# THE 2022 SOUTH AFRICAN CENSUS

A TECHNICAL REPORT PREPARED FOR THE SOUTH AFRICAN MEDICAL RESEARCH COUNCIL

TOM MOULTRIE AND ROB DORRINGTON CENTRE FOR ACTUARIAL RESEARCH (CARe) UNIVERSITY OF CAPE TOWN JUNE 2024











#### ACKNOWLEDGMENTS

The authors thank the Burden of Disease Research Unit (BODRU) of the South African Medical Research Council for financial support in the production of this Technical Report, as well as helpful comments during its drafting. We also thank the reviewers for their time and effort in reviewing the Report.

#### SUGGESTED CITATION:

Moultrie, Tom A and Rob E Dorrington. 2024. The 2022 South African Census. Technical report prepared for the South African Medical Research Council Burden of Disease Research Unit. Cape Town: South African Medical Research Council.

**ISBN:** 978-1-928340-77-5

#### THE 2022 SOUTH AFRICAN CENSUS

A TECHNICAL REPORT PREPARED FOR THE SOUTH AFRICAN MEDICAL RESEARCH COUNCIL

Tom Moultrie and Rob Dorrington Centre for Actuarial Research (CARe) University of Cape Town June 2024

# FOREWORD

Accurate census data matter. Knowing the size, demographic and socio-economic characteristics, and spatial distribution of our population is critical for both the planning and delivery of services, as well as ensuring the equitable allocation of scarce resources. A census is fundamental to all facets of governance, not least health and health related research.

Apart from demographic measures of mortality and fertility, particularly for smaller groups, reliable data on the size of the population guide our assessment of health needs and health related risk. The data provide denominators that allow the calculation of incidence and prevalence of rates of diseases, and death rates. In addition to providing the denominators, the census also provides sampling frames for researchers to conduct household surveys. These activities contribute not only to identifying inequities but also the research agenda to address them.

This Technical Report focuses on the recently released results of the 2022 South African Census, the fourth census to be conducted in post-Apartheid South Africa. It has been prepared by eminent demographers from the Centre of Actuarial Research (CARe) at the University of Cape Town (UCT) who have collaborated with researchers in the SAMRC Burden of Disease Research Unit for many years. This collaboration quantified the emerging mortality burden from HIV/AIDS at the start of this century. More recently, the collaboration set up a system to track number of weekly deaths in South Africa in near-to-real time, an invaluable tool during the Covid pandemic. Regrettably, the careful evaluation of the census results shows how the COVID-19 pandemic and the consequences thereof greatly impacted on the planning, operations, and logistics of an exercise as massive as the census.

The reported scale of the undercount in the 2022 South African Census not only compromises the census results at small-areas and among smaller populations, but requires urgent and deep reflection and analysis to ensure that the problems and concerns identified here do not re-occur. Furthermore, the indications that the population of 62.03 million might be overstated



by approximately one million people suggests that concerted efforts are needed to produce alternative estimates of the population as of early 2022 that can be used with greater confidence as a base from which to project estimates of the population in future years and in particular for the estimation of the burden of disease, and by researchers seeking to better understand the characteristics of the society we live in.

Not only should the census provide a snapshot of who we are as a nation; and assist in helping us appreciate the historical obstacles we have overcome, it should also aid in the identification of the challenges that still lie ahead of us. It is of great concern that this report reveals such substantive issues, especially given that we will have to wait until 2031 for the next census.



Professor Glenda Gray

President and CEO: South African Medical Research Council

## **EXECUTIVE SUMMARY**

The results of the 2022 South African census were released on 10 October 2023. This Technical Report investigates the reported count of the population inter alia by age, sex, and population group using data released by Statistics South Africa (Stats SA) to date, to inform the estimation of population numbers in future for policy and planning purposes as well as the determination of denominators that will inform estimates from the National Burden of Disease study.

Although logistical arrangements for the 2022 South African census were impacted severely by the repeated waves of SARS-CoV-2 infections, the National Treasury – unfortunately – declined a request to reallocate the census budget from the 2021/22 financial year to a subsequent financial year. This and other complications probably contributed to certain problems identified with the estimates produced from the census.

Important for policy and planning purposes, these are:

- Indications that the population of 62.03 million revealed by the 2022 Census might be overstated by approximately one million people.
- An underestimate and implausible age profile of the numbers of migrants implied by the questions on migration in the census.
- Incoherent and implausible results at a sub-national level, with the incoherence escalating at increasingly fine levels of spatial disaggregation.
- Serious concerns about the scope and scale of adjustments made to the census results in the light of the reported undercount of 31 per cent.

More specifically, the 2022 South African census appears to have:

• Overestimated the population aged 50 and older, and underestimated the population aged between 5 and 9 years of age (in particular those aged 5 last birthday).

- Estimates of the Indian/Asian and white populations that are too high.
- Estimates of the total population by sex that overestimate the number of men relative to women over age 60 (given that migration is low at these ages) and possibly underestimate men relative to women aged 25-39 (since one would not expect more female migrants than male migrants at these ages).
- A number of demographic inconsistencies in the reported age-sex structures of provincial populations.
- Inconsistencies between district populations estimated by the cand those of either the 2011 census or those projected by Stats SA.

Much of the data from the 2022 census, which would aid in understanding the results, is yet to be released. This includes the data on fertility, mortality, and more detailed data on migration based on place of residence at the previous census. These data should be released without delay.

Uncritical use of the 2022 census results for policy and planning purposes would be unwise, pending a full review of the 2022 census results. We are working to produce an alternative set of population estimates at the census date that will resolve – as far as possible – the inconsistencies identified.

Finally, it would be inadvisable for population projections (including the official mid-year population estimates (MYPE) to base projections on the 2022 census results without adjustment for the problems identified in this report.

## TABLE OF CONTENTS

ΕX	ECUTIVE SUMMARY	3
TA	BLE OF CONTENTS	4
LIS	T OF TABLES	5
LIS	T OF FIGURES	6
1.	INTRODUCTION	7
	1.1. Previous South African census undertakings	7
	1.2. Census 2022 timelines and preparation	8
	1.3. Data availability at the time of the official release of the Census 2022 results	8
2.	THE RESULTS FROM THE 2022 SOUTH AFRICAN CENSUS	9
	2.1. Introduction	9
	2.2. Balance equation analysis	9
	2.3. Comparison with previous censuses	13
	2.4. Concerns with district and municipal estimates	22
3.	CONCERNS ABOUT THE 2022 POST-ENUMERATION SURVEY	29
	3.1. The Post-Enumeration Survey (PES)	29
	3.2. The 2022 South African Census PES	29
	3.3. District-level concerns and conclusion	32
4.	WHAT HAPPENED?	33
5.	CONCLUSIONS	35
	5.1. Are the data fit for purpose?	35
	5.2. What next?	35
RE	FERENCES	37
AP	PENDIX 1: TIMELINE OF EVENTS LEADING TO THE RELEASE OF THE 2022 CENSUS	39
AP	PENDIX 2: FURTHER ANALYSIS OF THE FOREIGN-BORN POPULATION IN THE 2011 AND 2022 CENSUSES	41
AP TH	PENDIX 3: BACKGROUND TO POST-ENUMERATION SURVEYS (PESS), AND DETAILED CONCERNS REGARDING E PES IN THE 2022 SOUTH AFRICAN CENSUS	45
	A3.1 How to estimate who was missed in a census enumeration?	45
	A3.2 Specific concerns about the 2022 South African PES	46
	A3.3 Matching agreements	49
AP TC	PENDIX 4: DETAILED DESCRIPTION OF EVENTS AND PROCESSES THAT MAY HAVE CONTRIBUTED THE PROBLEMS WITH THE 2022 CENSUS	51
	A4.1 A major exogenous event: Covid-19	51
	A4.2 Endogenous events	52

# LIST OF TABLES

Table 1: Reconstruction of the South African population and       international migration using the balance equation, national	10	Table 9: Estimates of the total population, their standard       errors (SE), South Africa and its provinces, 2011 and 2022	31
Table 2: Reconstruction of the South African population and international migration using the balance equation, provincial	11	Table 10:     Timeline of important events in the preparation       and conduct of Census 2022	39
Table 3: Reconstruction of the South African population       and international migration using the balance equation,       national by sex	12	Table 11: Foreign-born population by sex and population       group, 2011 and 2022 censuses	42
Table 4: Reconstruction of the South African populationand international migration using the balance equation,national by population group	12	Table 12: Structure of a contingency table used to estimate       total population size in a capture-recapture exercise	45
Table 5: Intercensal growth rates     2011-2022, by district	22	Table 13: Estimates of the undercount (UC), their standard errors (SE), South Africa and its provinces as estimated directly from the post-enumeration survey23, 2011 and 2022	47
Table 6: District populations as shown by the mid-year       population estimates (interpolated to February 2022)       and the 2022 Census	26	<b>Table 14:</b> Estimates of the preliminary and final undercount (UC), their standard errors (SE), and coefficients of variation (CV), South Africa and its provinces, 2022 Census	47
Table 7: Populations aged 18 and over, numbers of registered voters, and selected ratios in 5 illustrative districts, 2011 and 2021/2	28	Table 15: Estimates of the total population, their standard       errors (SE), and coefficients of variation (CV), South Africa       and its provinces, 2011 and 2022	48
Table 8: Estimates of the undercount (UC) and their standard errors (SE), South Africa and its provinces as estimated directly from the post-enumeration survey18, 2011 and 2022	31	<b>Table 16:</b> Measures of matching from the PES in 2011 and 2021 by selected variables	50

# LIST OF FIGURES

Figure 1: Census 2022 numbers by age compared to numbers projected from past censuses and models, National	14	Figure 8: Population estimates of the Central Karoo District and City of Johannesburg, 2011 and 2022 censuses, by age group	24
Figure 2: Census 2022 estimated numbers by single years of age, National	15	Figure 9: Age structure of the population of WC053: Beaufort West Local Municipality, 2011 and 2022 censuses	25
Figure 3: Census 2022 ratio of numbers of males to females by age compared to the ratios projected from past censuses and models, National	16	Figure 10: Satellite image from Google Earth of Beaufort West town in January 2022, with areas developed since January 2010 outlined in red	26
Figure 4: Census 2022 estimated numbers by age compared to numbers projected from past censuses and models, Provinces (in thousands)	17	Figure 11: Numbers of foreign-born people reported in the 2011 and 2022 South African censuses, by age group and sex	41
Figure 5: Census 2022 ratio of numbers of males to females by age compared to the ratios projected from past censuses and models, Provinces	19	Figure 12: Age profiles of the foreign-born population by sex and population group, 2011 and 2022 South African censuses	43
Figure 6: Census 2022 numbers by age compared to numbers projected from past censuses and models, Population Groups	20	Figure 13: Ratio of coefficients of variation of population estimates (y-axis) plotted against the ratio of net undercount coefficients of variation (x-axis), 2011 and 2022 censuses	49
Figure 7: Census 2022 ratio of numbers of males to females by age compared to the ratios projected from past censuses and models, Population Groups	21	Figure 14: Weekly proportions testing positive for SARS-CoV-2 in South Africa, 2020-2023	52

### 1. INTRODUCTION

Even under near-ideal conditions, running a national census is a costly<sup>1</sup>, complex, and logistically-demanding undertaking. The US National Academies, in their assessment of the 2000 United States Census – entitled "The 2000 Census: Counting Under Adversity" – opens their executive summary of the volume as follows:

"The decennial census is the federal government's largest and most complex peacetime operation. Mandated in the U.S. Constitution to provide the basis for reapportioning seats in the U.S. House of Representatives, the census has many purposes today: redrawing congressional and state legislative district boundaries; allocating federal and state program funds; planning and evaluating federal, state, and local government programs; providing the basis for updated population estimates; and informing researchers, the private sector, the media, and the public about the characteristics of population groups and geographic localities and how they have changed over time." (National Research Council 2004: 1)

And, while many developed countries have replaced, or are in the process of replacing, their national censuses with snapshots of their population through linking of a variety of essentially-complete administrative databases and official registers<sup>2</sup>, the generally poor quality of administrative data in most low- and lower-middle income (and some upper-middle income) countries means that the census takes on even greater importance in these settings. Given the complexity described by the National Academies, we can see that there is a cruel and bitter irony associated with running censuses across the developing world: the countries that most desperately and urgently require reliable census data for policy and planning purposes are likely to be those least capable of mounting the complex logistical operations required for them to be successful.

# 1.1. Previous South African census undertakings

Previous South African censuses have been bedevilled by both common and unique issues. Prior to the end of apartheid, only two South African censuses, those conducted in 1936 and 1970, were regarded as having offered reasonably accurate enumerations of the entire population. In the quarter century from 1970 to 1996, the granting of 'independence' to 'homelands', the devolution of census operations to those authorities, and the increased illegitimacy of government and associated low-intensity unrest that prevailed almost unabated in South Africa from the mid-1980s to the early 1990s all served to render unreliable the censuses conducted in that period – in 1980, 1985, and 1991.

Censuses in the post-apartheid era were conducted in 1996, 2001, and 2011. While – for the first time since 1970 – the censuses sought to enumerate all South Africans, each of these censuses were characterised by challenges in enumerating the population. Post-Enumeration Surveys, a standard tool recommended by the United Nations Statistics Division (UN Statistics Division 2010) to estimate the extent of (and, often, to adjust for) a census undercount revealed undercounts of approximately 10, 17, and 14 per cent respectively.<sup>3</sup>

The reports from these censuses revealed particular difficulties in enumerating different sub-populations: people living 'behind high-walls' (who were reticent to admit census fieldworkers onto their premises); those living in 'informal' settlements (where mapping and household listings were often inadequate); and – more recently – international migrants (who might fear the data being used to identify and deport them, coupled with a growing fear of xenophobic persecution).

In comparison with undercounts estimated in other countries, these estimated undercounts were high;

<sup>&</sup>lt;sup>1</sup> The South African 2022 census is estimated to have cost around R3.2 billion (Parliamentary Monitoring Group 2023).

<sup>&</sup>lt;sup>2</sup> The United States is an important exception to this; Article 1 of the US Constitution requires that a census be conducted in order to allocate representation in the House of Representatives.

<sup>&</sup>lt;sup>3</sup> Comparison with subsequent censuses suggest the undercount was closer to 14%.

although it should be noted that many countries often either do not conduct, or do not put the results of their Post Enumeration Surveys in the public domain, even if one is conducted.

The undercount of 17 per cent in 2001 was a major contributor to the decision to shift the periodicity of South African censuses from a 5-year to a 10-year cycle, it being argued that frequent attempts at enumeration in conjunction with the long preparation and lead-times required to run a successful census was imposing an operationally-damaging burden on the government department entrusted with conducting the census, Statistics South Africa ('Stats SA').

