 2007 and CS 2016

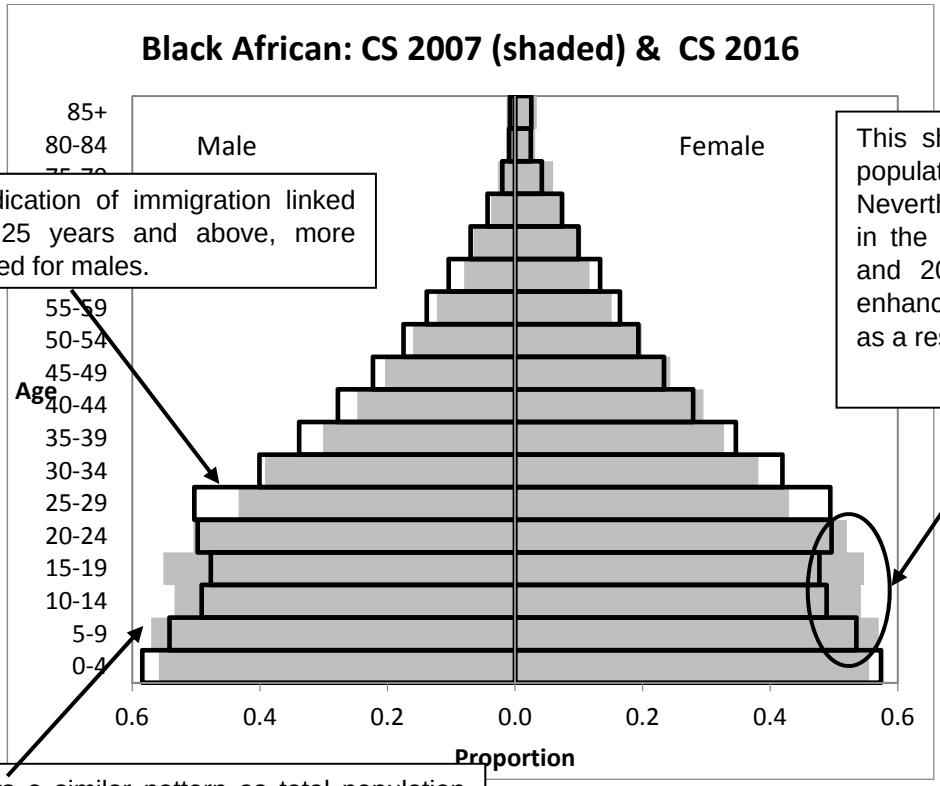


As expected from a young and growing population, both CS 2007 and 2016 populations narrow toward the top and widen at the bottom. This is typical of a growing population.

Ages 5–24 show fluctuating decreases between 2007 and 2016. While the youngest age group (0–4 years) show a minor increase for the same period, probably suggesting fertility increase observed in Census 2011 data or immigration with younger children.

The largest shift can be seen in the ages 10–19 in CS 2007 to ages 25–29 years in CS 2016 for both sexes.

Figure 4.3: Black African population structure, CS 2007 and CS 2016



Some indication of immigration linked to ages 25 years and above, more pronounced for males.

This shows an overall typical growing population, wide base and narrow top. Nevertheless, it shows an upward shift in the ages 5–19 years between 2007 and 2016. This is also indicative of enhanced coverage of younger children as a result of the usage of CAPI.

Shows a similar pattern as total population for both CS 2007 and 2016, a wide base and narrow top depicting the majority rule. No clear origin of depletion of 5–24 ages at face value, this might be linked to fluctuating fertility and emigration of families with children.

Figure 4.4: Coloured population structure, CS 2007 and CS 2016

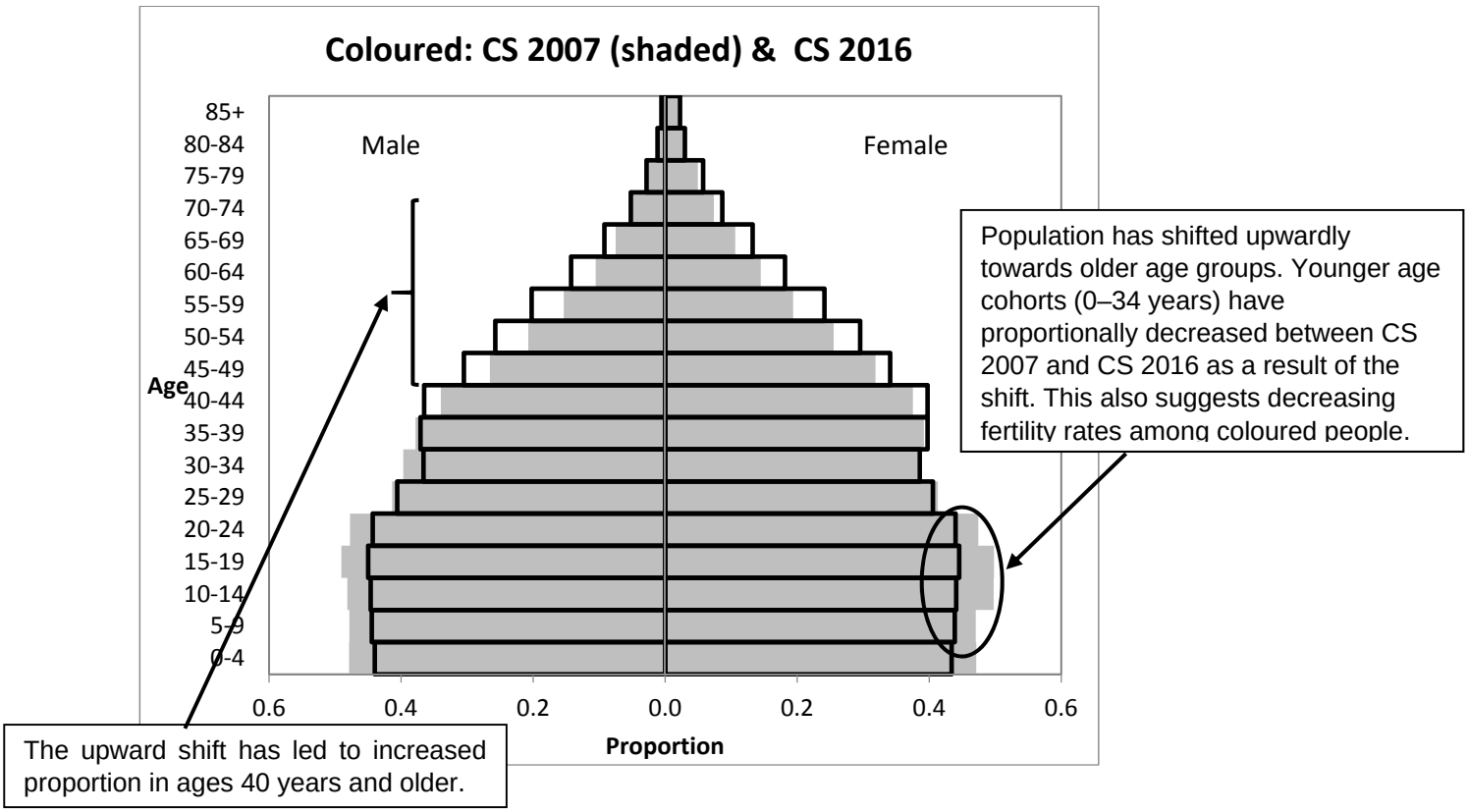
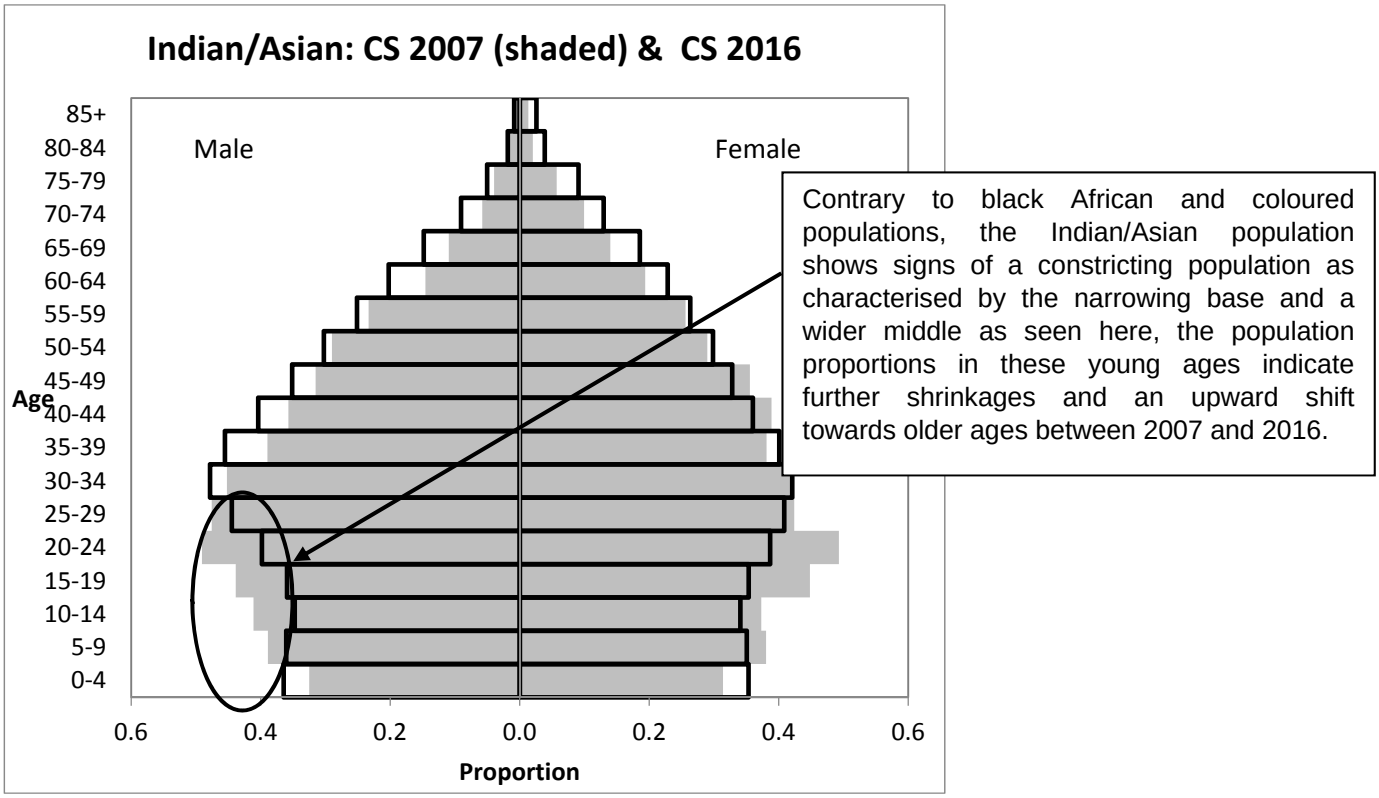
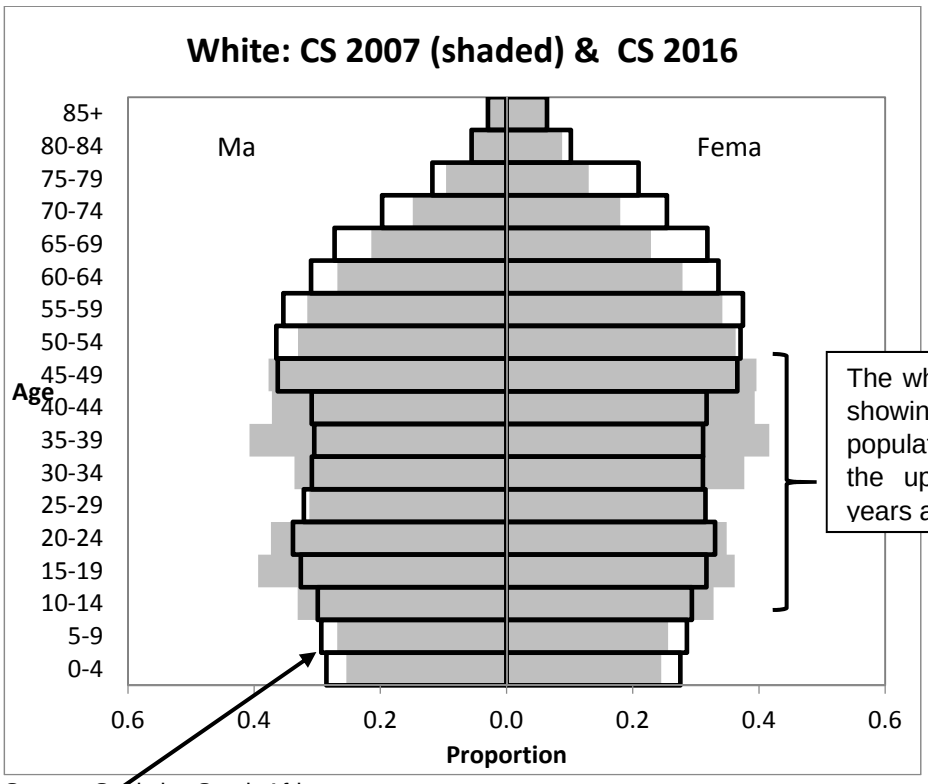


Figure 4.5: Indian/Asian population structure, CS 2007 and CS 2016



Source: Statistics South Africa

Figure 4.6: White population structure, CS 2007 and CS 2016



Source: Statistics South Africa

The white South African population is showing signs of a rapidly ageing population, as shown by this dent and the upward shift towards ages 50 years and older in the CS 2016 data.

CS 2016 shows a minor increase in the age 0–4 and 5–9 age groups. This may suggest increasing fertility rates among the whites.

4.2.1.2 Population growth between 2007 and 2016.

Table 4.1 presents the annual population growth rates by sex and total population between 2007 and 2016 as well as the sex ratios for both surveys. The results indicate that the population grew at approximately a rate of 1,5% annually since 2007. Overall, older age groups (35–59 and 60+) grew at a relatively higher rate (1,864 and 2,016 respectively) compared to younger age groups. Furthermore, older males (35–59 and 60+) grew at a much higher rate compared to females, at an average of 2,5% annually, on the other hand the female population aged 35–59 and 60+ grew at 1,4% and 1,7% respectively.

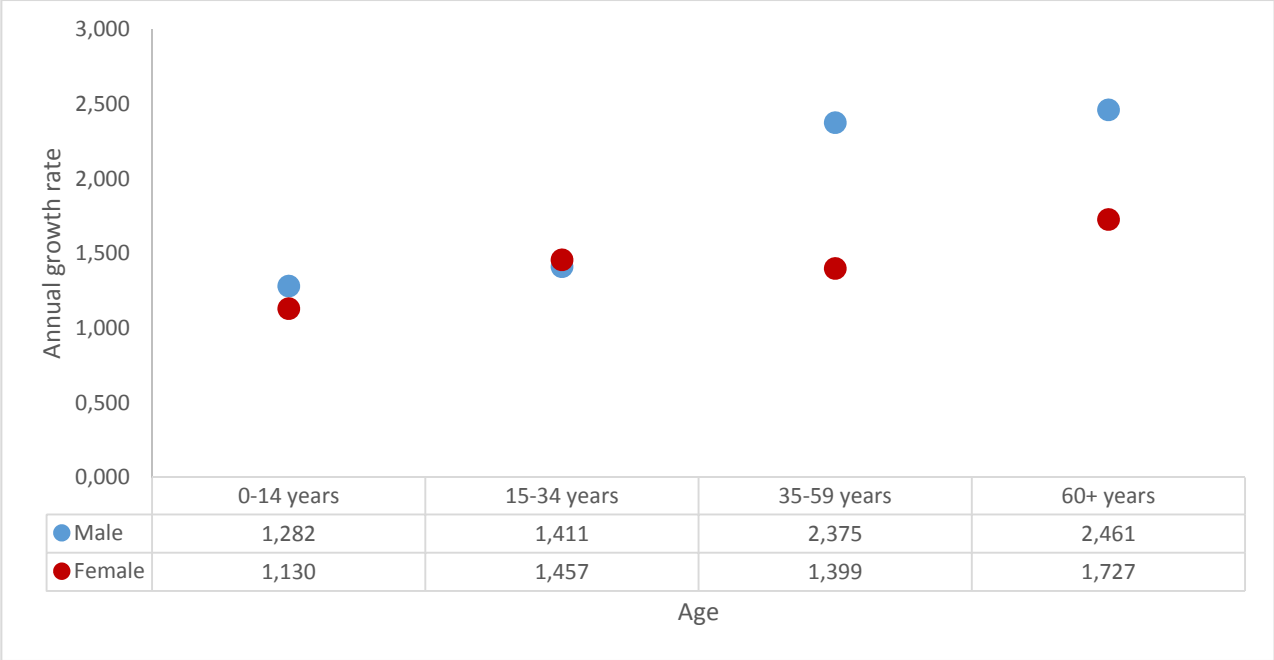
Table 4.1: Total population by sex, annual population growth rate by sex, and sex ratio, CS 2007 and CS 2016

	Male			Female			Total			Sex ratio	
Age	CS 2007	CS 2016	Growth rate	CS 2007	CS 2016	Growth rate	CS 2007	CS 2016	Growth rate	2007	2016
0-14	7 523 812	8 449 804	**1,282**	7 525 180	8 336 314	**1,130**	15 048 992	16 786 118	**1,206**	100,0	101,4
15-34	8 859 113	10 067 239	**1,411**	8 829 114	10 074 770	**1,457**	17 688 227	20 142 009	**1,434**	100,3	99,9
35-59	5 576 660	6 914 952	**2,375**	6 418 089	7 285 230	**1,399**	11 994 749	14 200 181	**1,864**	86,9	94,9
60+	1 452 479	1 815 231	**2,461**	2 317 616	2 710 114	**1,727**	3 770 095	4 525 346	**2,016**	62,7	67,0
Total	**23 412 064**	**27 247 226**	**1,675**	**25 089 999**	**28 406 428**	**1,371**	**48 502 063**	**55 653 654**	**1,519**	93,3	95,9

Source: Statistics South Africa

Moreover, as expected Table 4.1, shows a declining sex ratio as age increases for both the CS 2007 and CS 2016. However, the eight percentage point increase in the 35–59 years age group in the CS 2016 is a bit alarming. This increased male survival rate, however surprising is corroborated by the growth rates figures reported in Figure 4.7 which shows males in the ages 35 years and older growing at a much higher rate annually between 2007 and 2016 compared to their female counterparts.