As a result, no census was conducted in 2006, being replaced by a large-scale (an approximately 2 per cent sample of the total population) Community Survey conducted in 2007, approximately half-way between the 2001 and 2011 censuses, and again repeated in 2016.

# 1.2. Census 2022 timelines and preparation

A subsequent section of this report, detailing the logistical and other problems encountered in the census, delves into the preparation for Census 2022 in greater detail. For context and the record, a simple timeline of the major events leading to the start of the Census 2022 data collection exercise and culminating in the release of the census results in October 2023, reconstructed from Stats SA official press and media releases, appears in Appendix 1.

The difficulties experienced with implementation are apparent from the fact that while fieldwork was initially expected to run from 3-28 February 2022, the data collection period was extended by a further three weeks (to 20 March 2022), and then further extended for online enumeration nationally to 30 April 2022, and in the Western Cape for a further month, to 31 May 2022. Nationally, then, data collection took 87 days, rather than the anticipated 26 days. In the Western Cape, data collection took 118 days.

This, in addition, meant that rather than the Post-Enumeration Survey (PES) taking place as close in time as possible to the start of census (to allow close comparison of the results of the two), it took place some four months afterwards. To compound matters, the PES itself took some time to complete, further complicating comparison of the PES results to those from the census.

# 1.3. Data availability at the time of the official release of the Census 2022 results

The results of the 2022 census were released on 10 October 2023, some 20 months after the start of data collection. By comparison, the results of the 2011 census were released 12½ months after the start of data collection in that census.

However, in stark contrast with the release of the Census 2011 results, the quantity of data made publicly available at the official release were very scant. Only simple tabulations can be extracted from the published volumes (Statistics South Africa 2023c, e, f). One publication, South Africa at a Glance (Statistics South Africa 2023g), was released and almost immediately retracted, and has not been seen since.

Limited choice of tabulations (by 5-year age group and sex, and municipality) were available online at the time of the official release, but these could not be used to delve more deeply into the data. In addition, several issues were noted in those tabulations, with the data in the open age interval, 85 and over, including those for whom no age was determined; and numerous issues associated with recasting the 2011 data to match the 2022 municipal boundaries have been identified. These errors made precise analysis of the 2022 census results practically impossible, and it was only with the release of data on SuperWeb (Stats SA's online data portal) in early December 2022 that it was possible to construct multidimensional tables from the (still limited) data fields now made available<sup>4</sup>.

Two years after the census data were collected, much of the data collected has yet to be released. Again, this is a great pity – the longer the delay in making the data available, the harder it is to build an accurate snapshot of the population at the census date that is still relevant to current policy and other needs.

Finally, there is still no indication as to when the data of crucial demographic components (fertility, mortality, or the more detailed data on migration) will be made available, or indeed when a 10% public-use microsample of the data will be released.

<sup>&</sup>lt;sup>4</sup> Since the uploading of the 2022 census data, the SuperWeb portal (available at http://superweb.statssa.gov.za/) has been unstable, and often unavailable, for extended periods of time.

#### 2.1. Introduction

The census results released on 10 October 2023 headlined a population of 62.03 million at the census date in February 2022. The release of the report on the Post-Enumeration Survey (Statistics South Africa 2023e) on the same date estimated the undercount to be 31.06%, more than double that observed in 2011.

The reported estimated (after adjustment for the undercount) size of the population in South Africa produced by the 2022 census was quite a surprise to demographers, both in South Africa and internationally, particularly taking into account the assumed impact of the COVID epidemic on both mortality and migration. Prior to the release of the census, the estimates of the size of the South African population in 2022, based on population projections from a variety of credible sources ranged from the Institute for Health Metrics and Evaluation's 57.2 million to Stats SA's 2022 mid-year population estimate (MYPE) of 60.3 million (GBD 2021 Demographics Collaborators 2024; Statistics South Africa 2022g).

Figures 6 and 9 of the Post-Enumeration Survey (PES) Report (Statistics South Africa 2023e) indicate that the total of 62.03 million at the census data comprises an estimated true size of the population in households of 61.37 million, with a further 0.66 million people in "other collective living quarters" – hostels, prisons, university residence, and the homeless. The latter is surprising since the 2011 Census counted 1.79 million people (of a total of 51.77 million) living in "other collective living quarters"(Statistics South Africa 2012).

Thus, the overall Census 2022 population number is derived from an enumerated population of 36.59 million people (35.93 in households and 0.66 million in "other collective living quarters"), to which is added 16.78 million people identified from the PES but not recorded in the census enumeration and an estimated 8.66 million people who were estimated to be missed in both the PES and the census. **Given the magnitude of the undercount, it is more accurate to describe the results of the 2022 South African census as "estimates" rather than as "counts" of the population.**  While Statistics South Africa is to be commended for adhering to international best practice in conducting, and reporting the results of, a post-enumeration survey, the extent of the undercount of the census – in excess of 30% – is extreme. Among countries and censuses that follow this best practice, the undercount is the highest ever recorded (by an absolute margin of some 10%) among over 300 censuses with official PES-determined undercounts since the 1940s reported to the United Nations Population Division (United Nations Population Division 2023). We discuss our concerns with the PES in greater detail in Section 3.

#### 2.2. Balance equation analysis

One of the most basic tools for evaluating a census or population projection is the "balance equation". As far as censuses go, the balance equation asserts the logic that if all numbers are correct, the numbers in the current census (nationally, by age, by sex, by province, by population group, etc.) should be equal to those of the previous census, plus the births, less the deaths, plus the net in-migration (i.e., in-migration less out-migration) over the intercensal period.

The work of the SAMRC-UCT collaboration which monitored deaths in South Africa during the COVID epidemic has provided us with a solid understanding of the number of deaths in South Africa through to 2023. Likewise, the trend in the number of births is relatively well understood based on the results of previous censuses and surveys, in addition to the births registered by the Department of Home Affairs and captured by Statistics South Africa, as well as those reported by the Department of Health Information System (DHIS), corrected for late or under-reporting. Although the total numbers of intercensal births, and deaths are approximate<sup>5</sup>, they are probably accurate, nationally, to within one or two hundred thousand.

Using the data published in the 2022 census release, together with the results from the 2011 census and our estimates of the numbers of intercensal births and deaths in the country, we can reasonably accurately reconstruct the implied dynamics of the South African population over the intercensal period.

<sup>&</sup>lt;sup>5</sup> For a number of reasons, not least those relating to uncertainty about the level and timing of migration, and the fertility and mortality of migrants, but also data on fertility and mortality from the 2022 census and more up-to-date reports on vital registrations.

Table 1 below demonstrates this reconstructive process, at a national level.

## Table 1: Reconstruction of the South African population and international migration using the balance equation, national

Reported Census 2011 population [1]	51,770,561
+ Estimated intercensal births [2]	12,418,342
- Estimated intercensal deaths [3]	5,924,464
Implied immigration [4]	3,763,063
Reported Census 2022 population [5]	62,027,502
Reported immigration (place of residence) [6]	618,910
Reported immigration (place of birth) [7]	461,542

Reported/implied migration = [6]/[4]	16%
Reported/implied migration = [7]/[4]	12%

#### Notes, sources, and formulae:

[1] Statistics South Africa (2023c: Table 2.1 [page 3])

[2] Own estimate from recorded births corrected for late and under-registration (probably accurate to nearest 100 000)

[3] Own estimate from recorded deaths corrected for late and under-registration (probably accurate to the nearest 100 000)

[4] [5] - ([1]+[2]-[3])

[5] Statistics South Africa (2023c: Table 2.1 [page 3])

[6] Statistics South Africa (2023c: Table 3.1 [page 27]); the reported value of in-migration in Limpopo is mis-calculated based on the data shown<sup>6</sup>.

[7] Own estimate (subtracting estimated survivors of foreign-born identified in the 2011 census from those reported as foreign-born in the 2022 census)<sup>7</sup>.

Crucially, in order to reconcile the estimates from the two censuses and, as a starting point, assuming the estimates from the 2022 census are correct, **net immigration** (that is, those coming to South Africa from outside less those leaving South Africa) **over the intercensal period would have to have been approximately 3.7 million**<sup>8</sup>. Not only does this number seem high given the restriction of travel over the COVID epidemic, but it implies that **only 16% of these migrants identified themselves as such when asked where they were at the time of the previous census and even fewer (12%) when asked about place of birth.** This level of under- and mis-reporting is significantly higher than in the previous intercensal period<sup>9</sup>. Furthermore, the population estimates from the 2011 census are largely consistent with the results from earlier censuses and estimates of births and deaths in earlier intercensal periods. Accordingly, results from the 2022 South African census that give rise to concerns about the data reflect either errors or other problems with the estimates of the population from the 2022 census, or point to patterns of migration in the 2011-2022 intercensal period that are at variance with long-established historically observed patterns of migration by age (with net immigration concentrated in the young adult ages) and sex (Rogers and Castro 1981).

<sup>&</sup>lt;sup>6</sup> Data based on answering "Outside of South Africa" to Question P15c: "In which province did (name) live before moving to this place?" (Statistics South Africa 2022b: 10), which was filtered by Question P14 "Has (name) been staying in this place since October 2011?". Only those answering P14 in the negative, or who had been born after October 2011 and moved since, were asked Question P15c.

<sup>&</sup>lt;sup>7</sup> Data based on answering "Outside of South Africa" to Question P11: "In which province was (name) born?" (Statistics South Africa 2022b: 8)

<sup>&</sup>lt;sup>8</sup> This assumes that both census populations were accurately estimated. Previous work on the 2011 census estimates suggested that in all probability it overestimated the population by a little over 130 000, which, if corrected for, would increase the estimate of the number of migrants needed to produce the number estimated in the 2022 census by the same amount.

Balance equations applied to data from the 2001 and 2011 censuses indicated approximately 3.96 million (3.7 million if one corrects for errors in the censuses) immigrants in that intercensal period. While the quanta are broadly similar, it should be remembered that that the more recent intercensal period was characterised by the Covid pandemic: the impact of lock-down and the pandemic on border-control and international migration might be assumed to reduce overall migration.

<sup>&</sup>lt;sup>9</sup> It is of additional interest to compare these estimates to the 1.769m net immigrants over the period 2011-2021 assumed by Stats SA in their most-recently released population projections and repeated without comment in a recent profile report on migration in South Africa which also made extensive use of the 2022 census data (Statistics South Africa 2022g, 2024).

Table 2 provides the corresponding balance equations for each of the provinces. However, although the provincial numbers sum to the national numbers, the migration required to balance each of the equations at a provincial level comprises both net immigration from outside the country as well as net in-migration from each of the other provinces to each of the provinces<sup>10</sup>.

These estimates of provincial net in-migration (both internal and international) migration are also highly inconsistent with the numbers of migrants estimated from data on reported place of residence at the time of the previous (i.e., 2011) census at a provincial level, as can be seen from the comparison of the reported to implied migration in the last line. In Gauteng, for example, the number of people reporting that they moved to Gauteng from outside South Africa in the intercensal period is

282,000 while the number implied by subtracting the number who reported they moved to the province from another province in the intercensal period from the total number of in-migrants required to balance the equation is over 1 million.

Also, while it may be no surprise that most of the immigrants (both implied by the balance equation and reported in the census) are to Gauteng, it is perhaps questionable (given the reported sources of international migrants) that the second highest number is to be found in the Western Cape at a level almost equivalent to that in Gauteng.

Furthermore, excluding these two provinces, there is virtually no correlation between reported and implied net migration for the other provinces.

Table 2: Reconstruction of the South African population and internation	al migration
using the balance equation, provincial	

	EC	FS	GT	KZ	LM	MP	NC	NW	WC
Reported Census 2011 population [1]	6,562,053	2,745,590	12,272,263	10,267,300	5,404,868	4,039,939	1,145,861	3,509,953	5,822,734
+ Estimated intercensal births [2]	1,583,359	614,332	2,872,855	2,435,180	1,649,006	972,494	300,783	806,972	1,183,361
– Estimated intercensal deaths [3]	893,027	365,217	1,183,511	1,192,071	586,360	525,837	178,353	435,449	564,638
Implied net in-migration from balance equation [4]	(22,181)	(30,292)	1,137,815	913,498	105,206	656,728	87,655	(76,928)	991,562
Reported Census 2022 population [5]	7,230,204	2,964,412	15,099,422	12,423,907	6,572,720	5,143,324	1,355,946	3,804,548	7,433,019
Reported immigration (place of residence) [6]	(62,629)	(9,878)	399,353	(10,067)	(63,917)	12,650	12,094	47,275	294,029
Reported immigration (place of birth) [7]	33,698	14,707	282,842	47,344	50,411	37,219	5,672	36,376	110,641
Implied net in-migration [8]= [4] – ([6] – [7]	74,146	(5,707)	1,021,304	970,909	219,534	681,297	81,233	(87,827)	808,174
Reported/implied migration = [7]/[8]	45%	-258%	28%	5%	23%	5%	7%	-41%	14%

#### Notes, sources, and formulae:

[1] Statistics South Africa (2023c: Table 2.1 [page 3])

[2] Own estimate

[3] Own estimate

[4] [5] – ([1]+[2]-[3])

[5] Statistics South Africa (2023c: Table 2.1 [page 3])

[6] Statistics South Africa (2023c: Table 3.1 [page 27]); the reported value of in-migration in Limpopo is mis-calculated based on the data shown.

[7] Statistics South Africa (2023b: Table 3.1 [page 27])

[8] This estimate reflects the migration implied after allowing for reported immigration determined from place of birh and place of residence.

<sup>10</sup> The sum of inter-provincial net migration is, by definition, zero.

#### Sex profiles of international migrants

Applying the balance equation to the population by sex (Table 3) indicates that more than half of the implied net number of immigrants were female, while based on data from previous intercensal period the expectation was that the net number of male immigrants would outnumber the net number of female immigrants.

# Table 3: Reconstruction of the South African population and international migration using the balance equation, national by sex

	MALES	FEMALES
Reported Census 2011 population [1]	25,188,791	26,581,770
+ Estimated intercensal births [2]	6,255,393	6,162,949
– Estimated intercensal deaths [3]	3,068,872	2,855,591
Implied immigration [4]	1,703,445	2,059,618
Reported Census 2022 population [5]	30,078,757	31,948,745

Notes, sources, and formulae: [1] Statistics South Africa (2023c: Table 2.1 [page 3]) [2] Own estimate [3] Own estimate [4] [5] - ((1)+[2]-[3]) [5] Statistics South Africa (2023c: Table 2.1 [page 3])

# Population group profiles of international migrants

Even though we lack information that could provide guidance as to levels of fertility and mortality over the intercensal period by population group<sup>11</sup>, it is still possible to provide approximate balance equations by applying extrapolated fertility and mortality rates from past trends to the average of the census populations.

The approximate balance equations are provided in Table 4<sup>12</sup>.