Figure 4.7: Annual population growth rate by sex and broad age groups, CS 2007 and CS 2016

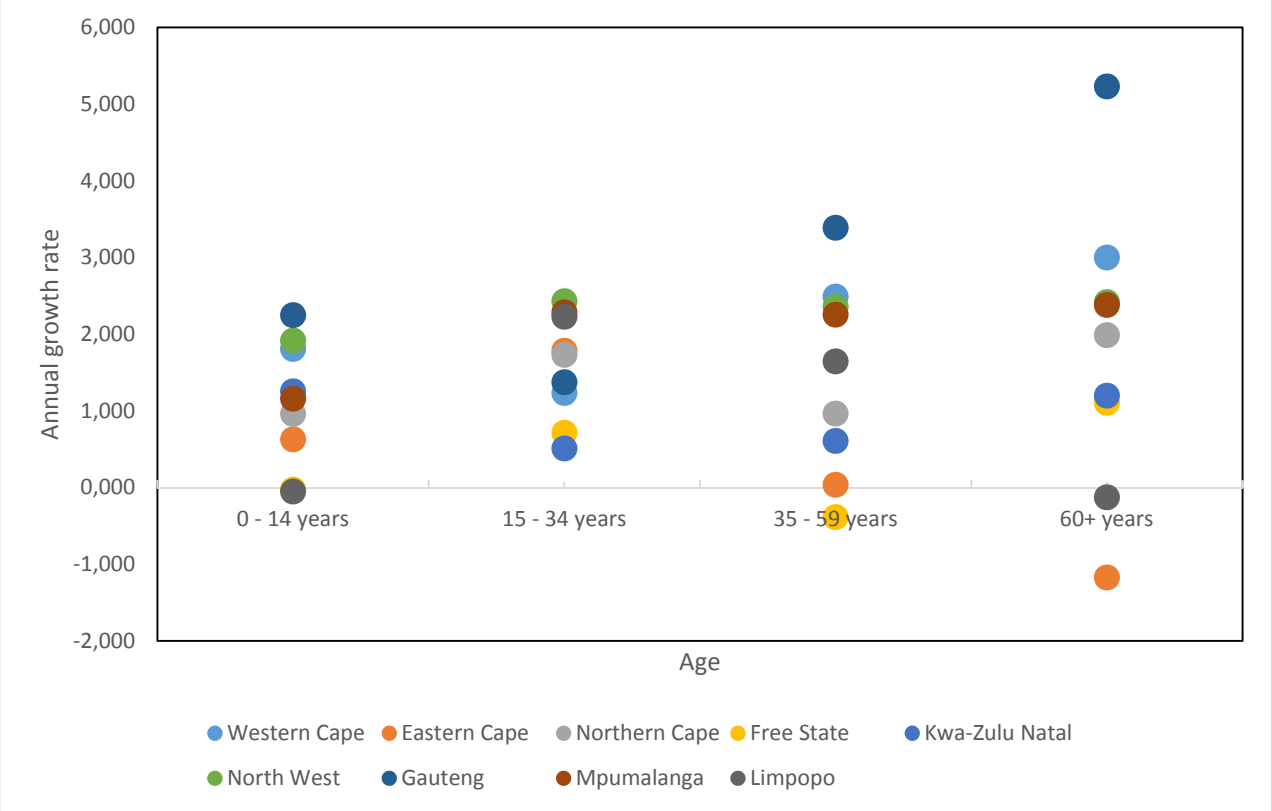


Source: Statistics South Africa

Figure 4.7 presents annual population growth by sex and broad age groups for the years between 2007 and 2016. The data shows consistent growth rates for males and females in the two younger age groups (0–14 years=1,2%; 15–34 years=1,4%). Nonetheless, the results further indicate that males in the older age (35–59 and 60+ years) groups grew at a much higher rate compared to females in the same ages.

Consistent with migration trends, Gauteng, the economic hub of the country shows the largest annual population growth rate for the 35–59 and 60+ years with 3,4% and 5,2%, respectively as shown in Figure 4.8. Eastern Cape and Limpopo on the other hand indicate a negative growth rate for the oldest age group. Furthermore, Free State showed a negative growth for the 0–14 and 35–39 years age group. Data further indicates that the youngest age cohort in Limpopo has been growing at a negative rate of -0,05% annually.

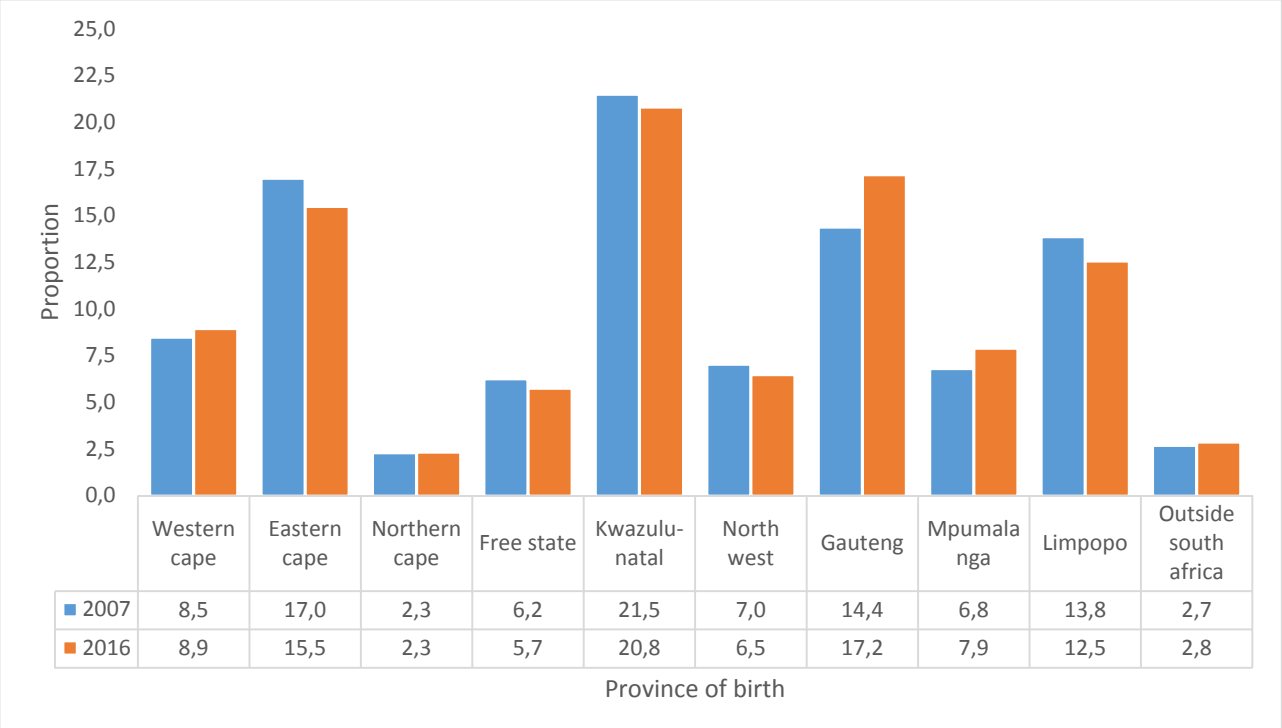
Figure 4.8: Annual population growth rate by broad age groups and province, CS 2007 and CS 2016



Source: Statistics South Africa

4.2.1.3 Province of birth

Figure 4.9: Proportion distribution of population by province of birth, CS 2007 and CS 2016



	Western cape	Eastern cape	Northern cape	Free state	Kwazulu-natal	North west	Gauteng	Mpumalanga	Limpopo	Outside south africa
2007	8,5	17,0	2,3	6,2	21,5	7,0	14,4	6,8	13,8	2,7
2016	8,9	15,5	2,3	5,7	20,8	6,5	17,2	7,9	12,5	2,8

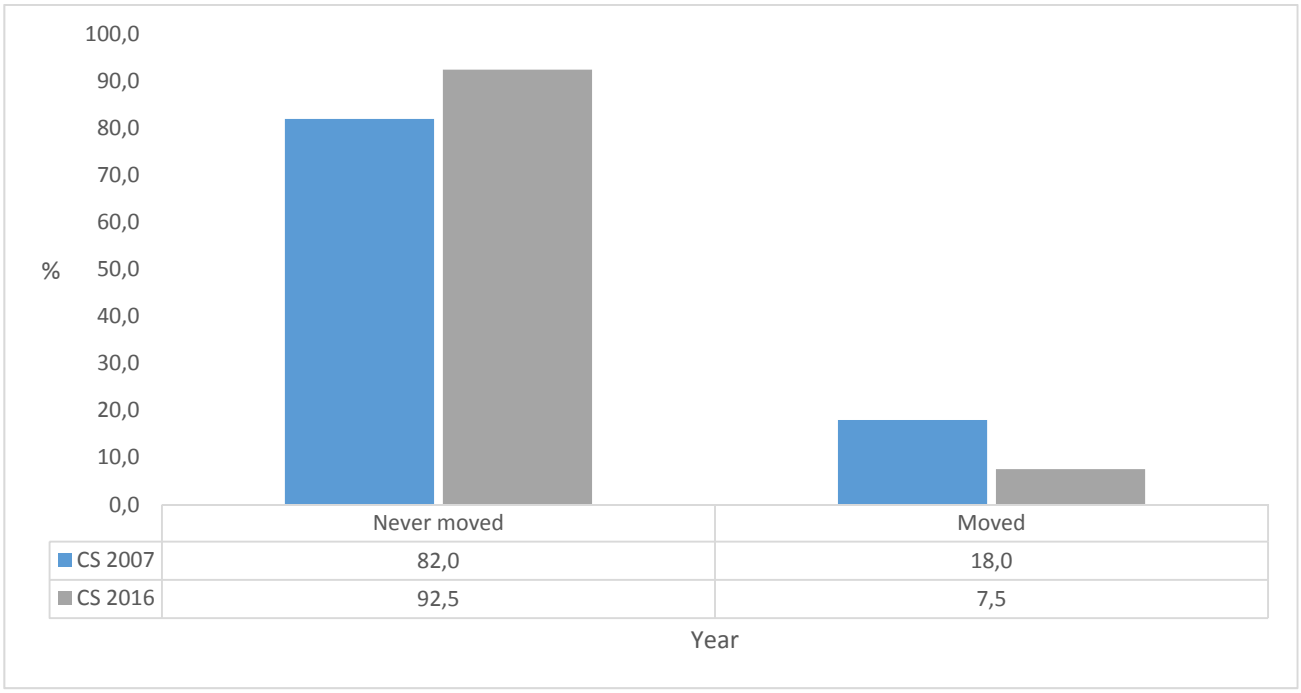
Source: Statistics South Africa

Migration trends confirm that people move from rural and least urbanised parts of the country to the big cities, especially those in Western Cape and Gauteng in search for better opportunities. Once in the receiving areas respondents tend to misreport information on province of birth, more often than not they opt to report that they were actually born at the enumeration area when that is not the case. Hence Figure 4.9 shows an increased number of respondents reported Gauteng, Western Cape and Mpumalanga as their provinces of birth in 2016 as compared to 2007, whilst fewer respondent reported Eastern Cape and Limpopo as their provinces of birth in the same period. Nonetheless, fewer documents exist on CS 2007 enumeration procedures to substantiate otherwise.

4.2.1.4 Internal movement five years ago

According to Figure 4.10, the proportion of the population reporting that they have not moved in the five years preceding the respective surveys increased by 10,5 percentage points, from 82% in 2007 to 92,5% in 2016. However, it must be noted that the sudden increase in reported movements in CS 2016 could be a result of the way in which the question was posed between the two surveys. In CS 2007 it asked "was (*the person*) living in this dwelling in October 2001?" whereas in CS 2016 it asked "has (*the person*) been staying in this place since October 2011?" These may have been understood differently by respondents.

Figure 4.10: Percentage of population by internal movement in the last five years, CS 2007 and CS 2016

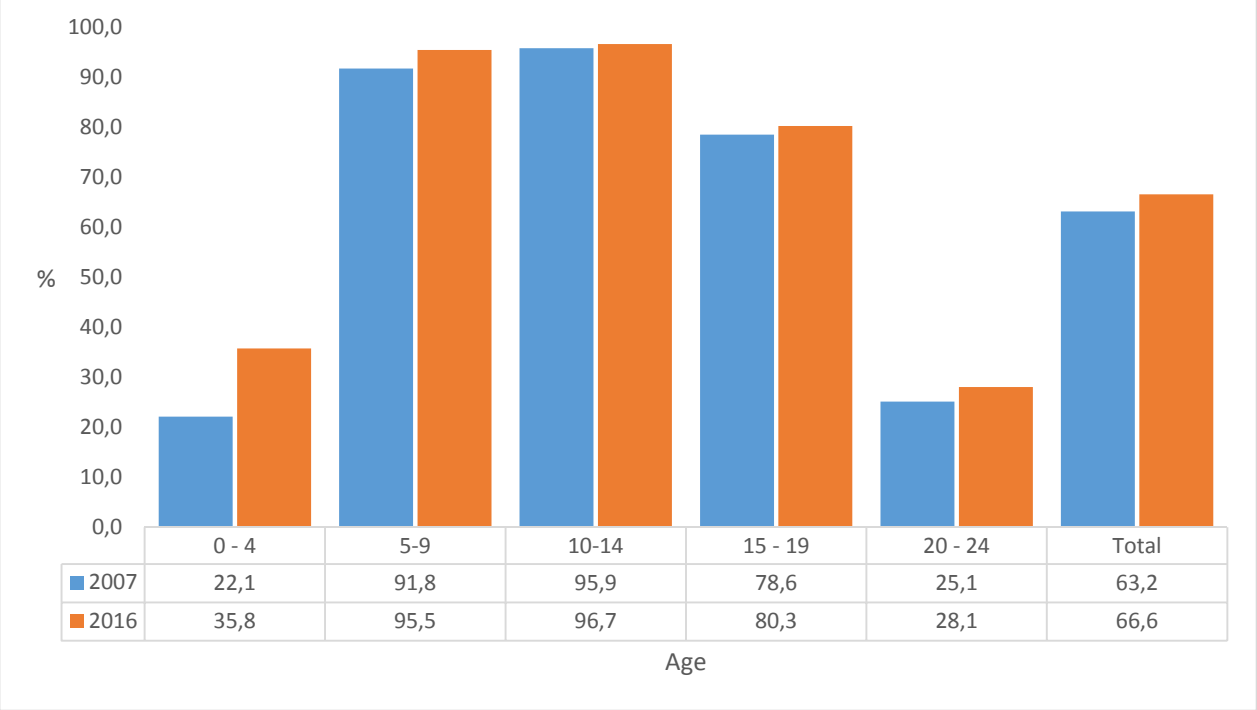


Source: Statistics South Africa

4.2.1.5 School attendance

Figure 4.11 presents percentage of persons aged 0–24 years by school attendance in 2007 and 2016. The 3,4 percentage points (63,2% to 66,6%) total increase in the overall (0–24 years) school attendance in 2016 compared to that of 2007 can be attributed to the school attendance increases in the 0–4 (22,1% to 35,8%), 5–9 (91,8% to 95,5%) and 20–24 (25,1% to 28,1%) age groups. Attendance in the 10–14 and 15–19 years ages remained relatively stable in the two surveys as shown in Figure 4.11.

Figure 4.11: Percentage of population aged 0-24 years by school attendance and age group, CS 2007 and CS 2016

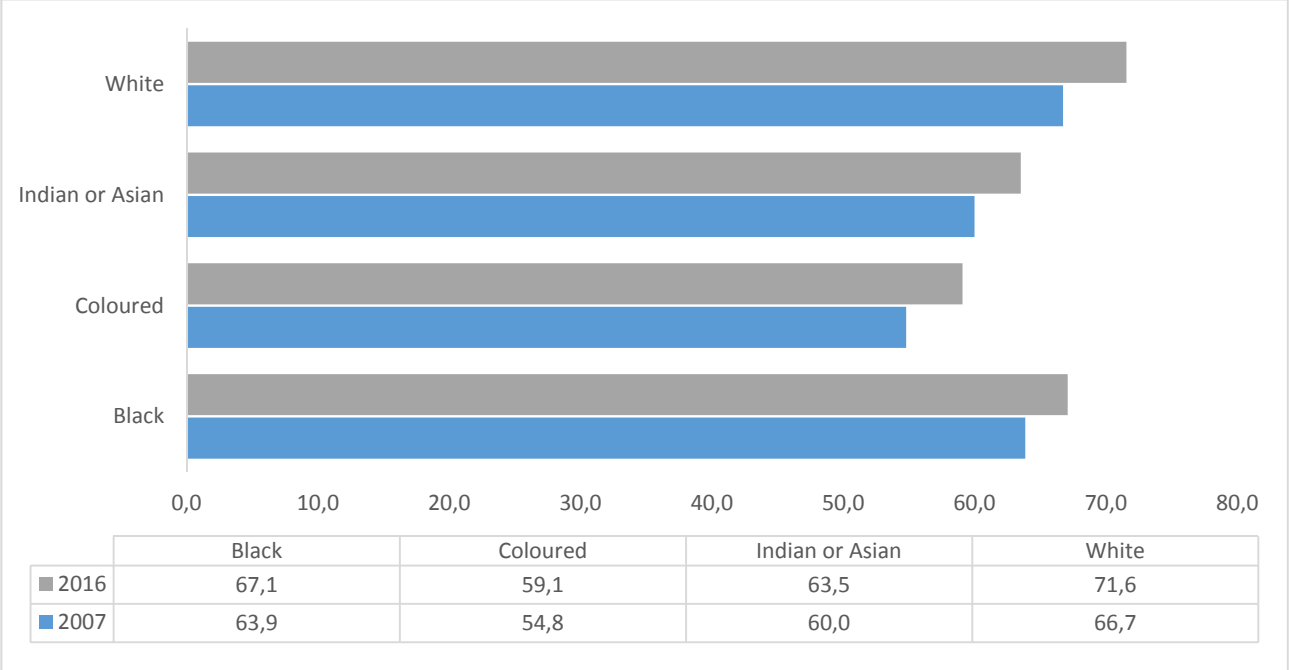


	0 - 4	5-9	10-14	15 - 19	20 - 24	Total
2007	22,1	91,8	95,9	78,6	25,1	63,2
2016	35,8	95,5	96,7	80,3	28,1	66,6

Source: Statistics South Africa

The increase in the 20–24 years age group can be attributed to the government's introduction of the Technical and Vocational Education and Training (TVET) (formerly Further Education and Training (FET)) colleges. This greatly increased access to tertiary education especially for those who did not meet the minimum requirements for acceptance at a university or university of technology. Furthermore, according to Figure 4.12 there has been a consistent increase in school attendance for the different population groups between 2007 and 2016. The white population showed the highest attendance rates while coloured population exhibited the lowest for both surveys.