# Table 4: Reconstruction of the South African population and international migration using the balance equation, national by population group

	AFRICAN	COLOURED	INDIAN	WHITE
Reported Census 2011 population [1]	41,224,260	4,640,540	1,293,940	4,611,821
+ Estimated intercensal births [2]	10,826,513	947,837	201,054	476,746
– Estimated intercensal deaths [3]	4,893,936	471,279	126,254	436,476
Implied immigration [4]	3,563,836	(41,951)	337,142	(126,293)
Reported Census 2022 population [5]	50,720,672	5,075,148	1,705,882	4,525,800

#### Notes, sources, and formulae:

[1\*] Statistics South Africa (2023c: Table 2.4 [page 6]), apportioning "Other" to the population groups

[2] Own estimate

[3] Own estimate

[4][5\*]-([1\*]+[2]-[3]

[5\*] Statistics South Africa (2023c: Table 2.4 [page 6]), apportioning "Other" to the population groups

<sup>12</sup> An additional assumption is required here: namely that there have not been intercensal changes in self-identification norms and practices in answering this question. While such self-reclassification may indeed have occurred, it is unlikely to have taken place in sufficient numbers to distort the estimates unduly.

<sup>&</sup>lt;sup>11</sup> Births are not classified by population group, and although deaths are, data on these are incomplete, need to be adjusted for under-reporting of deaths (which differ by population group) and are only available up to 2018.

From this we can see that the estimate of the Indian/ Asian population has increased by nearly one-third between 2011 and 2022 (particularly concentrated in KwaZulu-Natal, where the Indian/Asian population apparently increased by over 50% since the 2011 census<sup>13</sup>). This would require an improbably high level of net immigration of people self-identifying as Indian/ Asian compared to the limited immigration observed in the past in this population group. Given the extremely high level of undercount (72.30%) of the Indian/Asian population in the 2022 census identified by the PES, it is much more likely that **the estimate of the Indian/Asian population in the 2022 census is unreliable.** 

In summary, comparison of the 2022 census estimates for the Indian/Asian population with those from population projections (Figure 6 below) suggest that the 2022 census numbers in this population group could be overestimated by around 190 000. Similarly, the 2022 census estimates of the white population may be overestimated by approximately 290 000 people, and implying that net emigration in this group is probably more than double that indicated in Table 4.

# 2.3. Comparison with previous censuses

A second investigation that is always conducted in the assessment of a census is to compare the results from the most recent census to those from previous censuses. If censuses are conducted exactly 10 years apart, those aged 15-19 years (for example) at the time of the earlier census would be aged 25-29 years at the time of the second census. The sources of any observed discrepancy between the two numbers must arise from one or more of the following:

- Errors in the earlier census
- Errors in the later census
- Errors in assumed mortality in that cohort
- Errors in assumed migration in that cohort

On the other hand, consistency in the numbers of the population by age observed in successive censuses, allowing for mortality and migration, points to a general coherence of the data.

In addition, demographers make use of cohortcomponent projection models to provide a counterfactual population based on assumed patterns and trends in mortality, fertility, and migration. Demographers can estimate these components with reasonable accuracy (less so for migration) applying an array of demographic techniques (such as those described in Tools for Demographic Estimation (Moultrie, Dorrington, Hill et al. 2013)) to data from past censuses, national demographic surveys, and vital registration systems.

Again, consistency between the results of past censuses and population projections with the results from a more recent census increases confidence in the reliability of the more recent census data. Conversely, where the results from a more recent census depart markedly from a coherent and consistent prior set of results and population projections, the reliability of the more recent census is called into question.

#### National

Figure 1 shows the national population of South Africa by age as observed in the four post-apartheid censuses, as well as that projected by the UCT-SAMRC collaboration, and Statistics South Africa (as released in their mid-year population estimates).

<sup>13</sup> Statistics South Africa (2023b: Table 2.4 [page 6]); Statistics South Africa (2023c: Table 2.4 [page 6])



Figure 1: Census 2022 numbers by age compared to numbers projected from past censuses and models, National

There is a strong congruence<sup>14</sup> between the estimates of the population (by age in 2022) from previous censuses (up to 2011) allowing for mortality and migration, and the two sets of population projections. The identified undercount of children aged under five in 1996 (who would be aged 26 to 30 in October 2022) is apparent (Dorrington 1999) as well as, to a lesser extent, those aged 5-9 in 1996. The mid-year estimates produced by Statistics South Africa, especially between ages 25 and 39, are somewhat higher (and more in line with those from the 2022 census) than suggested by the UCT population projection. Although the Stats SA methodology for producing their population projections is unclear as to the basis for many assumptions made as well as the parameterisation of the model used to produce the projections, this difference is likely mostly attributable to different assumptions regarding migration.

When one considers the age distribution of the difference between census numbers and those expected based from either of the projection models (Figure 1), only a little more than 40% of the difference is to be found in the typical age range of migration (i.e., as evidenced from the previous intercensal period), namely people aged 20-39 at the 2022 census. This again, creates doubt as to the reliability of the data (either of the adjustment for undercount, or of reported migrants or, more likely, both) provided by the most recent census.

The comparisons of the current census results to those projected from past censuses highlight that earlier census estimates by age are not always entirely consistent with later censuses (for example the population aged under 5 in the 2001 census), demonstrating that census numbers are not always accurate at every age in every census.

<sup>&</sup>lt;sup>14</sup> Congruence between projections of numbers from past censuses and that of the UCT-SAMRC projection is to be expected since past census numbers are projected using projection factors (deaths and migration in this case) used to produce the UCT-SAMRC projection. Any differences from the UCT-SAMRC projection thus reflect issues with the numbers in the original censuses.

While the comparison for the country as a whole clearly reflect an excess in the 2022 census relative to the past censuses or either of the projection models – much of which could possibly be explained by net immigration, especially among those aged 20-39 – it also shows a number of other features that are of concern.

The first, also observed in most provinces, is an **apparent underestimate in the 2022 census of the numbers of children aged 5-9.** There is no evidence to support the extent of declines in fertility and/or child survival that would be required to produce such a low number of children aged 5-9. Initially, noting that the PES adjusted the population aged 5-14 as a single entity, we thought it may be the case that the adjustment for the population aged 5-14 was unable to fully identify or correct for a differentially significant undercount in the 5-9 and 10-14 age groups.

Once Stats SA released data (via SuperWeb) by single years of age, closer inspection revealed significant underestimations of the population at ages 5 and 19 last birthday (Figure 2). When asked, Stats SA were unable to provide any explanation for these distortions.

While the shortfall at age 19 last birthday appears to be compensated for by excesses in the 15-18 age group (since the number in the 15-19 age group in Figure 1 is in line with expectation), that at age 5 last birthday is not similarly balanced by excesses in the 6-9 age range<sup>15</sup>

The second feature evident from Figure 1 is the **excess** of the population in the 2022 census aged 50 and older. Although not evident from the figure, since the numbers in a population decline rapidly with age, the proportional excess remains fairly constant from age 55 and older. This excess is very unlikely to be due to net immigration since there is consistency between the projections at these ages, and a difference that is proportional to the numbers in the population at each age is much more likely due to an error in the overall adjustment for undercount than it is to migration. Thus, it is likely that the 2022 census overestimated the numbers age 50 and older.



#### Figure 2: Census 2022 estimated numbers by single years of age, National

<sup>15</sup> An alternative explanation is that the missing children at age 5 were deliberately miscounted as being aged 0-4 (hence the excess in this age group) in order to avoid the general health and functioning, and most of the educational attainment, modules of the census questionnaire that had to be asked of household members aged 5 and older (Statistics South Africa 2022b).

#### Sex ratios

A second comparison relates to the sex ratio (the number of men per 100 women) as observed in the 1996, 2001, 2011, 2022 censuses, and the two sets of population projections.

Figure 3 compares the ratio of male to female numbers by age for the 2022 census to those produced by the various projections. From this we can see that with the exception of the anomaly in the ratio in the 20-24 age group (those who were approximately 10-14 at the 2011 census), there is broad consensus on the ratio of the numbers of male to females by age for all the projections. **By comparison, the ratio of estimates from the 2022 census shows problems in the 25-39 and 60+ age ranges.** (It is important to note that although the deviations in the ratio in the 25-39 age range are visually small, deviations in this age range represent sizeable numbers in the population.)

The problem in the 25-39 age range is, in all likelihood, connected to the observation that the net numbers of immigrants necessary to reconcile the 2011 and 2022 census numbers require a higher number of women than men immigrating into the country in the intercensal period. However, without more detailed data it is difficult to determine with any certainty whether it is the numbers of women, men, or both, that is at fault.

Regarding the higher ratio for ages 60 and over, based on the data from previous censuses and projections, it is likely that this suggests that the numbers of men over the age of 60 have been overestimated relative to the numbers of women.

#### Provincial age structures

The third investigation is to analyse the data at a sub-national level. Given the centrality of the population numbers by province to the determination of the National Treasury's Equitable Share Formula, our focus here is on the revealed age structure by province.

Figure 4 shows comparisons at a provincial level, similar to those presented in Figure 1 at national level. These show that the undercount of the 5-9-year-olds is apparent in all provinces and the excesses over age 50 in all provinces except Gauteng (again, this is interesting, since this is the province to which most international immigrants are most likely to migrate) and North-West. The comparisons also highlight inconsistencies between the more recent censuses, and the earlier censuses of 1996 (particularly) and 2011 (to a lesser extent).

Other notable province-specific features in Figure 4 are:

Apparent in-migration to the Eastern Cape in the ages
20-34. Historically, there has been out-migration of work-seekers from this province. Although the number of children under age 10 is not inconsistent with the apparent in-migration recorded by the census for this province, comparison with survivors of estimates of the number of births in the province over the intercensal period suggest that the 2022 census estimate of children aged 0-4 is too high.







Figure 4: Census 2022 estimated numbers by age compared to numbers projected from past censuses and models, Provinces (in thousands)

- Although, apart from the problems common to all provinces, the census estimates for the Free State correspond reasonably closely to the projections for adults, this is not so for children under 15. This suggests that the census may have underestimated the numbers of children under 15. The alternative explanations (rapid decline in fertility or significant out-migration of children) both seem implausible given the available data (from the DHIS and from vital registration on births in the province.
- In-migration into Gauteng in the age range 20-39 seems probable (and is supported by Stats SA's mid-year estimates). In that case, however, it is possible that the numbers of children under age 10 are underestimated by the census unless there was higher-than-expected out-migration of these children than already allowed for in the projection models.
- In-migration to KwaZulu-Natal over the whole of the adult age range, when in the past the pattern has been out-migration mainly to seek work. In addition, the estimates of the numbers of children under age 10 are a bit low, implying, even if the census estimates of the numbers of adults were reduced, a drop in fertility rates that is inconsistent with the numbers of births occurring in the province.
- Although the census estimates for Limpopo also imply in-migration over the whole adult age range, it is more difficult to dismiss this out of hand since inmigration (both of foreign-born as well, at the older ages, of South African-born returning to Limpopo as home province) is possible.
- Similarly, one needs to keep an open mind about net migration of adults to Mpumalanga. However, in this case comparison to estimates of the numbers of births in the five years prior to the census suggest that the 2022 census overestimates the numbers of children aged under 5 in the province.
- There is a great deal of uncertainty about the estimates for the Northern Cape, in both past and present censuses. Not only are the censuses inconsistent with one another, both as to level and shape (i.e., distribution by age), but so are the projections from models.
- Although the 2022 census shows no excess over age 50 for North-West, the 2022 census numbers are inconsistent with all past censuses (particularly 2001

and 2011<sup>16</sup>) and both population projections for ages below 50. Without more information, it is hard to say to what extent there has been significant out-migration and to what extent these differences reflect poor estimation in the 2022 census. However, comparison of the numbers of births 5 to 14 years prior to the census suggest that **the census underestimates the numbers of children in the province aged 5-14**.

There appears to have been significant in-migration into the Western Cape. Although in-migration into the province is probably not unexpected, it is unlikely that the bulk of the migrants would be newly-arrived immigrants. The extent to which these numbers depend on immigration from outside the country needs further investigation. Comparison of the numbers of births 5 to 9 years prior to the census suggest that the census underestimates the numbers of children in the province aged 5-9.

In summary, the 2022 census estimates of the total population and by age to varying extents are probably reasonable approximations for the Free State, Gauteng, North-West and Western Cape (assuming some in-migration at older ages), possibly less so for Limpopo and Mpumalanga, and poor estimates for the Eastern Cape (unlikely to have been net immigration into the province in the 20-39 age group), KwaZulu-Natal (overstated for all adult ages, since usually experiences out-migration), and unknown for the Northern Cape (which has proven difficult to estimate in the past as well).

#### Provincial sex ratios

Considering the ratio of males to females by age for the provinces (Figure 5) there is far less consistency than at the national level. There are problems with the sex ratio of the census numbers for all provinces except Gauteng and the Western Cape, and especially for KwaZulu-Natal, Mpumalanga and the Northern Cape. In particular, the following features (roughly) are apparent:

- With the exception of KwaZulu-Natal there is broad consensus between the ratios found in past censuses and the UCT-SAMRC projection, with ratios produced by the 2022 census in particular and by the mid-year population estimates (MYPEs) to a lesser extent, deviating from this pattern.
- <sup>16</sup> Much of the excess implied by the comparison of the census estimates to those produced by projecting the 1996 census to 2022 is due to changes in the provincial boundaries between 1996 and 2001.

- For the Free State, Mpumalanga, the Northern Cape, and North-West, and to a lesser extent KwaZulu-Natal the male to female ratio is low over the age range 20-44 or a bit older.
- For the Eastern Cape, KwaZulu-Natal, Limpopo, and Mpumalanga and to a lesser extent the Free State, Northern Cape and North-West, the male to female rates are too high from about age 55 or 60 onwards.
- There is a noticeable lack of smoothness in the ratio over age in Limpopo and to a lesser extent KwaZulu-Natal, Mpumalanga, and North-West.

Further investigation will be needed to determine whether the problem is with the estimates of the number of males, females or both.

# Figure 5: Census 2022 ratio of numbers of males to females by age compared to the ratios projected from past censuses and models, Provinces



#### Population group age structures

A possible explanation for the excess of the census estimates over the projections for ages over 40 is provided by a comparison of the census estimates by population group<sup>17</sup> to those projected from past censuses and by the UCT-SAMRC collaboration and Stats SA in Figure 6.

If one assumes that the excess of the census estimates of the African population relative to the projections in the age range 20-39 is due to migration unaccounted for by the projections, the estimates from the census for the African and coloured population groups are close to the numbers from the projections. Thus, the bulk of the excess seen in the national population above age 50 is to be found in the Indian and white population groups.

Given that the estimated undercount of these two population groups was exceedingly high (72% of the Indian and 62% of the white population group were not counted according to the Post-Enumeration Survey (Statistics South Africa 2023e)), it is quite likely that much of the excess of the census above the projections seen at the





<sup>17</sup> Those recorded as "Other" or "Unspecified" (accounting for about 0.5% of the total population) were distributed to the four specified population groups in proportion to the numbers recorded for each population group.

national level above age 50 is due to an over-adjustment for undercount (with the true numbers lying somewhere between the UCT-SAMRC and Stats SA projections). The excess of the census over the projections over the whole age range amounts to 24% of the projected population for the Indian group and 14% for the white population group.

#### Population group sex ratios

The ratio of the numbers of males to females by age for the four population groups is shown in Figure 7. From this we see that there is a broad consensus between the census estimates and projections for the African and coloured population groups (except that for the African population the MYPE ratio is lower for 60-79, and for the coloured population both projections are slightly higher for adult ages under 60).

For the Indian population group the 2022 census ratio appears to be too low in the 30-39 age range and too high for ages 70 and over, and while the ratios are broadly consistent for the white population group, the ratio for the census is also high (though consistent with the ratio for the MYPE) for ages 70 and over.

The MYPEs also differ from those of the 2022 census but again there are some discrepancies between the MYPE ratios and those of past censuses.



# Figure 7: Census 2022 ratio of numbers of males to females by age compared to the ratios projected from past censuses and models, Population Groups

# 2.4. Concerns with district and municipal estimates

Census data inform policy and planning at all spatial levels of the country. The provincial population estimates are a key component of the Equitable Share Formula used to allocate disbursements from the national government to the provinces. At district and municipality levels, population estimates might inform service delivery (especially of health and education), infrastructural needs analysis and investment, as well as affect financial transfers to districts and municipalities.

Although the focus of this report is on the estimates at a national and provincial level it is useful to highlight the extent of the uncertainty about the estimates at a subprovince level. At this stage, the work is illustrative: to determine alternative district- and municipality-level population estimates is not a simple exercise. Not only is there much greater uncertainty about the demographic components of fertility, mortality and migration, but the most difficult component to estimate, migration, is demographically orders of magnitude more significant for less populous areas. Thus, comparisons with population numbers from other sources (such as the records of births by district, children in school, antenatal clinic attendance, registered voters, etc.) are important in helping produce a best estimate of these sub-provincial populations. Three distinct investigations are presented below to illustrate the extent of the problem. The first examines the intercensal growth rate in the total population by district. The second compares the district estimates from the census with those produced by Statistics South Africa in their mid-year population estimates (MYPE). The third examines the 2011 and 2022 census data on those aged 18 and over in five local municipalities in South Africa, comparing the results from the censuses with voter registration data from the Independent Electoral Commission (IEC).

# Intercensal population growth rates by district, 2011-2022

With the release of the data from the 2022 census, Statistics South Africa recast the 2011 district population estimates to the 2022 boundaries. Theoretically, at least, this exercise should remove discrepancies in population estimates from the census at a district level arising from changes in boundaries.

Table 5 shows the implied annual intercensal growth rates between the 2011 and 2022 censuses for each of the 52 district municipalities and metros, ranked by descending magnitude.

CODE	DISTRICT NAME AND PRO	VINCE	2011	2022	INTERCENSAL GROWTH 2011-2022
DC4	Garden Route	WC	574,261	838,457	46.0%
DC5	Central Karoo	WC	71,010	102,173	43.9%
DC3	Overberg	WC	258,175	359,446	39.2%
TSH	City of Tshwane	GT	2,921,490	4,040,315	38.3%
DC25	Amajuba	KZN	499,836	687,408	37.5%
DC32	Ehlanzeni	MP	1,691,267	2,270,897	34.3%
DC21	Ugu	KZN	592,543	773,402	30.5%
DC42	Sedibeng	GT	916,490	1,190,688	29.9%
BUF	Buffalo City	EC	755,200	975,255	29.1%
DC29	iLembe	KZN	606,814	782,661	29.0%
DC6	Namakwa	NC	115,841	148,935	28.6%
DC35	Capricorn	LM	1,130,296	1,447,103	28.0%
EKU	Ekurhuleni	GT	3,178,472	4,066,691	27.9%
CPT	City of Cape Town	WC	3,740,024	4,772,846	27.6%
DC24	Umzinyathi	KZN	510,848	649,261	27.1%
DC1	West Coast	WC	391,766	497,394	27.0%

#### Table 5: Intercensal growth rates 2011-2022, by district

CODE	DISTRICT NAME AND PRO	VINCE	2011	2022	INTERCENSAL GROWTH 2011-2022
DC10	Sarah Baartman	EC	422,281	533,253	26.3%
DC33	Mopani	LM	1,092,512	1,372,873	25.7%
DC43	Harry Gwala	KZN	451,288	563,893	25.0%
DC47	Sekhukhune	LM	1,076,318	1,336,805	24.2%
ETH	eThekwini	KZN	3,442,360	4,239,901	23.2%
DC30	Gert Sibande	MP	1,043,196	1,283,459	23.0%
DC28	King Cetshwayo	KZN	833,181	1,021,344	22.6%
DC48	West Rand	GT	820,995	998,466	21.6%
DC31	Nkangala	MP	1,308,130	1,588,968	21.5%
DC22	Umgungundlovu	KZN	1,017,759	1,235,715	21.4%
DC45	John Taolo Gaetsewe	NC	224,801	272,454	21.2%
DC8	ZF Mgcawu	NC	236,791	283,624	19.8%
DC23	Uthukela	KZN	668,074	789,092	18.1%
DC27	Umkhanyakude	KZN	625,812	738,437	18.0%
DC26	Zululand	KZN	803,577	942,794	17.3%
DC44	Alfred Nzo	EC	801,349	936,462	16.9%
DC7	Pixley ka Seme	NC	186,340	216,589	16.2%
DC9	Frances Baard	NC	382,080	434,343	13.7%
DC19	Thabo Mofutsanyana	FS	736,237	831,421	12.9%
DC14	Joe Gqabi	EC	349,768	393,048	12.4%
DC36	Waterberg	LM	679,340	762,862	12.3%
DC38	Ngaka Modiri Molema	NW	842,692	937,723	11.3%
DC15	OR Tambo	EC	1,364,947	1,501,702	10.0%
DC39	Dr Ruth Segomotsi Mompati	NW	463,817	508,192	9.6%
DC2	Cape Winelands	WC	787,489	862,703	9.6%
MAN	Mangaung	FS	747,432	811,431	8.6%
JHB	City of Johannesburg	GT	4,434,829	4,803,262	8.3%
DC18	Lejweleputswa	FS	627,629	679,746	8.3%
DC16	Xhariep	FS	121,943	131,901	8.2%
DC34	Vhembe	LM	1,531,389	1,653,077	7.9%
DC37	Bojanala	NW	1,507,501	1,624,428	7.8%
DC40	Dr Kenneth Kaunda	NW	695,933	734,203	5.5%
DC20	Fezile Dabi	FS	488,032	509,912	4.5%
DC13	Chris Hani	EC	800,260	828,387	3.5%
NMA	Nelson Mandela Bay	EC	1,152,114	1,190,496	3.3%
DC12	Amathole	EC	854,966	871,601	1.9%

The results reveal a somewhat surprising range with a number of unexpected intercensal growth rates. While the national population is reported to have grown by just over 20% over the intercensal period, seven districts show growth in excess of 30% over the intercensal period, including in some of the most rural and poorest parts of the country (e.g. the Central Karoo and Overberg). At the other extreme,

what was once certainly the most populous city in South Africa, the City of Johannesburg, grew by only 8% over the intercensal period. However, in contrast, adjacent metros to the City of Johannesburg grew by 28% (Ekurhuleni) and 38% (Tshwane) respectively, so to some extent this difference might be due to intercity migration.



Figure 8: Population estimates of the Central Karoo District and City of Johannesburg, 2011 and 2022 censuses, by age group

The age structures of the Central Karoo District and the City of Johannesburg in the 2011 and 2022 censuses are shown in Figure 8. While the age structure in each district is broadly consistent between censuses (with the City of Johannesburg showing a structure heavily influenced by past in-migration of working-age adults), it is hard to comprehend how the population of the Central Karoo has increased in the manner shown. Neither fertility nor mortality can account for the dynamics over the intercensal period (and the population structure shows lesser growth in the population aged under 10, those who would have been born between the two censuses), meaning that the difference must be accounted for by an improbable age-pattern of migration into the district, or by material issues with the adjustments made to population counts at a district level.

#### Case study: Beaufort West Local Municipality

Clarity as to the nature of the problem is provided by considering the single largest municipality in the Central Karoo, Beaufort West Local Municipality (WC053). This Local Municipality accounted for 69.8% of the population of the Central Karoo District Municipality in 2011, and 71.4% of the district population in 2022. According to the census, the population of this local municipality grew by 47.2% (3.7% per annum) over the intercensal period, from 49.585 to 72,972 people. The age structure of the population at the time of the 2011 and 2022 censuses is shown in Figure 9.

The increase in the population is spread across all age groups, and again points more to migration than fertility or mortality as the driver of this growth. Proportionately the largest increase was in the 65-69 age group, which grew by nearly 6 percent per annum over the intercensal period.

A significant undercount of this population in the 2011 census cannot be a plausible explanation, being largely consistent with prior censuses as well as with the number of households in the municipality. Figure 10 shows a satellite image from Google Earth of the town of Beaufort West (itself accounting for approximately 70% of the population of the Local Municipality, and thus about half the population of the entire Central Karoo district) in 2022 with the growth in the town from January 2010 outlined in red. As is readily apparent, the increase in the number of dwellings simply cannot account for the increased population, especially when the average household size in the local municipality is reported by Statistics South Africa to have remained constant at 3.8 persons per household. In addition, there is no evidence of densification of the urban space, or other changes in the town or municipality's morphology that could explain the purported intercensal increase in the municipal population.

Figure 9: Age structure of the population of WC053: Beaufort West Local Municipality, 2011 and 2022 censuses



Figure 10: Satellite image from Google Earth of Beaufort West town in January 2022, with areas developed since January 2010 outlined in red



Similar investigations in other districts, as well as the broad consistency between the 2011 census data at a district level with earlier census results after allowing for boundary changes, lead to similar conclusions: **the district and local municipality populations estimated by the 2022 census are likely to be quite unreliable and should be treated with extreme caution**.

# Comparison with Statistics South Africa's mid-year population estimates

The second investigation pursued is to compare the 2022 census results at a district level with the district-level population estimates released by Statistics South

Africa as part of its 2022 mid-year population estimates (MYPEs). Exponential interpolation between the district population estimates for mid-2021 and mid-2022 to the census date (2 February 2022) permits direct comparison of these numbers with those from the census.

As with the analysis in the preceding section, discrepancies between the MYPE and census results at a district level cannot be taken as prima facie evidence of errors in the census: the MYPE may have mis-estimated any of the demographic components, especially migration, at a district level. Nevertheless, significant differences require investigation in order to understand why or how those differences might have occurred.

## Table 6: District populations as shown by the mid-year population estimates (interpolated to February 2022) and the 2022 Census

CODE	DISTRICT NAME AND PRO	VINCE	MYPE	CENSUS 2022	DIFFERENCE CENSUS: MYPE
DC5	Central Karoo	WC	75,420	102,173	35.5%
DC4	Garden Route	WC	628,042	838,457	33.5%
DC6	Namakwa	NC	116,162	148,935	28.2%
DC32	Ehlanzeni	MP	1,817,221	2,270,897	25.0%

CODE	DISTRICT NAME AND PRO	OVINCE	МҮРЕ	CENSUS 2022	DIFFERENCE CENSUS: MYPE	
BUF	Buffalo City	EC	798,976	975,255	22.1%	
DC42	Sedibeng	GT	975,691	1,190,688	22.0%	
DC25	Amajuba	KZN	564,590	687,408	21.8%	
DC13	Chris Hani	EC	705,087	828,387	17.5%	
DC3	Overberg	WC	306,606	359,446	17.2%	
DC34	Vhembe	LM	1,438,364	1,653,077	14.9%	
DC24	Umzinyathi	KZN	565,108	649,261	14.9%	
DC14	Joe Gqabi	EC	342,521	393,048	14.8%	
DC23	Uthukela	KZN	690,795	789,092	14.2%	
DC33	Mopani	LM	1,205,492	1,372,873	13.9%	
DC44	Alfred Nzo	EC	823,878	936,462	13.7%	
DC29	iLembe	KZN	688,951	782,661	13.6%	
DC43	Harry Gwala	KZN	503,660	563,893	12.0%	
DC12	Amathole	EC	784,122	871,601	11.2%	
DC10	Sarah Baartman	EC	483,000	533,253	10.4%	
DC47	Sekhukhune	LM	1,212,792	1,336,805	10.2%	
DC35	Capricorn	LM	1,314,239	1,447,103	10.1%	
DC22	Umgungundlovu	KZN	1,128,332	1,235,715	9.5%	
DC19	Thabo Mofutsanyana	FS	771,528	831,421	7.8%	
DC28	King Cetshwayo	KZN	955,315	1,021,344	6.9%	
DC26	Zululand	KZN	882,227	942,794	6.9%	
DC39	Dr Ruth Segomotsi Mompati	NW	477,334	508,192	6.5%	
DC27	Umkhanyakude	KZN	694,981	738,437	6.3%	
DC1	West Coast	WC	469,584	497,394	5.9%	
ETH	eThekwini	KZN	4,024,390	4,239,901	5.4%	
TSH	City of Tshwane	GT	3,875,752	4,040,315	4.2%	
DC9	Frances Baard	NC	417,575	434,343	4.0%	
DC7	Pixley ka Seme	NC	208,330	216,589	4.0%	
DC48	West Rand	GT	968,079	998,466	3.1%	
DC38	Ngaka Modiri Molema	NW	909,394	937,723	3.1%	
DC18	Lejweleputswa	FS	662,744	679,746	2.6%	
DC30	Gert Sibande	MP	1,260,336	1,283,459	1.8%	
CPT	City of Cape Town	WC	4,724,256	4,772,846	1.0%	
DC36	Waterberg	LM	756,091	762,862	0.9%	
DC16	Xhariep	FS	130,879	131,901	0.8%	
DC8	ZF Mgcawu	NC	285,299	283,624	-0.6%	
DC20	Fezile Dabi	FS	515,574	509,912	-1.1%	
EKU	Ekurhuleni	GT	4,116,347	4,066,691	-1.2%	
DC45	John Taolo Gaetsewe	NC	276,687	272,454	-1.5%	
DC15	OR Tambo	EC	1,529,533	1,501,702	-1.8%	
NMA	Nelson Mandela Bay	EC	1,212,603	1,190,496	-1.8%	
DC31	Nkangala	MP	1,621,103	1,588,968	-2.0%	
MAN	Mangaung	FS	837,577	811,431	-3.1%	
DC21	Ugu	KZN	801,476	773,402	-3.5%	
DC2	Cape Winelands	WC	962,607	862,703	-10.4%	
DC40	Dr Kenneth Kaunda	NW	822,621	734,203	-10.7%	
DC37	Bojanala	NW	1,956,030	1,624,428	-17.0%	
JHB	City of Johannesburg	GT	6,041,207	4,803,262	-20.5%	

Comparison of these estimates to those from the 2022 census (Table 6) show that the difference between the MYPEs (interpolated to the census date) at a district level and those from the census was greater than 10% for nearly half (21/52) of the districts. The biggest difference (35% discrepancy, at the same notional date) was again in the Central Karoo. At the other extreme, the City of Johannesburg's population in the census was more than 20% lower than that predicted by the MYPE.