Figure 4.12: Percentage of population aged 0–24 years by school attendance

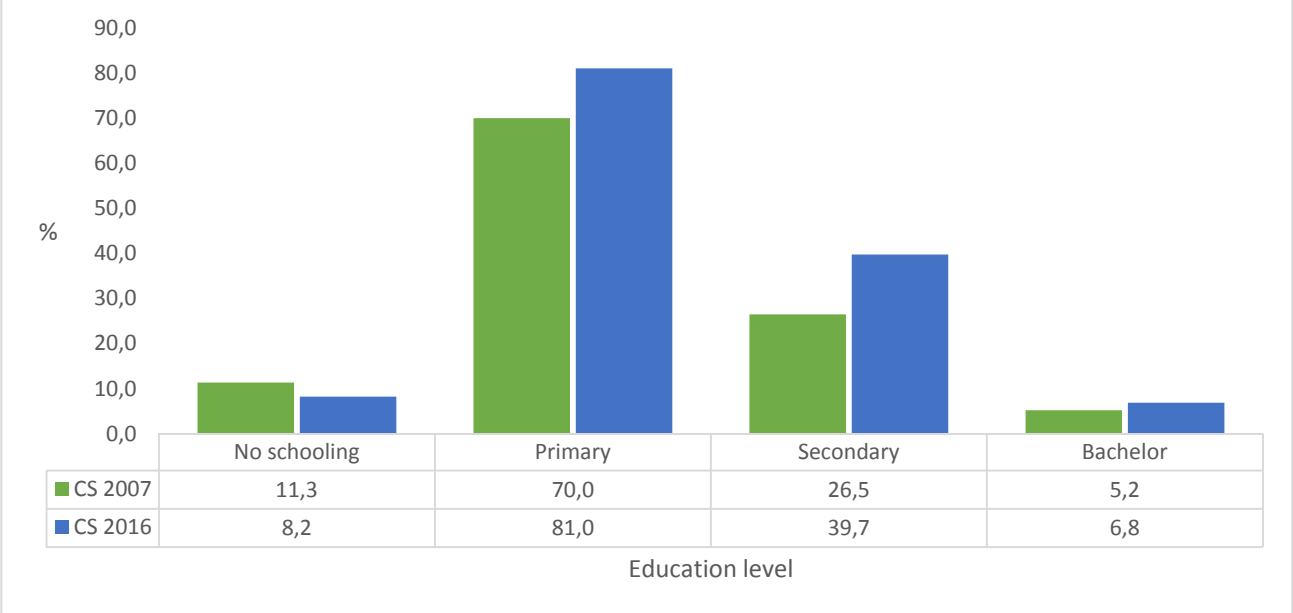


	Black	Coloured	Indian or Asian	White
2016	67,1	59,1	63,5	71,6
2007	63,9	54,8	60,0	66,7

Source: Statistics South Africa

4.2.1.6 Educational attainment

Figure 4.13: Percentage of persons aged 25 years and older by educational attainment



	No schooling	Primary	Secondary	Bachelor
CS 2007	11,3	70,0	26,5	5,2
CS 2016	8,2	81,0	39,7	6,8

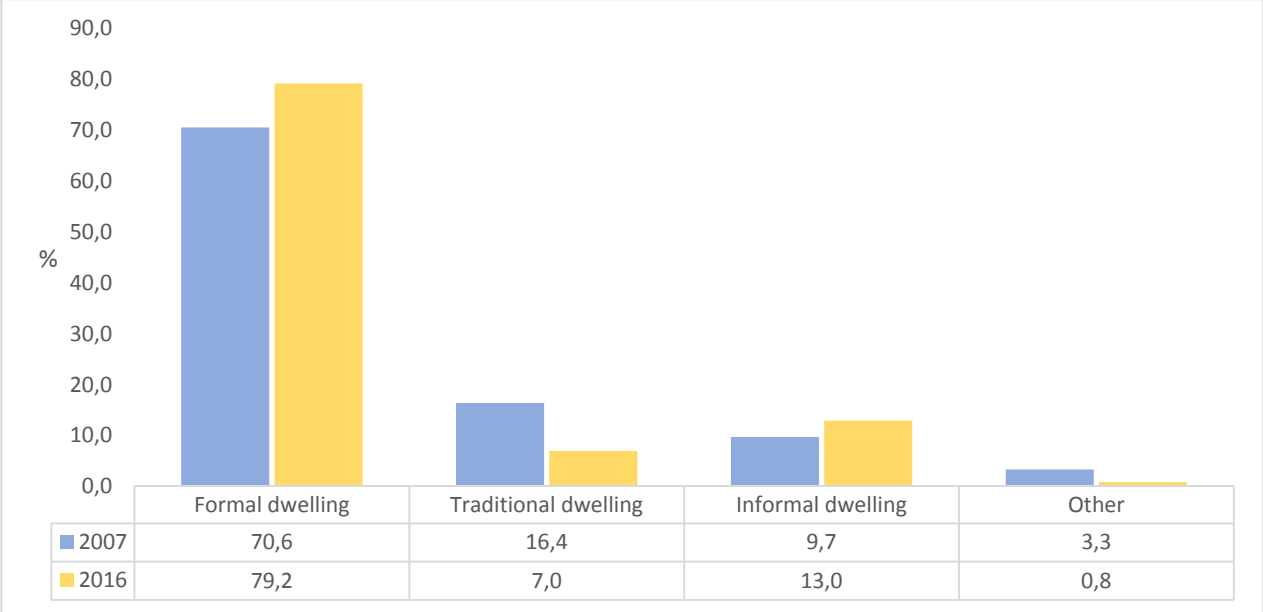
Source: Statistics South Africa

Whilst the percentage of persons aged 25 years and older who have no schooling decreased in 2016 relative to 2007, those who have attained primary, secondary and university education consistently increased. Those who reported having attained secondary education reflect the largest increase of about 13,2 percentage points.

4.2.2 HOUSEHOLD PROFILE

4.2.2.1 Type of main dwelling

Figure 4.14: Per cent of households by type of main dwelling, CS 2007 and 2016



	Formal dwelling	Traditional dwelling	Informal dwelling	Other
2007	70,6	16,4	9,7	3,3
2016	79,2	7,0	13,0	0,8

Source: Statistics South Africa

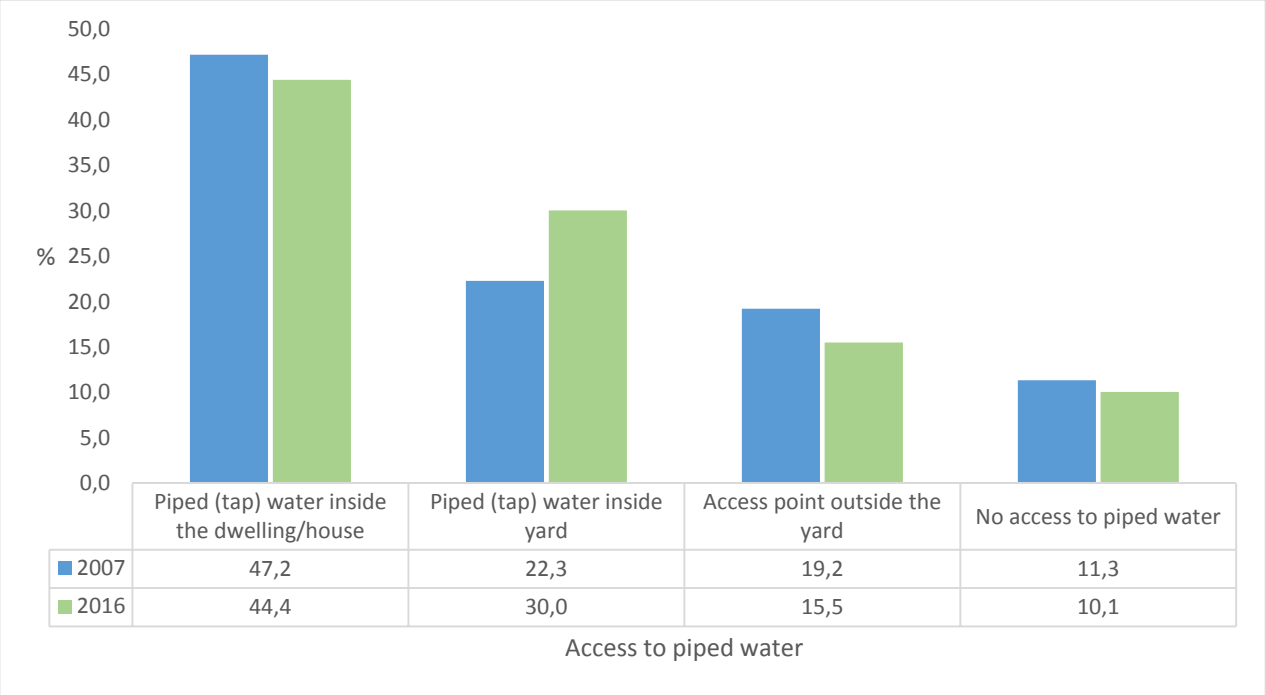
Figure 4.14 shows an increase in formal and informal dwellings in 2016 compared to 2007; while traditional dwellings decreased by 9,4 percentage points. This is consistent with urbanisation trends in South Africa. The increase in the informal dwellings could be explained by the movement of people from rural areas to bigger cities, which is complemented by an upsurge in the need for human settlement. In the absence of alternatives, people seek shelter in the form of informal dwellings.

4.2.2.2 Basic service delivery

CS 2016, data show a decrease in the proportion of households accessing piped water inside their dwellings/houses; this unexpected decrease can partly attributable to the way in which the question was posed in the two surveys (see Figures 4.16 and 4.17). Furthermore, South Africa has been experiencing one of the worst droughts in decades which led to water restrictions in some parts of the country, especially in the metropolitan areas. This may have adversely contributed to the responses given for this question in the CS 2016. Nevertheless, Figure 4.15 shows a minor decrease (1,2%) in the proportion of households that had no access to piped water in 2016 as

compared to 2007, a more accurate observation given the increase in the number of households and a subsequent delivery of basic services.

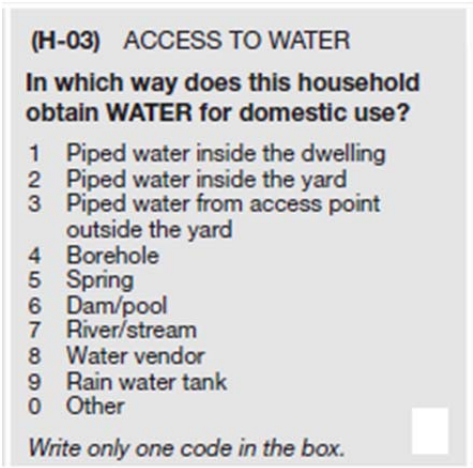
Figure 4.15: Percentage of households by access to piped water, CS 2007 and CS 2016



Source: Statistics South Africa

Note*: the question on access to water was asked differently in the two surveys. In 2007 the question asked for sources of water for domestic use, while in 2016 it asked for the main source of water for drinking.*

Figure 4.16: Snapshot of water source question, CS 2007



(H-03) ACCESS TO WATER

In which way does this household obtain WATER for domestic use?

1 Piped water inside the dwelling
2 Piped water inside the yard
3 Piped water from access point outside the yard
4 Borehole
5 Spring
6 Dam/pool
7 River/stream
8 Water vendor
9 Rain water tank
0 Other

Write only one code in the box.

Figure 4.17: Snapshot of water source question, CS 2016

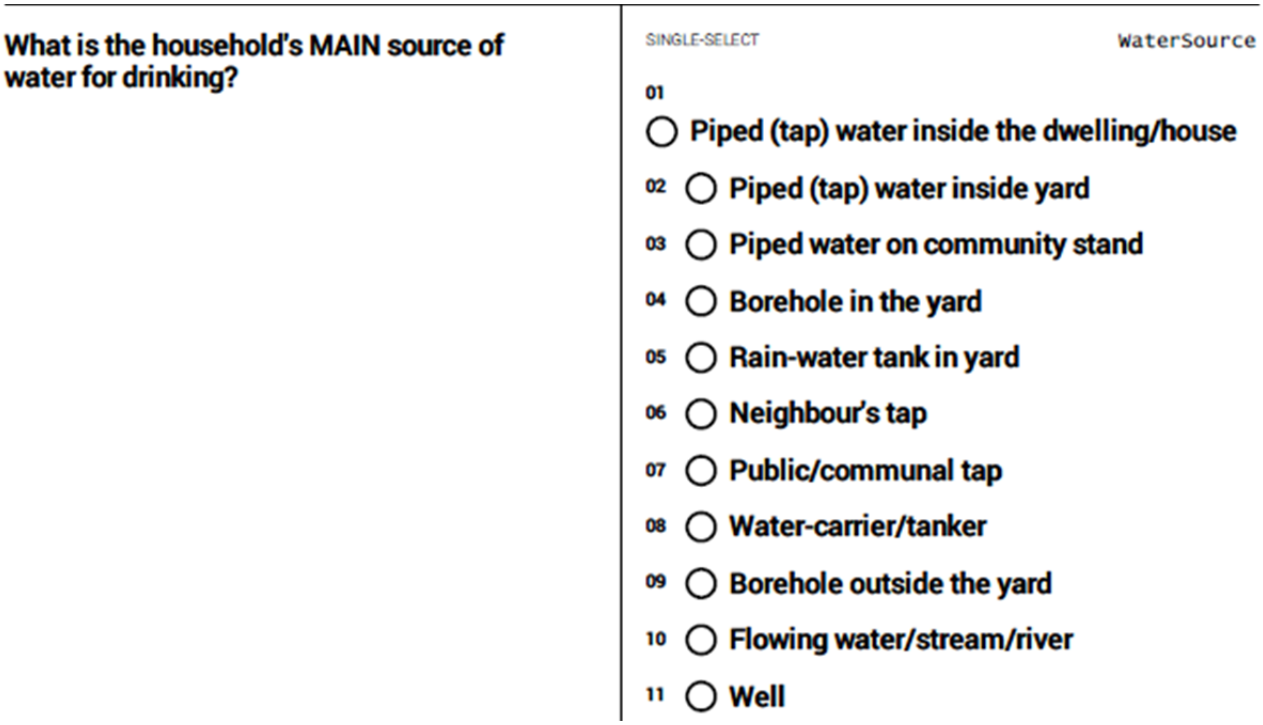
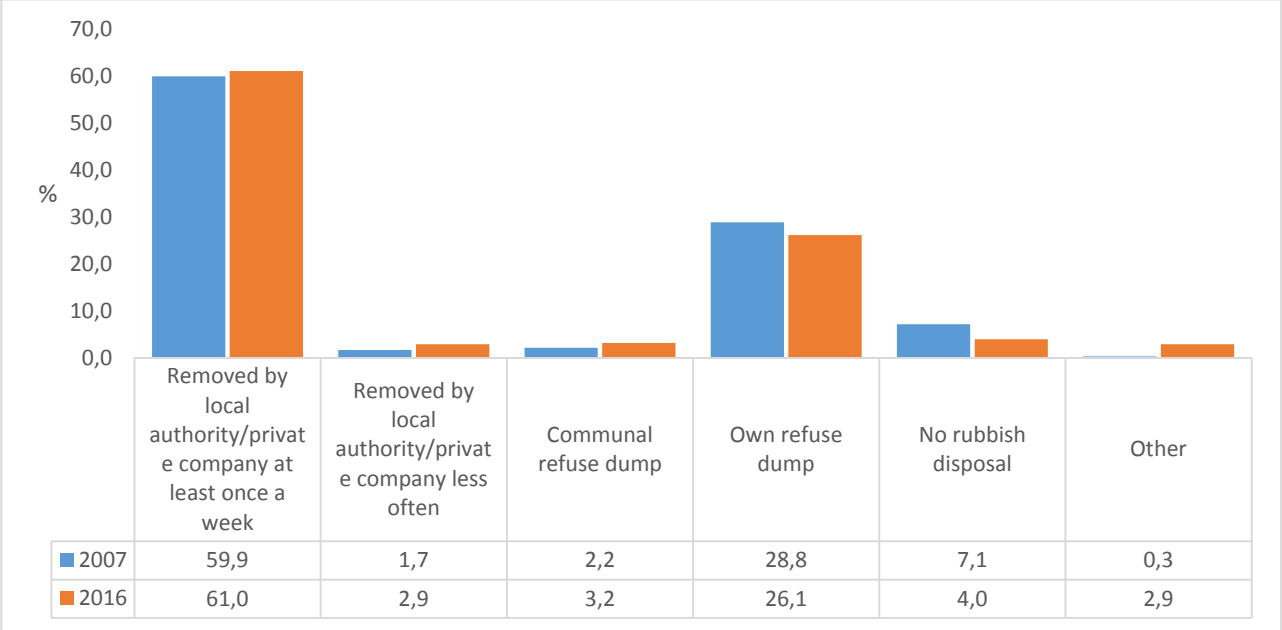


Figure 4.18 shows the percentage of households by type of refuse removal for the CS 2007 and CS 2016, the results indicate a consistent one percentage point increase in the proportion of households whose refuse was removed by local authorities/private company and those using communal refuse dumps. On the other hand, the proportion of households that used their own refuse dumps and those that had no means of disposing their refuse had to some extent decreased in 2016 as compared to 2007.

Figure 4.18: Percentage of households by type of refuse removal, CS 2007 and CS 2016