These differences are likely to have a material impact on resource allocation, as well as policy planning, formulation, and implementation at district levels. The fact that the fastest intercensal growth (and often the biggest discrepancies between the MYPE recast to the census date and the census) is seen in some of the most rural and least hospitable parts of the country again leads us to caution against the use of the census data at subprovincial levels.

#### Comparison of local municipality population estimates in 2011 and 2022 with voter registration data from the 2011 and 2021 Local Government Elections

The third investigation is to compare the changes in local municipality populations aged 18 and over as revealed by the last two censuses with those of the voter registration data released by the South African Independent Electoral Commission (IEC). There are more than 200 local municipalities in the country, and the intercensal growth rates between 2011 and 2022 have been released by Stats SA (Statistics South Africa 2023b). Excluding those municipalities that may have been subject to significant boundary changes between the two censuses, the growth rates range from 102.4% (a more-than-doubling in just over a decade) in Richtersveld (NC061, Northern Cape), to -23.6% in Thabazimbi (LIM361, Limpopo).

Of course, not all adults over the age of 18 in a given municipality are registered voters, and there may be wide variation in the proportion of adults registered to vote across municipalities. However, one would anticipate that the proportions of registered voters relative to all adults over the age of 18 in any area would be roughly stable. Table 7 shows the populations aged 18 and over, the numbers of voters registered by the IEC, and selected metrics in five local municipalities with amongst the biggest and smallest percentage increases in their census populations, and which did not experience material changes to their boundaries in the intercensal period.

Both metrics presented (the proportion of registered voters relative to the proximate census population of adults and the increase in census population compared with the number of registered voters disclosed by the IEC) point to significant inconsistencies between the two data sources. In Richtersveld, for example, the census adult population is reported to have increased by over 100%, yet the number of registered voters increased by less than 5% over the same period. Conversely, in Thabazimbi in Limpopo, according to the census the population of adults decreased by nearly a quarter, yet the number of registered voters increased by nearly a fifth.

As we explain in Section 3, we further believe that these anomalous results are caused by incorrect correction for undercounts which are the likely consequence of the combination of poorly planned and executed census and post-enumeration survey operations.

		BEAUFORT WEST WC053	MOSSEL BAY WC043	OVERSTRAND WC032	RICHTERSVELD NC061	THABAZIMBI LM361
2011	Census – 18+	31,253	64,878	60,250	8,581	64,471
2011	LGE RV	23,812	52,416	45,298	7,096	38,917
2022	Census – 18+	48,418	109,422	103,283	17,369	49,278
2021	LGE RV	26,878	64,310	59,979	7,420	45,688
Proportion	2011	76.2%	80.8%	75.2%	82.7%	60.4%
RV	2021/2	55.5%	58.8%	58.1%	42.7%	92.7%
Increase 2011:2021/2	Census	54.9%	68.7%	71.4%	102.4%	-23.6%
	RV	12.9%	22.7%	32.4%	4.6%	17.4%

## Table 7: Populations aged 18 and over, numbers of registered voters, and selected ratios in 5 illustrative districts, 2011 and 2021/2

Notes: LGE=Local Government Elections; RV = Registered Voters

### 3. CONCERNS ABOUT THE 2022 POST-ENUMERATION SURVEY

This section identifies a number of concerns with the 2022 Post-Enumeration Survey (PES) which was used to establish the extent of the undercount, and thence to adjust the enumerated population to be more reflective of the population of South Africa in February 2022. The concerns relate both to the internal coherence of the PES itself, as well as the adjustments indicated by the PES that were used to produce the final estimate of the 2022 South African Census population.

#### 3.1. The Post-Enumeration Survey (PES)

The United Nations Statistics Division (UN Statistics Division 2010) recommends that Post-Enumeration Surveys are carried out shortly after a census to estimate the extent of (and, often, to adjust for) a census undercount. It defines a PES as:

"the complete re-enumeration of a representative sample of the census population and matching each individual who is enumerated in the post-enumeration survey with information from the main enumeration. The objectives of the post enumeration survey can be summed up as follows:

- (a) To assess the degree of coverage during census enumeration.
- (b) To examine the implications of coverage deficiencies, if any, on the usefulness of the census data.
- (c) To obtain information for the design of future censuses and surveys.
- (d) To examine the characteristics of persons who may have been missed during census enumeration."
  (UN Statistics Division 2008: 86)

A post-enumeration survey – even though conducted as a single exercise – has two distinct elements:

 Using the result of a matching exercise between the census and the PES data to estimate the number of persons missed in the census, in a capturerecapture process.  Using the PES to determine if the census population sampled in the PES was indeed correctly enumerated, in terms of both content (e.g. population group, sex, or age) and coverage (with elements being neither duplicated nor incorrectly included in the census sampling frame nor incorrectly located) (US Bureau of the Census 1985).

Importantly, it can be seen that the PES exercise is not an attempted complete re-enumeration of a national population (which would be as time consuming and logistically complicated and expensive as running another census), but only a partial sample of the population.

Matching respondents in a census to respondents in a PES entails a complex exercise to establish whether one has indeed 'recaptured' a person previously 'captured' in the census. Discrepancies in names, dates of birth, can arise, and complicate matters further. We consider the implications of the matching exercise undertaken in South Africa further in Section 3.5.

In addition, the UN recommends that a PES has to be "carried out within a few months of the end of the census to ensure that the impact of natural population changes (births, deaths and migration) and lapses in respondent recall do not hopelessly complicate the exercise" (UN Statistics Division 2008: 87).

The function and design of the PES is set out in more detail in Appendix 3.

#### 3.2. The 2022 South African Census PES

The UN's recommendation that the PES fieldwork should follow shortly after the main census fieldwork is significant in the context of the 2022 South African Census. A major complicating aspect in this regard was the extended duration of the census fieldwork itself – which ended up lasting for nearly four months (from early February to the end of May 2022, but always with a reference date of 2 February 2022). Thus, even though the PES began shortly after the census fieldwork ended, the extended period of data collection in the census compromised the integrity of the PES, and complicated the matching process while increasing the potential impact of migration, fertility, and mortality between the census date (which was the reference date for all census data collection) and the PES.

As noted in Section 2.1, the estimated undercount in the 2022 census was 31%, more than twice that observed in 2011, and extremely high even among developing countries, where PES-estimated undercounts tend to fall well below 10%. An undercount of this magnitude is consequential – not only in terms of the greater uncertainty surrounding estimates of the total population size, but also because of the implications for the reliability of estimates for smaller populations, particularly those defined at a granularity finer than the dimensions used to stratify the PES.

It goes without saying, and this is evident from the estimates of the undercount, that the PES-derived weights are doing much more heavy lifting in producing an estimate of the population in 2022 than would be the case if the undercount was lower. In other words, in the 2022 census, the population estimates will have ended up looking far more like the adjustment factors than in previous censuses, and hence it becomes all the more important to examine the plausibility of those adjustment factors.

#### Sampling concerns

A more detailed exposition of our concerns around methodological aspects of the PES is included in Appendix 3. Unfortunately, the description of the 2022 PES sampling methodology is opaque, and the tabulations provided are – in many respects – not directly comparable with those provided as part of the PES reports from earlier censuses.

With an undercount rate more than twice the 2011 undercount rate, our concern is that the sample size of the 2022 Census PES may have turned out to be far too small to reliably estimate the undercount.

# Estimates of the undercount and the population size, and their standard errors, 2011 and 2022

As we noted in Section 2.1, it is important to re-emphasise that the output from many censuses (South Africa's included) should be treated as estimates rather than counts of the population. This is so, because, the counts, quite understandably, are adjusted for an estimate of the extent to which the census was under- (or, less likely, over-) counted. The extent to which the census count is an under-/over-estimate is determined by the Postenumeration Survey (PES).

Our key concern in this regard is illustrated by comparison of two tables from each of the 2011 and 2022 PES reports.

The first comparison (presented in Table 8) shows estimates of undercounts and uncertainty about those estimates derived from the results of the survey. From this we see that the estimates of the undercount from the PES were about twice as high in 2022 than in 2011. However, what is of greater interest is that the estimates of uncertainty about these estimates (the standard error (SE)) in the 2022 PES are, as might be expected, markedly higher than those of the 2011 PES. Table 8: Estimates of the undercount (UC) and their standard errors (SE), South Africa and its provinces as estimated directly from the post-enumeration survey<sup>18</sup>, 2011 and 2022

	20	11	20	22
	EST. UC (%)	SE (%)	EST. UC (%)	SE (%)
WC	18.5	0.542	35.56	1.56
EC	12.9	0.196	30.56	0.59
NC	13.4	0.318	29.11	2.27
FS	10.1	0.362	18.52	2.62
KZN	16.7	0.379	31.19	1.08
NW	14.9	0.532	17.92	5.39
GT	14.7	0.174	30.79	2.07
MP	15.5	0.473	34.00	4.51
LP	10.0	0.135	23.47	1.02
ZA	14.6	0.132	29.6	0.82

Source: 2011 PES Report (Table 10); 2022 PES Report (Table 1)

Based on the PES, the 95% confidence interval for the undercount in 2011 was narrow (14.34%; 14.86%). In 2022, the equivalent interval was, as would be expected, wider (27.99%; 31.21%). However, as can be seen from Table 9, at the point where the report provides 'final estimates' of the population having corrected for the undercount, the uncertainty surrounding the census estimate for 2022 has all but disappeared, nationally, and is only higher (relative to 2011) in the 2022 census in the Western Cape and comparable (but still lower) in KwaZulu-Natal and Gauteng.

Table 9: Estimates of the total population, their standard errors (SE), South Africa and its provinces, 2011 and 2022

	2011		2022	
	POPULATION	SE	POPULATION	SE
WC	5,822,734	40,830	7,430,000	70,000
EC	6,562,053	125,810	7,230,000	60,000
NC	1,145,861	82,466	1,350,000	30,000
FS	2,745,590	117,567	2,960,000	40,000
KZN	10,267,300	109,994	12,390,000	100,000
NW	3,509,953	166,754	3,800,000	40,000
GT	12,272,263	106,023	15,120,000	90,000
MP	4,039,939	219,299	5,160,000	60,000
LP	5,404,868	251,244	6,570,000	60,000
ZA	51,770,560	997,560	62,030,000	120,000

Source: 2011 PES Report (Table 14); 2022 PES Report (Table 5)

Unfortunately, the two PES reports are not of any help in estimating the likely uncertainty about the estimates of the Indian or white populations. While the 2022 PES report does provide estimates, of the standard errors of the undercount of 0.82% for the Indian population and 0.42% for the white population (both of which are clearly implausible given the extent of the undercount in these two populations), the 2011 PES report does not, which makes it impossible to gauge the scale of the possible error in these estimates.

In short, though, it is difficult to understand how it can be that the two censuses (2011 and 2022) had broadly equivalent PES sampling fractions but – with the undercount in 2022 roughly twice as high – it can be said that one can be 95% certain that the true size of the population in 2022 is within 235,000 of the estimated population size, whereas in 2011, we could only be 95% certain that the true population was within 1.955 million either side of the estimated population size. In the absence of any other information, this would appear to be a computational error.

<sup>&</sup>lt;sup>18</sup> Table 10 of the 2011 report is described as "in-scope sub-universe" while Table 1 of the 2022 report is described as "(p)reliminary net undercount rate for persons".

# 3.3. District-level concerns and conclusion

The extent of the documented undercount together with our concerns about the scope and power of the PES to fully or reliably adjust the census data for that undercount might also be responsible for many of the sub-provincial anomalies identified in the 2022 census data.

Principally, the PES was not designed to correct or adjust the population at levels of granularity finer than those used to stratify the PES sample (which, spatially, only went down to 26 strata: 9 provinces, and by type of place of residence (urban, farms, and traditional areas – in the Western Cape, the 'farm' and 'traditional areas' strata were combined on account of the small number of 'traditional' areas in that province).

As we have noted, the PES covered only 840 of the more than 130,000 enumerator areas (EAs). In five provinces (Northern Cape, Free State, North West, Mpumalanga, and Limpopo) the PES was conducted in fewer than 100 EAs. As a result, areas in those strata with broadly similar demographic composition (by age, sex, and population group, for example) would have received similar adjustments in respect of the undercount, even though we know from the Table 1 of the 2022 PES report that the width of the 95% confidence interval of the preliminary undercount exceeded 5% in six of the nine provinces, the exceptions being the Eastern Cape, KwaZulu-Natal and Limpopo.

We have already suggested that the size and scale of the PES was inadequate to provide reliable estimates of the population, and this may well account for the rather odd results evident at district (and even more so, municipal) levels, as well as the implausible reported population numbers of certain categories identified in this report: for example, Indian/Asians and whites; in municipal and district populations and even the total estimated population in some provinces.

All things considered, we are extremely concerned that the results of the PES may have been mis-represented with there being considerably greater uncertainty as to the numbers than has been reported. In this regard, the census results cannot be relied upon for policy, planning or other purposes at sub-provincial levels spatial disaggregation, or (at the very least) to provide meaningful insight into those other subpopulations mentioned.

### 4. WHAT HAPPENED?

Had they been consulted, most demographers would have strongly motivated for the census to be deferred to 10 October 2022, or even 10 October 2023, to ensure that processes and implementation were not rushed, that it was on an anniversary of previous censuses, and that the census took part when the population was most stable (in terms of migration and other interruptions).

While a full analysis of all the factors that may have contributed to the poor execution of the 2022 South African Census is beyond the scope of this Technical Report, and would require access to internal documents that are not in the public domain, it is nonetheless possible to offer indicators as to some of the factors that impeded the successful conduct of the census.

Appendix 4 expands in much greater detail on, and provides further evidence relating to, many of the aspects summarised below.

SARS-CoV-2 was first detected in South Africa in March 2020, which resulted in lockdowns of variable severity and repeated waves of infection and excess mortality that lasted until mid-2022. Although it is not clear to what extent preparations for a census in October 2021 were on track in March 2020, there is no doubt that the outbreak of

the Covid-19 epidemic was a major interruption, forcing Stats SA to delay the census.

Unfortunately, the threat of the withdrawal of funding by National Treasury if the census was not undertaken in the financial year ending March 2022 offered Stats SA Hobson's choice – either carry out the census in that financial year or lose the funding for the census, possibly to 2031. Thus, Stats SA was forced to undertake the census while not ready to do so, and entered the enumeration period in a state of significant unreadiness, particularly with regards to:

- The recruitment and training of field staff.
- Being unready and ill-prepared to mount a significant field work operation, having perhaps relied excessively on the hope that untried remote data collection (e.g. via Web or telephonic enumeration) would make things easier.
- Logistics and IT systems not being fully stress-tested.

Every indication is that this decision impelled Statistics South Africa to set the census night (2 February) in order to have a realistic prospect of completing (at least the bulk) of census enumeration and PES fieldwork before the end of the fiscal year on 31 March.







































## 5. CONCLUSIONS

This section sets out the conclusions that we draw from the analysis presented in the previous sections. We concentrate on two fundamental questions: first, to offer an opinion on whether the census data are fit for purpose; and second, to outline the next steps that need to be taken.

#### 5.1. Are the data fit for purpose?

Despite pronouncements by both the Statistician-General and the Chair of Statistics Council that Census 2022 is fit for purpose, there is sufficient doubt about the quality and content of the data produced by Stats SA from Census 2022 to conclude that – as they currently stand – the data, at least in terms of estimates of the numbers of people, are probably not fit for purpose to assist with fiscal allocation, or for national, provincial, or particularly local government planning and resource allocation, or for projects such as estimating the National Burden of Disease.

Both public and private sectors should exercise caution in making long-term plans or drawing policy conclusions based on these data. We are particularly concerned that – as they currently stand – the estimates of the population (by age, sex, province, population group) may lead to significant misallocation of resources through (for example) the Equitable Share Formulae; or of education and health resources at national, provincial, or local government levels.

At the most basic level, we believe that the estimated count of the population, at least on the dimensions described in the previous paragraph, is substantially incorrect. The extent of the undercount revealed by the PES, as well as our doubts about the accuracy of and uncertainty about the PES adjustments applied to the enumerated populations to derive the final estimated counts, suggest that this census was not a success.

Although this investigation is, at this stage, perforce preliminary, it would appear that the population of South Africa may have been overestimated by as much as a million people, with about half of that number accounted for by significant overestimates of the white and Indian/ Asian populations, particularly<sup>19</sup>. Furthermore, this overall excess is concentrated in the age groups aged 50 and over, an age range unlikely to have been affected by substantial in-migration and the excess appears to be a little higher for men than women.

That said, we also appreciate that some proportional measures derived (such as most rates and ratios) using the Census 2022 data in both numerator and denominator may well be more robust, as the extent of errors in the numerator and denominator used to derive those proportional measures may well cancel out to some extent. However, where rates and ratios are derived using numerators from a source other than the census, but using the census data as a denominator, these could be subject to substantial error.

In the interim, we urge all users of the 2022 census data to exercise caution in their use and interpretation, especially for planning and implementation at district and municipal levels.

It would be inadvisable for population projections (including the official mid-year population estimates (MYPE) to base projections on the 2022 census without adjusting for the problems identified in this report. As past practice has been for Stats SA to base any projections (and reconciliations with past census estimates) on the most recent census, it will be necessary for an alternative 'unofficial', but at least coherent, set of population estimates to be produced – down to at least a district level – for projects such as the estimation of the National Burden of Disease Study.

#### 5.2. What next?

At this stage, we have raised more questions than we can answer. In part, this is because we remain greatly constrained by the limited data that has been released by Stats SA. Census data on fertility and mortality, and much of the data on migration, are still outstanding. The data on fertility and mortality is of particular importance given recent delays in releasing the regular annual data on registered births and, particularly, deaths.

<sup>19</sup> The number of people over age 50 estimated by the 2022 census exceeds the number from the projections by close around 985 000 and it is quite possible/likely that net immigration 20-39 was less than required to support the 2022 census estimates of the population in this age range.

The release of those data takes on a particular urgency, so as to better understand what has happened with Census 2022, and to be able to come to better, and betterinformed, conclusions about the Census 2022 results – not only because it may help with further understanding and (perhaps) additional recommended adjustments to the data (although our concerns about the PES remain, and may therefore impede any such exercise), but also because those insights and conclusions are essential to set in place mechanisms and corrective processes to ensure that the 2031 census (or any other future census) is not compromised. Ideally, we would have liked to engage with Statistics South Africa to rework or directly rehabilitate the data from the 2022 South African census. Unfortunately, in meetings in December 2023, they were of the view that the census numbers are what they arrived at following the required processes, and that any disagreement was up to 'analysts' to sort out.

Finally, there is an urgent need for alternative population estimates that may better describe the South African population in the mid-2020s than those in the Census 2022 data. Those estimates might then be used to better inform evidence-based resource allocation and planning in order to benefit the lives of all South Africans.

### REFERENCES

- Dorrington, Robert E. 1999. "Who was counted in? Some possible deficiencies with the 1996 South African Census results," Paper presented at Actuarial Society of South Africa Convention. Johannesburg, 2-3 November 1999.
- GBD 2021 Demographics Collaborators. 2024. "Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950–2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the Global Burden of Disease Study 2021", *The Lancet* <u>https://dx.doi.org/10.1016/S0140-6736(24)00476-8</u>
- Moultrie, Tom A., Rob E. Dorrington, Allan G. Hill, Kenneth Hill et al. (eds). 2013. Tools for Demographic Estimation. Paris: International Union for the Scientific Study of Population. Available at demographicestimation.iussp.org.
- National Research Council. 2004. The 2000 Census: Counting Under Adversity. Panel to Review the 2000 Census. Citro, Constance F., Daniel L. Cork and Janet L. Norwood (eds), Committee on National Statistics, Division of Behavioural and Social Sciences and Education. Washington DC: National Academies Press. <u>www.nap.edu/catalog/10907.html</u>.
- Parliamentary Monitoring Group. 2023. Briefing by Statistician-General on the census 2022 results; with Deputy Minister. Cape Town: Parliamentary Monitoring Group. <u>https://pmg.org.za/committee-meeting/37818/</u>. Accessed: 3 May 2024.
- Rogers, Andrei and L.J. Castro. 1981. Model Migration Schedules. Laxenburg, Austria: International Institute for Applied Systems Analysis. <u>http://webarchive.iiasa.ac.at/Admin/PUB/Documents/RR-81-030.pdf</u>.
- Statistics South Africa. 2012. Post-enumeration Survey. Report 03-01-46. Pretoria: Statistics South Africa. http://www.statssa.gov.za/census/census\_2011/census\_products/Census\_2011\_PES.pdf. Accessed: 30 March 2024.
- Statistics South Africa. 2018a. Census 2021 New Methodologies Test. 8 August 2018. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?page\_id=11389. Accessed: 30 January 2024.
- Statistics South Africa. 2018b. Updating geo-frame for Census. 6 November 2018. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?page\_id=11703. Accessed: 30 January 2024.
- Statistics South Africa. 2019. Stats SA begins process of questionnaire consultation for Census 2021. 24 May 2019. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=12161</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2020. South Africa in line for first-ever digital census. 10 August 2020. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=13537. Accessed: 30 January 2024.
- Statistics South Africa. 2021a. Car hire service providers needed for Census 2022 project. 2 December 2021. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=14983</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2021b. Census 2022 remote data collection, a safety measure against Omicron as South Africa's population count gets underway in 50 days. 14 December 2021. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15029</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2021c. Pilot Census to make use of remote data collection only. 9 July 2021. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=14505. Accessed: 30 January 2024.
- Statistics South Africa. 2021d. Registration for Census 2022 data collection is open. 1 December 2021. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=14943. Accessed: 30 January 2024.
- Statistics South Africa. 2021e. Stats SA extends closing date for Census 2022 Recruitment Drive to 31 January 2022. 29 December 2021. Pretoria: Statistics South Africa. <u>https://www.statssa.govza/?p=15073</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2021f. Stats SA goes into field for Pilot Census. 17 May 2021. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=14318. Accessed: 30 January 2024.
- Statistics South Africa. 2021g. Stats SA resumes face-to-face data collection for Pilot Census from 13 August. 11 August 2021. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=14576</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2021h. Stats SA to conduct Census 2022 in 100 Days. 25 October 2021. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=14866. Accessed: 30 January 2024.
- Statistics South Africa. 2021i. Stats SA urges households and institutions to register for Pilot Census. 24 June 2021. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p-14466</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022a. Car hire required for Census 2022 project. 7 February 2022. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=15137. Accessed: 30 January 2024.
- Statistics South Africa. 2022b. Census 2022 CAPI CATI Household Questionnaire. 25 November 2021. Pretoria: Statistics South Africa. https://census.statssa.gov.za/assets/documents/2022/Census\_2022\_CAPI\_CATI\_Household\_questionnaire.pdf. Accessed: 10 November 2023.
- Statistics South Africa. 2022c. Census 2022 data collection in Western Cape extended to 14 May. 28 April 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.govza/?p-15360</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022d. Census 2022 online questionnaire completion now available. 20 April 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15349</u>. Accessed: 30 January 2024.

- Statistics South Africa. 2022e. Census 2022: Statement from Chairperson of Council and Statistician-General. 22 February 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15192</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022f. Final call to Western Cape Residents: Census 2022 data collection will run until 31 May. 13 May 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15377</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022g. Mid-year population estimates, 2022. Report P0302. Pretoria: Statistics South Africa. https://www.statssa.gov.za/publications/P0302/P03022022.pdf.
- Statistics South Africa. 2022h. Online questionnaire completion for Census 2022 closing on 25 February 2022. 22 February 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15188</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022i. Post-enumeration Survey (PES) 2022 kicks off to measure Census 2022 coverage. 1 August 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15437</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022j. Post Enumeration Survey kicks off to measure Census 2022 coverage. 7 June 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15437</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2022k. Stats SA kicks off Census 2022 mop-up phase. 29 March 2022. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=15271. Accessed: 30 January 2024.
- Statistics South Africa. 2022l. Stats SA to undertake nationwide Census in eight days. 25 January 2022. Pretoria: Statistics South Africa. https://www.statssa.gov.za/?p=15097. Accessed: 30 January 2024.
- Statistics South Africa. 2022m. Urgent appeal to Western Cape households to get counted in Census 2022. 3 May 2022. Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/?p=15363</u>. Accessed: 30 January 2024.
- Statistics South Africa. 2023a. Annual Report 2022/23 (Book 1). Pretoria: Statistics South Africa. <u>https://www.statssa.gov.za/</u> <u>publications/AnnualReport/StatisticsSouthAfricaAnnualReport202223\_Book1.pdf</u>. Accessed: 12 February 2024.
- Statistics South Africa. 2023b. Census 2022 Municipal Fact Sheet. Report 03-01-82. <u>https://census.statssa.gov.za/assets/documents/2022/Census\_2022\_Municipal\_factsheet-Web.pdf</u>. Accessed: 30 March 2024.
- Statistics South Africa. 2023c. Census 2022 Statistical Release. Report P0301.4. Pretoria: Statistics South Africa. https://census.statssa. gov.za/assets/documents/2022/P03014\_Census\_2022\_Statistical\_Release.pdf. Accessed: 30 March 2024.
- Statistics South Africa. 2023d. How the Count was Done. Report 03-01-45. Pretoria: Statistics South Africa. <u>https://census.statssa.gov.</u> <u>za/assets/documents/2022/How\_the\_count\_was\_done2022.pdf</u>. Accessed: 30 March 2024.
- Statistics South Africa. 2023e. Post-Enumeration Survey Statistical Release. Report P0301.5. Pretoria: Statistics South Africa. https://census.statssa.gov.za/assets/documents/2022/P030152022.pdf, Accessed: 20 October 2023.
- Statistics South Africa. 2023f. Province [sic] at a Glance. Report 03-1-43. Pretoria: Statistics South Africa. https://census.statssa.gov.za/assets/documents/2022/Provinces\_at\_a\_Glance.pdf. Accessed: 5 November 2023.
- Statistics South Africa. 2023g. SA at a Glance. Pretoria: Statistics South Africa. https://census.statssa.gov.za/assets/documents/2022/Census2022\_SA\_at\_a\_Glance.pdf. Accessed: 11 October 2023.
- Statistics South Africa. 2024. Migration Profile Report for South Africa. Report 03-09-17. Pretoria: Statistics South Africa. https://www.statssa.gov.za/publications/03-09-17/03-09-172023.pdf. Accessed: https://www.statssa.gov.za/publications/03-09-172023.pdf.
- UN Statistics Division. 2008. Principles and Recommendations for Population and Housing Censuses v.2. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/STAT/SER.M/67/Rev2. http://unstats.un.org/unsd/publication/SeriesM/Seriesm\_67rev2e.pdf.
- UN Statistics Division. 2010. Post-Enumeration Surveys: Operational Guidelines Technical Report. New York: United Nations Statistics Division. <a href="http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manual\_PESen.pdf">http://unstats.un.org/unsd/demographic/standmeth/handbooks/Manual\_PESen.pdf</a>.
- United Nations Population Division. 2023. World Population Prospects 2022 [Data Sources]. New York: United Nations Population Division. <u>https://population.un.org/wpp/Download/Files/4\_Metadata/WPP2022\_F02\_METADATA.XLSX</u>. Accessed:30 January 2024
- US Bureau of the Census. 1985. Evaluating Censuses of Population and Housing. Statistical Training Document ISP-TR-5. Washington, D.C: US Bureau of the Census. <u>https://www.census.gov/content/dam/Census/library/working-papers/1985/adrm/rr85-24.pdf</u>.

38

### APPENDIX 1: TIMELINE OF EVENTS LEADING TO THE RELEASE OF THE 2022 CENSUS

#### Table 10: Timeline of important events in the preparation and conduct of Census 2022

DATE	DESCRIPTION	SOURCE*
First half of 2018	Preparation for Census 2021 begins	(2021b)
August 2018	Testing of new data collection methods	(2018a)
November 2018	Updating of geo-frame for Census 2021 begins	(2018b)
May-June 2019	Consultation on census content	(2019)
May 2020	Feedback given by UCT on 2020 Pilot questionnaire	
September 2020	Census Trial Run	(2020)
17 May 2021	Census Pilot, to 30 June 2021 – subsequently extended to 31 August 2021	(2021f, g)
25 October 2021	Announcement of census night of 2 February 2022	(2021h)
31 December 2021	Closing date for recruitment of field staff – subsequently extended to 31 January 2022	(2021e)
2-3 FEBRUARY 2022	CENSUS NIGHT	
25 February 2022	Planned completion of online enumeration	(2022h)
28 February 2022	(Planned) completion of enumeration	(2022 )
20 March 2022	Revised date for completion of face-to-face and telephonic enumeration	(2022e)
24 March 2022	Commencement of 'mop-up' phase, to 14 April 2022.	(2022k)
20 April 2022	Re-opening of online registration and enumeration nationally, to 30 April 2022	(2022d)
28 April 2022	Extension of data collection, including online, in the Western Cape, to 14 May 2022	(2022c)
13 May 2022	Extension of data collection, including online, in the Western Cape, to 31 May 2022	(2022f)
8 June 2022	Post-Enumeration Survey (PES) commences, to 26 July 2022	(2022j)
27 July 2022	Post-Enumeration Survey (PES) 'mop-up' activities, to 6 August 2022	(2022i)
10 OCTOBER 2023	RELEASE OF CENSUS 2022 RESULTS	

\*All references are to Statistics South Africa (date). Full references are provided at the end.