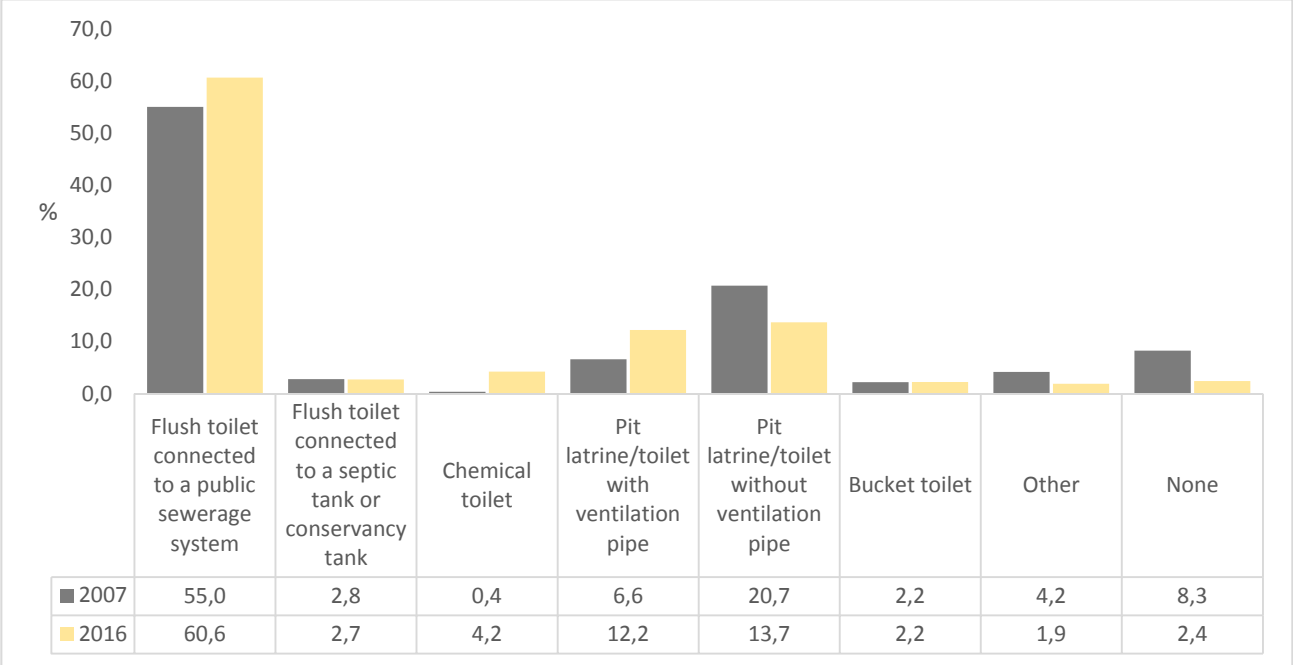


	Removed by local authority/private company at least once a week	Removed by local authority/private company less often	Communal refuse dump	Own refuse dump	No rubbish disposal	Other
2007	59,9	1,7	2,2	28,8	7,1	0,3
2016	61,0	2,9	3,2	26,1	4,0	2,9

Source: Statistics South Africa

According to Figure 4.19 the proportion of households that had access to flush toilets connected to a public sewerage system increased from 55% in 2007 to 60,6% in 2016, whilst those using pit latrine/toilets without ventilation pipes decreased by seven per cent in the same period from 20,7% to 13,7%. Furthermore, the usage of chemical toilets increased from less than one per cent in 2007 to 4,2% in 2016; and although not the ideal form of toilet, but nonetheless more dignified than the bucket toilet.

Figure 4.19: Percentage of households by type of toilet facility, CS 2007 and CS 2016



	Flush toilet connected to a public sewerage system	Flush toilet connected to a septic tank or conservancy tank	Chemical toilet	Pit latrine/toilet with ventilation pipe	Pit latrine/toilet without ventilation pipe	Bucket toilet	Other	None
2007	55,0	2,8	0,4	6,6	20,7	2,2	4,2	8,3
2016	60,6	2,7	4,2	12,2	13,7	2,2	1,9	2,4

Source: Statistics South Africa

Figure 4.20: Percentage of households by main source of energy for lighting, CS 2007 and CS 2016

	Electricity	Gas	Paraffin	Candles	Solar	Other	None	Unspecified
2007	80,1	0,2	5,3	13,7	0,2	0,5	0,0	0,0
2016	90,2	0,2	2,7	5,9	0,6	0,1	0,2	0,2

Source: Statistics South Africa

In addition, data presents a 10% increase in the proportion of households using electricity as their main source of energy for lighting in 2016 (90,2%) compared to 80,1% in 2007. This however, could be as a result of the way in which responses were categorised between CS 2007 and CS 2016 energy source questions. In the former, response categories did not specify different types of electricity while the latter did specify and even offered examples. Thus, households using a generator as their main source of energy for lighting in 2007, for example would have selected

"other" while the same household would have chosen "other sources of electricity (e.g. generator, etc.)". Figure 4.20 consequently indicates a decrease in the proportion of households using paraffin and candles for lighting in the CS 2016.as compared to the CS 2007.

4.3 Differences between CS 2007 and CS 2016

This subsection discusses some of the methodological and other differences and the impact they may have on the quality of data for CS 2007 and CS 2016 in no particular order.

4.3.1 CS 2007 and CS 2016 enumeration interval since last Census

According to the Stats Act (Act No. 6 of 1999) a population census must be conducted every five years in South Africa. However, as a result of fiscal constraints and the massive proportion of logistical mobilisation involved, censuses are conducted every ten years. Consequently, Community Surveys are supposed to be conducted in the five years mid-point between two censuses to bridge the information gap and update the primary data as a base for population projections. With regards to CS 2007 as well as CS 2016 the five-year mark was missed by a few months. Whilst the reference night for the two censuses was on the night of the 9th–10th October in the respective years, CS 2007 was on the night of 14th–15th February which was three and a half months past the stipulated five years. On the other hand, CS 2016 was conducted on the 6th–7th March, eight months short of five years.

One of the main reasons Censuses are conducted in the month October is because the South African population, a highly mobile one, has been found to be less mobile in October, thus easier and more practical to enumerate. In CS 2007; 238 067 out of 274 348 sampled DUs completed questionnaires whilst the response rate for CS 2016 was 90,5% which was well within the minimum acceptable threshold of 80% response rate at Stats SA.

4.3.2 Sampling procedures

Different sampling procedures were adopted for each of the two surveys. A two-stage stratified random sampling process was adopted for the CS 2007. Stage one involved the selection of the EAs, and stage two entailed the selection of the DUs. Since data was expected to be disseminated at local municipal level (category A and B), each municipality was considered as a stratum. The Census 2001 dwelling frame was used because it gave a full geographic coverage of the country without any overlap. Out of 80 787 EAs countrywide, 79 466 were considered in the frame. A total of 919 institutions and 402 recreational areas in the frame were excluded.

In the first stage, the EAs within each municipality were ordered by geographic type and EA type. The selection was done by using systematic random sampling. For the second level, a DU listing exercise was undertaken before the selection of the DUs. The listing exercise provided a completed list of DUs in the selected EAs. Only structures classified as DUs were considered for selection. A total of 2 511 314 DUs were listed from the exercise, 274 348 were sampled and of those 238 067 responded.

On the other hand, a single-stage sample design was adopted for the CS 2016. Owing to developments in technology, listing, as undertaken in 2007 was not done or needed for the CS 2016. Rather, a geo-referenced dwelling frame indicating a GPS location point spatially, where each point was assigned to a structure, stand or a yard depending on the settlement type was used. While the 2007 methodology included selection of EAs, in the CS 2016 all eligible Census 2011 EAs were included in the initial frame. EAs which did not have any DUs and those with a very small number of eligible DUs were excluded. This was done to ensure a representative sample. A total of 93 427 EAs and 1 370 809 DUs made up the final sample of the CS 2016.

4.3.2 Data items and questions asked in the Community Survey 2007 and 2016

Table 4.2 shows that questions on service delivery more than quadrupled between CS 2007 and CS 2016. This is attributed mainly to the changes in national agenda; focus on building infrastructure maintenance and quality of services rendered to the citizenry. Planners, including government require extensive indicators to assess the impact of programmes and interventions targeting addressing past imbalances. Also, newly-incorporated questions on emigration influenced the total number of CS 2016 data items. Data items increased from 78 in the CS 2007 to 134 in the CS 2016 as shown in Table 4.2. Given that the questionnaire was electronic with built-in skipping instructions, questionnaire length and interviewer fatigue were greatly minimised. Thus, the number of data items holds very little bearing when it comes to the impact on data quality. However, type of information collected increased. For the first time, in CS 2016 the questionnaire included questions on perceptions on service delivery.

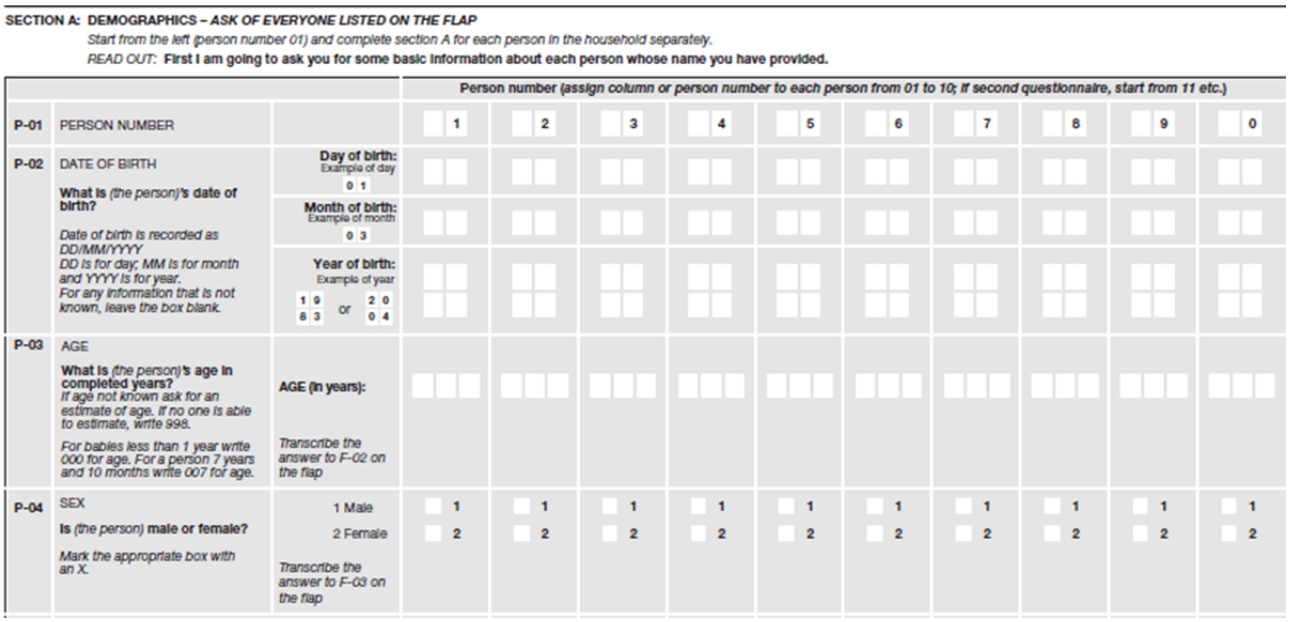
Table 4.2: Distribution of data items by section, CS 2007 and CS 2016

Section	CS 2007	CS 2016
Demographics	8	**9**
International Immigration and Internal Migration	12	**14**
General health & functioning	3	**2**
Parental Survival Status	4	**6**
Education	4	**7**
Employment	17	**18**
Income & Social Grants		**4**
Fertility & Breastfeeding Practices	8	**12**

Section	CS 2007	CS 2016
Households, Access & Quality of Basic Services, Perception on Crime & Experiences, Agriculture & Food Security	10	**46**
Emigration	0	**9**
Mortality	9	**7**
TOTAL	**78**	**134**

Source: Statistics South Africa

Figure 4.21: Snapshot of some demographics questions, CS 2007



SECTION A: DEMOGRAPHICS – *ASK OF EVERYONE LISTED ON THE FLAP*

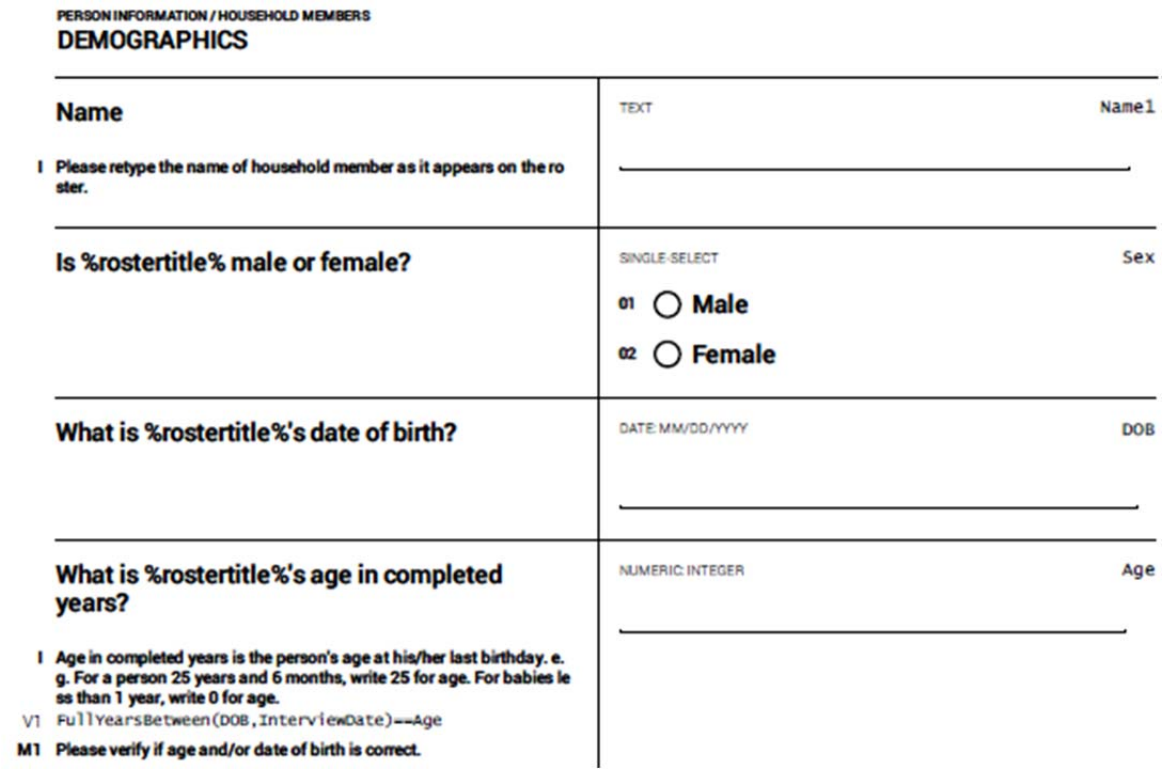
Start from the left (person number 01) and complete section A for each person in the household separately.

READ OUT: **First I am going to ask you for some basic information about each person whose name you have provided.**

			Person number (assign column or person number to each person from 01 to 10; if second questionnaire, start from 11 etc.)									
P-01	PERSON NUMBER		1	2	3	4	5	6	7	8	9	0
P-02	DATE OF BIRTH **What is *(the person)*'s date of birth?** *Date of birth is recorded as DD/MM/YYYY. DD is for day; MM is for month and YYYY is for year. For any information that is not known, leave the box blank.*	**Day of birth:** Example of day 0 1										
		Month of birth: Example of month 0 3										
		Year of birth: Example of year 1 9 8 3 or 2 0 0 4										
P-03	AGE **What is *(the person)*'s age in completed years?** *If age 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 10 months write 007 for age.*	**AGE (in years):** *Transcribe the answer to F-02 on the flap*										
P-04	SEX **Is *(the person)* male or female?** *Mark the appropriate box with an X.*	1 Male 2 Female *Transcribe the answer to F-03 on the flap*	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2	1 2

The demographics section of the population data set remains the back bone of such data. This is mainly because demographic analysis estimates are often considered a reliable standard for judging the accuracy of census/survey information collected at any time. Questions on population characteristics remained consistently comparable between CS 2007 and CS 2016 as shown on Figures 4.21 and 4.22. This ensured that a population characteristics trend analysis can be done between the surveys. Obviously, though, the move from CAPI to PAPI enhanced the accuracy of the CS 2016, given the in-built checks. This is reflected in the imputation rates, especially for the age variable, whereas in the CS 2016 less than one per cent (0,005%) of the age responses were imputed.

Figure 4.22: Snapshot of some demographics questions, CS 2016



PERSON INFORMATION / HOUSEHOLD MEMBERS

DEMOGRAPHICS

Name — TEXT — Name1

I Please retype the name of household member as it appears on the roster.

Is %rostertitle% male or female? — SINGLE-SELECT — Sex

01 ○ Male

02 ○ Female

What is %rostertitle%'s date of birth? — DATE: MM/DD/YYYY — DOB

What is %rostertitle%'s age in completed years? — NUMERIC: INTEGER — Age

I Age in completed years is the person's age at his/her last birthday. e.g. For a person 25 years and 6 months, write 25 for age. For babies less than 1 year, write 0 for age.

V1 FullYearsBetween(DOB,InterviewDate)==Age

M1 Please verify if age and/or date of birth is correct.

4.3.3 CAPI vs. PAPI

Evidently, the CS 2016 was the second large sample household survey undertaken after the CS 2007. In 2016, data were collected electronically using the paper-less CAPI system as opposed to the PAPI method used in CS 2007. Furthermore, other than the financial and time saving benefits associated with the usage of the CAPI system used in CS 2016, data quality enhancement was paramount. In CS 2016 high accuracy in collected data was warranted as a result of the built-in checks and the minimum processability rules imbedded into the questionnaire. This is confirmed by the minimal imputations (see Chapter 2) subjected to the data set, post collection. Moreover, time needed for data editing in CS 2016 was greatly reduced as a result compared to CS 2007. Furthermore, as a result of CAPI greater timeliness was obtained in CS 2016; as a result the results were released approximately two months after completion of collection whilst the CS 2007 results were released seven months after data collection.

4.4 Conclusion

Evidently, key indicators compare well between CS 2007 and CS 2016 as one may expect. Furthermore the usage of the CAPI method played a significant role in improving quality, especially the precision and accuracy of collected information in CS 2016.

SECTION 5: CONCLUSION

Survey data quality assessment is never a clear-cut process that requires one to decide on the method and level of analysis. For the purpose of this report a diversified level of analysis including calculating imputation rates and confidence interval, and comparing of output data for selected indicators for CS 2007 and CS 2016 was adopted. To deduce the true level of CS 2016 data quality, it was deemed necessary to compare it to another large sample survey rather than the recent 2011 Census. Although different methodologies were applied in the two surveys, especially the introduction of CAPI in 2016 and sampling method that ensured selection of DUs in all EAs in the dwelling frame, nevertheless, the two were similar in more ways than one on the other hand. Key variables remained comparable between the two surveys. Introduction of new questions as a result of increase in user needs for information saw data items quadruple in CS 2016 as compared to CS 2007; this however, had little or no effect on all on data quality.

Furthermore, the built-in checks in the electronic questionnaire used in CS 2016 warranted minimal imputations in the data. However, “Year of death of last child born”, “educational institution type” and “maintenance of toilet facility” variables had imputation rates higher than five per cent. A large majority of the variables had imputation rates statistically significant at one per cent, which is exceptional, far exceeding the 10% acceptable threshold recommended by the UN. Moreover, CS 2016 confidence intervals inferred a higher precision and accuracy of the data as compared to CS 2007 which showed wider interval ranges for most of the variables. Overall, we can safely conclude that the change in the sampling method and the adaption of the paper-less collection method has yielded data of higher quality, not only compared to CS 2007 but by any standard.

References

Statistics South Africa (2007), *Community Survey Statistical Release, 2007 (Revised version)*. Pretoria: South Africa

Statistics South Africa (2016), *Community Survey Technical Report, 2016*. Pretoria: South Africa