### APPENDIX 2: FURTHER ANALYSIS OF THE FOREIGN-BORN POPULATION IN THE 2011 AND 2022 CENSUSES

#### Sex profiles of international migrants

Using the data on country of birth, at the time of the 2011 census, 40.08% of the 2.2m people born outside South Africa were female; this proportion had increased slightly to 42.19% (of 2.4m) in the 2022 census. Figure 11 shows the reported numbers of foreign-born people (by age group and sex) reported in the two most recent censuses. While self-reported immigrant status may well be underreported (out of fear of xenophobic violence or the information being provided to the authorities), the data from the two censuses are barely consistent, showing significant increases in the male foreign-born population between ages 35 and 54, and in the female foreign-born population between ages 30 and 49.

The upwards shift in the age profiles of foreign-born people aged between 25 and 54 in the 2022 census is indicative of a generalised aging of that population over the intercensal period, a feature consistent with the very small increase in the reported foreign-born population between the 2011 and 2022 censuses. This observation, however, is inconsistent with the balance equation analysis presented in Section 2, and further amplifies our concern as to the quality of the 2022 census data.

The number of reported foreign-born males increased by just under 80 000 between the 2011 and 2022 censuses, while that of reported foreign-born females increased by nearly 140 000. These are net increases: some of the foreign-born population reported in 2011 may have either died, or left South Africa before the 2022 census, meaning that actual net intercensal immigration would have been higher, as shown in row [7] of Table 1. Nevertheless, the age profiles give cause for some concern, suggesting that the increase in the foreign-born population has occurred in age groups that one would not anticipate being those most likely to immigrate.



#### Figure 11: Numbers of foreign-born people reported in the 2011 and 2022 South African censuses, by age group and sex

#### Age, sex, by population group sex profiles of foreign-born persons in the 2011 and 2022 censuses

The foreign-born population, by sex and population group, identified in the 2011 and 2022 censuses is shown in Table 11.

These data indicate a number of perhaps-surprising trends:

- The African population reported being foreign-born increased by 26.5%, with the increase in women being (surprisingly) twice that of men.
- The coloured population reported being foreignborn fell by 35%, but the numbers are small and the classification 'coloured' is the vaguest of the four.
- The Indian/Asian population reported being foreignborn fell by 4%, with a decrease of 17% among women almost offset by a 1% increase among men. In aggregate, this is entirely inconsistent with the balance equation analysis presented in Section 2.
- The white population reported as foreign-born fell by over 70% (or nearly 280 000 people) over the intercensal period. While this may, to some degree, reflect emigration, the much greater reduction in the population of white foreign-born women (86%) relative to men (62%) suggests that there might be other problems in the data. In 2011, the sex ratio in the white foreign-born population was 97 (indicating roughly equal numbers of men and women); whereas

in 2022, there were 261 white foreign-born men for every 100 white foreign-born women.

- Those foreign-born who specified their population group as 'Other' increased by 67%, or nearly 94 000 people, the overwhelming proportion of these being women.
- The age profile of the foreign-born population by sex and population group identified in the 2011 and 2022 censuses is shown in Figure 12. From this, we can see that the data indicate:
- The age profile of the African foreign-born population has aged between the two censuses.
- Among the Coloured foreign-born population, for both men and women below the age of 20, and above age 50, the numbers are approximately stable. There has been a big drop in the numbers of foreign-born coloured men between 20 and 50, while that of foreign-born coloured women increased in that age group.
- A generalised ageing of the Indian foreign-born population among both men and women.
- A dramatic (and implausible) shift in the numbers and age structure of the foreign-born white population for both men and women.
- A completely different (and implausible) age-sex profile of those foreign-born who declared their population group as "Other". In part this might be attributable to the small numbers of people enumerated in this category, and to which incorrect weights have been applied.

	AFRICAN		AFRICAN COLOURED INDIAN		WHITE		OTHER		TOTAL			
	2011	2022	2011	2022	2011	2022	2011	2022	2011	2022	2011	2022
Foreign born	1,569,409	1,985,506	20,045	12,919	91,637	87,314	378,135	98,336	140,644	234,123	2,199,871	2,418,198
Sex ratio	157	136	113	120	220	269	97	261	235	94	150	137
% increase (2011-2022)												
		African	(	Coloured		Indian		White		Other		TOTAL
Total		26.5%		-35.5%		-4.7%		-74.0%		66.5%		9.9%
Female		38.1%		-37.7%		-17.3%		-85.8%		188.0%		15.7%
Male		19.1%		-33.7%		1.0%		-61.9%		14.7%		6.0%

#### Table 11: Foreign-born population by sex and population group, 2011 and 2022 censuses

Figure 12: Age profiles of the foreign-born population by sex and population group, 2011 and 2022 South African censuses















### APPENDIX 3: BACKGROUND TO POST-ENUMERATION SURVEYS (PESS), AND DETAILED CONCERNS REGARDING THE PES IN THE 2022 SOUTH AFRICAN CENSUS

This appendix is presented in two parts. The first is a brief exposition of how capture-recapture methods can be used to estimate the population missed in a census. The second sets out specific and detailed concerns with the results of the PES conducted after the 2022 South African census, and explains what implications these concerns may have for our understanding of the reliability and robustness of the census results.

#### A3.1 How to estimate who was missed in a census enumeration?

To simply illustrate how the process of estimating an undercount is arrived at, consider the question of how to estimate the population of fish living in an isolated lake, where the fish have no means of entering or exiting. To do so requires what is termed a 'capture-recapture' exercise: at a (first) point in time, investigators may go out onto the lake, and capture and tag all the fish they encounter in a fixed period of time.

A short while (perhaps a few days, or weeks) later, the investigators again head out onto the lake, and (in the same fixed period of time) count how many fish they now identify, by whether they had been tagged at the first visit (A) or not (B). We would also know from the first exercise how many fish tagged in the first expedition were NOT found in the second visit (C).

The most important assumptions of a capture-recapture exercise are that the two expeditions are independent of each other, and that the fish had no mechanism or reason to actively seek to avoid capture and tagging (or counting) in either visit, and that the time elapsed between the two visits was sufficiently short to be able to discount the effects of changes in the population caused by birth, death or migration.

If these assumptions hold, the total population of fish can be estimated algebraically using the data presented as a contingency table: Table 12: Structure of a contingency table used to estimate total population size in a capture-recapture exercise

		Captured at the second visit		
		YES	NO	
Captured and tagged	YES	A	С	
at the first visit	NO	В	D	

The total population of fish would be given by the sum of the observed quantities A, B, and C (as described above), together with the (as yet) unestimated quantity, D. We estimate D on the basis of assuming that the probabilities of being caught on each visit are independent, and thus the ratios of captured fish at the second visit being caught to not being caught at the first visit (i.e. A/B) should be the same as that of not being caught at the second visit, relative to the first visit (i.e. C/D).

If this assumption holds, then we can derive an estimate of the size of D, those missed in both the census and the PES:

$$\frac{A}{B} = \frac{C}{D} > D = \frac{C.B}{A}$$

So, if on our first visit we had tagged 279 fish, and on the second visit we found 207 fish with tags (A), and 58 fish with no tags (B). and knowing that we did not find (279-207=) 72 fish from the first visit at the second visit (C), we estimate the number of missed fish, D, to be (72 x 58)/207 = 20 (rounded), and hence the total population of fish to be 207+58+72+20 = 357. The proportion of all fish counted in the first visit is 279/357 = 78.2%.

The same logic applies to an application to a census except that people are not uniformly distributed, nor do they move about randomly. Thus, in this case the first 'sample' is replaced by the attempt to enumerate the whole population (in dwelling units<sup>20</sup>), and the second sample has to be chosen in such a way as to reflect

 $^{\rm 20}\,$  No attempt is made to estimate under- or over-count of those not in dwelling units.

the whole population (i.e. proportional to listed dwelling units). In addition, unlike fish which may be tagged, this is clearly unethical with human subjects, and so much more complex matching exercises have to be undertaken to ensure that the same person has indeed been identified in both the census and the PES. We discuss our concerns about the matching exercise in Section A3.3.

# A3.2 Specific concerns about the 2022 South African PES

#### Sampling concerns

The description of the sampling methodology in the PES report (Statistics South Africa 2023e) is somewhat opaque. The 2022 PES report describes the PES sample size as being determined by a desired precision level of the estimates. This is not unreasonable. However, the description then notes that the determination of the sample size was based on the preliminary net undercount rate from 2011, "allowing for a 15% reduction in the coefficients of variation for that indicator for PES 2022" (Statistics South Africa 2023e: 50).

This suggests that the expectation was that the undercount would perhaps be a little better (if reducing the coefficients of variation by 15% was due to reducing the standard error by the same amount in relative terms) or somewhat worse (if the reduction in the CV by 15% was due to increasing the mean value of the undercount rate by 15% in relative terms). It is not clear which element was assumed to be modified; or, if the latter, it is unclear how the undercount would have been known at the time of planning the 2022 PES. Regardless, with an undercount rate more than twice the 2011 undercount rate, this suggests that the sample size of the 2022 Census PES turned out to be far too small to reliably adjust for the undercount that was encountered.

The How the Count was Done report issued by Statistics South Africa (Statistics South Africa 2023d) provides some information on the geography of the country at an Enumeration Area (EA)<sup>21</sup> level. However, the information is somewhat hard to interpret. At page 13, the report records that the 2011 EA boundaries were not altered for the 2022 Census, but certain EAs that grew over the intercensal period were sub-divided: "The sub-divided 2011 EAs were termed sub-EAs".

The number of EAs (presumably including sub-EAs) in 2022 was 133 694, an increase of 30 118 from the 103 576

EAs demarcated in the 2011 census, implying (assuming each EA was sub-divided once only) that some 30% of 2011 EAs were sub-divided for the purposes of sampling in 2022.

The PES sample for the 2022 survey was drawn from 840 "sub-EAs", meaning that the PES sample fraction at an EA level in 2022 was 0.628% (Statistics South Africa 2023e)<sup>22</sup>. In 2011, the PES consisted of 608 EAs, giving a sampling fraction at an EA level of 0.587% (Statistics South Africa 2012). These are broadly comparable, with a slightly greater sampling fraction in the more recent census.

# Concerns about the uncertainty surrounding the estimate of the undercount

It is important to re-emphasise that the output from many censuses (South Africa's included) are estimates rather than counts of the population. This is so, because, the counts, understandably, are adjusted for an estimate of the extent to which the census was under- (or, less likely, over-) counted. The extent to which the census count is an under-/over-estimate is determined by the Postenumeration Survey (PES).

Where the undercount is low the impact on the estimate of the population is not material (and is often ignored) but where it is large, it can be very material, in which case it is important to have some sense of the uncertainty about the extent and effects of the adjustments applied.

Our key concern (and indeed our principal concern regarding the PES) in this regard is illustrated by comparison of two tables from each of the 2011 and 2022 PES reports.

The first comparison (Table 13) is of the estimates of undercounts and uncertainty about those estimates derived from the results of the PES in 2011 and 2022 From this we see that the preliminary estimates of the net undercount from the PES were about twice as high in 2022 than they were in 2011. However, what is of greater importance is that the estimates of uncertainty about these estimates (the standard error (SE)) in the 2022 PES are, as might be expected, markedly higher than those of the 2011 PES. Based on the PES, the 95% confidence interval for the undercount in 2011 was narrow (14.34%; 14.86%). In 2022, the equivalent interval was, as would be expected, wider (27.99%; 31.21%).

<sup>&</sup>lt;sup>21</sup> An Enumeration Area (EA) is defined as "the smallest geographical unit (piece of land) into which the country is divided" for census enumeration purposes (Statistics South Africa 2012: 72)(Statistics South Africa 2012).

<sup>&</sup>lt;sup>22</sup> Assuming that this was drawn from the 133 000 EAs, and not only the sub-divided EAs, and that the terminology in the 2022 PES report is inaccurate.

Table 13: Estimates of the undercount (UC), their standard errors (SE), South Africa and its provinces as estimated directly from the postenumeration survey<sup>23</sup>, 2011 and 2022

	20	11	20	22
	EST. UC (%)	SE (%)	EST. UC (%)	SE (%)
WC	18.5	0.542	35.56	1.56
EC	12.9	0.196	30.56	0.59
NC	13.4	0.318	29.11	2.27
FS	10.1	0.362	18.52	2.62
KZN	16.7	0.379	31.19	1.08
NW	14.9	0.532	17.92	5.39
GT	14.7	0.174	30.79	2.07
MP	15.5	0.473	34.00	4.51
LP	10.0	0.135	23.47	1.02
ZA	14.6	0.132	29.6	0.82

Source: 2011 PES Report (Table 10); 2022 PES Report (Table 1)

At a population group level, unfortunately, the two PES reports are not any help in estimating the likely uncertainty about the estimates of the Indian or white populations. While the 2022 report does provide estimates, namely standard errors (of the estimates of undercount) of 0.82% for the Indian population and 0.42% for the white population (both of which are clearly implausible given the extent of the undercount), the 2011 PES report does not, which makes it impossible for users to gauge the change, and indeed possible error, in these estimates.

Table 14 below shows the change in the estimated undercount from the preliminary to final estimates in the 2022 census. The methodology of how the final estimates of the undercount were determined are not described (and an equivalent table was not provided in 2011). From the table, we note the (smallish) increase in the final estimate of the 2022 undercount (from 29.6% to 31.1%) but the much smaller standard errors around the final estimates of the undercount, which indicate that the final undercount was 95% certain to lie within the range of 31.06 ± 0.02%.

## Table 14: Estimates of the preliminary and final undercount (UC), their standard errors (SE), and coefficients of variation (CV), South Africa and its provinces, 2022 Census

	2022 (PRELIM)				2022 (FINAL)	RATIO OF CVS	
	EST. UC (%)	SE (%)	cv	EST. UC (%)	SE (%)	CV	(PRELIM:FINAL, %)
WC	35.56	1.55	0.0436	35.58	0.02	0.00056	1.3
EC	30.56	0.59	0.0193	32.40	0.01	0.00031	1.6
NC	29.11	2.27	0.0780	33.41	0.03	0.00090	1.2
FS	18.52	2.62	0.1415	20.95	0.02	0.00095	0.7
KZN	31.19	1.08	0.0346	34.67	0.01	0.00029	0.8
NW	17.92	5.39	0.3008	23.33	0.02	0.00086	0.3
GT	30.89	2.07	0.0670	30.79	0.01	0.00032	0.5
MP	34.00	4.51	0.1326	35.26	0.02	0.00057	0.4
LP	23.77	1.02	0.0429	23.53	0.01	0.00042	1.0
ZA	29.6	0.82	0.0277	31.06	0.01	0.00032	1.2

Source: 2022 PES Report (Tables 1 and 3)

<sup>23</sup> An Enumeration Area (EA) is defined as "the smallest geographical unit (piece of land) into which the country is divided" for census enumeration purposes (Statistics South Africa 2012: 72)(Statistics South Africa 2012). This huge reduction in the uncertainty of the percentage undercount between the preliminary and final estimates of the undercount finds its way into the estimates of the uncertainty around the population count.

#### Concerns about the uncertainty surrounding the population estimates after applying the results of the 2022 PES to the 2022 census enumeration

Table 15 compares the final estimates of the population, by province and nationally, from the 2011 and 2022 censuses, after the adjustment for the undercount has been applied to the enumerated population. In the 2011 census, the final population size was estimated to be 51.77 million, with a 95% confidence interval of 1.955 million people either side of that estimate. By contrast, in the 2022 census, the final estimated population size was 62.03 million, with a 95% confidence interval some ten times smaller (allowing for the growth of the population), suggesting that the true population size is 95% certain to lie within 235 000 of the final estimated population size.

Unfortunately, the data for 2022 presented in the table in the 2022 PES report are rounded to the nearest 10,000, so complete precision is impossible. Nevertheless, we can see that the coefficient of variation (CV; the ratio of the standard error to the central estimate) nationally in 2022 is one-tenth of what it was in 2011; and that only in the Western Cape is the CV higher in 2022 than it was in 2011, and the CV in six of the nine provinces in 2022 is less than half of what it was in 2011.

These results make little sense given the extent of the undercount, the significantly greater uncertainty surrounding the preliminary estimates of the undercount in 2022 relative to 2011 and the likely underpowering of the PES.

The combined effects of the changes in the coefficients of variation of the preliminary estimates of the undercount, and those of the final population estimates are shown in Figure 13: the coefficients of variation of the net (preliminary) undercounts are typically 3-5 times greater in 2022 than they were in 2011; while the coefficients of variation of the final population estimates in 2022 are typically 10-40 percent of those from 2011. The complete absence of any correlation between the two series shown amplifies our concern that the standard errors of the population estimates in 2022 have been miscalculated.

In short, it is difficult to understand how it can be asserted that one can be 95% certain that the population in 2022 lies within plus or minus 235,000 of the estimated population, while in 2011 the range for that certainty is 1.955 million, when in 2011 – with similar PES sampling fractions – the undercount was half that in 2022.

Table 15: Estimates of the total population, their standard errors (SE), and coefficients of variation (CV), South Africa and its provinces, 2011 and 2022

	POPULATION	SE	CV	POPULATION	SE	CV	CV22/CV11 (%)
WC	5,822,734	40,830	0.007	7,430,000	70,000	0.009	134.4
EC	6,562,053	125,810	0.019	7,230,000	60,000	0.008	43.3
NC	1,145,861	82,466	0.072	1,350,000	30,000	0.022	30.9
FS	2,745,590	117,567	0.043	2,960,000	40,000	0.014	31.6
KZN	10,267,300	109,994	0.011	12,390,000	100,000	0.008	75.3
NW	3,509,953	166,754	0.048	3,800,000	40,000	0.011	22.2
GT	12,272,263	106,023	0.009	15,120,000	90,000	0.006	68.9
MP	4,039,939	219,299	0.054	5,160,000	60,000	0.012	21.4
LP	5,404,868	251,244	0.046	6,570,000	60,000	0.009	19.6
ZA	51,770,560	997,560	0.019	62,030,000	120,000	0.002	10.0

Source: 2011 PES Report (Table 14); 2022 PES Report (Table 5)





#### A3.3 Matching agreements

In 2011, the PES report (at Table 5, for example) provided detailed information on the criteria used to establish whether a match by age of people enumerated in both the 2011 census and the 2011 PES was established. This information is not provided with the 2022 PES report.

However, it would appear that either the matching criteria (for many variables) were strengthened in 2022 relative to 2011, or that the quality of the two exercises was markedly different, judging from a comparison of the two consistently-defined metrics provided in the PES reports, as shown in Table 16<sup>24</sup>.

<sup>24</sup> The third is the gross difference rate, the complement of the second.

VARIABLE	CENSUS	INDEX OF INCONSISTENCY	RATE OF AGREEMENT
Age	2011	7.4%	93.72%
	2022	13.61%	86.92%
Sex	2011	2.5%	98.77%
	2022	8.19%	95.76%
Population group	2011	9.9%	96.31%
	2022	7.52%	98.23%

#### Table 16: Measures of matching from the PES in 2011 and 2021 by selected variables

Source: 2011 PES Report (Table 33); 2022 PES Report (Tables 25, 22, and 28)

For two of the three crucial variables that might affect the determination of the adjustment factors in the PES, the 2022 showed significantly higher indices of inconsistency (and lower levels of agreement). This is concerning, as it indicates that while not extreme, the matching process delivered markedly inferior results in 2022 than it did in 2011, especially with regard to sex and age.

In turn, the deterioration in the indices of inconsistency for two of the three crucial matching indicators further calls into question the relative reliability of the census estimates in 2022 relative to those from the 2011 census, even as the confidence interval around the numbers has apparently reduced so much.

### APPENDIX 4: DETAILED DESCRIPTION OF EVENTS AND PROCESSES THAT MAY HAVE CONTRIBUTED TO THE PROBLEMS WITH THE 2022 CENSUS

Apart from the usefulness of documenting the process and hurdles encountered in completing the census in 2022, it is important to identify some of the lessons to be learnt so as not to repeat the same mistakes in future.

While this section is largely speculative, we can draw on the many media statements and press releases issued by Stats SA relating to Census 2022 issued from 2018 onwards to make inferences about what might have contributed to the problems identified in this report. Here we consider separately factors external to the Census 2022 exercise, as well as those emanating from within Stats SA. In many senses, of course, this distinction is artificial: exogenous events will have had an impact on internal processes, but enforcing the separation is of assistance in identifying those aspects that might have been outside of Stats SA's control, and to assess how they might have impacted on those that were within their control.

#### A4.1 A major exogenous event: Covid-19

If the challenges identified in previous South African censuses, resulting in the reported estimated undercounts, were severe, it is an understatement to say that the 2022 South African Census ('the 2022 Census') was conducted at an inauspicious time.

In keeping with the timing of the three post-apartheid censuses of 1996, 2001, and 2011, each of which had a specified 'census-night' of the night of 10-11 October, Stats SA's intention was to conduct the most recent census with a reference date of 10-11 October 2021. While not crucial, there is a demographic convenience in adopting a consistent census date, especially if they are spaced five or ten years apart. If this is done, identifying shifts in the population age structure is made easier: all individuals at a single year of age, or in fiveyear age groups, would be an exact number of years older at the more recent census than they were at the preceding census. In keeping with best practice, preparations for this census had commenced in 2018, many years before the census date: listing and geo-mapping of households, questionnaire design, and logistics preparation all require substantial lead times. Press-releases released by Stats SA in 2018 and 2019 (see for example, Statistics South Africa (2018a, 2019)) make repeated reference to a census to be conducted in 2021.

Preparations for the 2021 census, however, were substantially affected by the Covid-19 pandemic. The first case of SARS-CoV-2 was identified in early March 2020, and the first official death from the virus was reported at the end of March 2020. The government response to the pandemic saw the country locked-down entirely (with a 24-hour curfew) from the last week of March 2020 to the end of April 2020, with partial restrictions continuing to the end of May 2020. Periodic partial lockdowns were imposed with each of the major waves of infection occurring through early 2022, and many – if not most – organisations only saw a substantial process of 'returning to work' (as opposed to remote working) in 2022.

The weekly proportions of tests returning positive for SARS-CoV-2 in South Africa as monitored by the South African National Institute for Communicable Diseases (NICD) from 2020 to mid-2023 is shown in Figure 14. Distinct waves of infection are evident, with the census date (shown by the red vertical line) falling between the fourth and fifth waves.

The Omicron variant of SARS-CoV-2 had been identified in late October 2021, with the proportion testing positive peaking in late December 2021; an ill-timed occurrence given the 2 February 2022 census date. A fifth wave of infections hit as Stats SA was attempting to complete the data collection.

While we do not have detailed information as to what working arrangements were at Stats SA from March 2020 to February 2022, it is likely that those arrangements disrupted planning for the census at every stage and every level in that period.



Figure 14: Weekly proportions testing positive for SARS-CoV-2 in South Africa, 2020-2023

Note: The census date (2 February 2022) is indicated by the red vertical line

#### A4.2 Endogenous events

While Covid-19 was almost certainly the essential underlying contributor to the problems observed in the census, there were specific aspects that were particularly impacted by the disruptions occasioned by the pandemic.

To our mind, these impacts were of such a nature that it might be considered foolhardy in the extreme for Statistics South Africa to have even attempted to conduct the census in the form, and when, they did, as described below.

#### New modes of data collection

From the outset, Census 2021 (as it was then described) was heralded as implementing a major shift in the manner in which census data was to be collected. Previous censuses had relied on door-to-door paper-based enumeration by teams of fieldworkers. Collecting the data in this fashion is certainly time-consuming and error-prone: in 1996 the census forms had to be manually captured by means of double-entry processes (meaning each form was captured de novo twice each). Manual

capture significantly slowed the release of the results from that census. In 2001 and 2011, the same mode of data collection was employed, but optical scanning was used to produce the digital census data. This was not without its own problems (in 2001, the paper on which the census questionnaires were printed was too thin, requiring the urgent procurement of 'blue-light' scanners to avoid inadvertent capture of both sides of the forms in a single image), and scanning often struggled with optical character recognition – for example confusing badly-written 1s with 7s, and vice versa.

Direct-to-digital data collection<sup>25</sup> avoids these pitfalls and problems, but introduces others. First, a tri-modal data collection exercise is more complex, having to ensure that web, telephone, and personal interfaces are equivalently and consistently programmed; field staff need to be trained differentially to perform CATI and CAPI using technology that may be unfamiliar to both trainers and trainees; and requiring server-side data handling and management protocols for a high-volume data collection exercise. Equally, back-office functions

<sup>25</sup> Direct-to-digital data collection routinely takes three forms, making use of computer-assisted web, telephone, and personal interviewing. These are usually referred to by their acronyms as CAWI, CATI, and CAPI respectively. Even face-to-face interviewing (i.e. CAPI) is conducted as a direct-to-digital exercise, with responses being recorded by the enumerator on a digital device (notebook; PDA; tablet) and automatically uploaded to a server. and support systems to cope with direct-to-digital data collection are fundamentally different from those required with paper-first data collection exercises.

In August 2018, at the start of the census project, Stats SA flagged that it was testing new methodologies (Statistics South Africa 2018a) for the census in 2021. A common thread runs through many of the releases that appear to give particular prominence to CAWI – where individuals would log-in directly to a Stats SA server, and complete the census questionnaire online, with minimal fieldworker interface.

The technological innovations were, it would seem, first spurred by the need to move aw ay from paper-based data collection, and then amplified by the deleterious impact of SARS-CoV-2 on conventional fieldwork exercises.

Indeed, as the Covid epidemic unfolded, with repeated waves of infections associated with SARS CoV-2 at roughly six-month intervals, it is understandable that greater hope was placed in the ability to collect data remotely – i.e. through CAWI or CATI. Indeed, even as the census date neared, Stats SA was still touting how the census would "showcase Stats SA's new technological advances as it leaps into a new era of digital data collection." (Statistics South Africa 2022l).

#### The role played by National Treasury

The timing of the census, and the setting of the census date as 2 February 2022 is important. Once the census date had been announced, the initial indications were that the plan was to complete the data collection by the end of February 2022 (Statistics South Africa 2022h, U. Stats SA had been allocated funds to run a census, planned for October 2021, in the government financial year running from 1 April 2021 to 31 March 2022. The Statistics South Africa annual report for 2022/2023 noted:

"Census 2022 has been impacted by various challenges; therefore, the project's timelines were revised to the 2022/23 financial year. As a result of the Census 2022 extension into the 2022/23 financial year, Stats SA overspent on Programme 6 by R807,987 million after applying virements, thus increasing the unauthorised expenditure balance to R980,140 million.

A rollover and an unforeseeable and unavoidable expenditure funding request for R283,324 million and R737,155 million respectively to cater for Census 2022 extended activities was submitted to National Treasury. A roll-over of R194 million was approved, which is R89 million less than the requested amount. The unforeseeable and unavoidable expenditure funding request was, however, not approved resulting in the project being underfunded" (Statistics South Africa 2023a: 18)

From this it is clear that the National Treasury chose to take a particularly hard-line with Stats SA regarding the possibility of rolling over census-related expenditure from the 2021/22 to the 2022/23 financial year, and denied that request.

While prudential management of government funds is to be strongly encouraged, it is not beyond the realm of improbability that the National Treasury decision to abide by its regulations compelled Stats SA to rush the 2022 Census process so that most of it might be complete by the end of the 2021/22 financial year. Informally, we have been told that Stats SA was informed that they had to "use it or lose it" – if funds were not expended in 2021/22, funds would not be rolled over to 2022/23, possibly leaving the census unfunded for at least another decade.

If this is so, it represents a significant 'own-goal' for both the National Treasury and Stats SA: inflexibility regarding the funding of a project of this significance – we draw attention again to the description of the complexity of the US 2020 Census quoted in opening of the Introduction – might have resulted in the active waste of public funds.

A full public accounting of the decisions in this regard would be useful to help learn lessons for the future funding of censuses.

Had they been consulted, most demographers would have strongly motivated for the census to be deferred to 10 October 2022, or even 10 October 2023, to ensure that processes and implementation were not rushed. The already-severe impact of the Covid-19 pandemic on organisational processes, and the National Treasury decision that compelled Stats SA to expend the census budget in the 2021/22 financial year, had a number of further knock-on effects.

# Finalisation of census instruments and testing

Nearly half a year into the South African epidemic, in a video message to South Africans to make themselves available to participate in the 2021 census Trial Run in September and October 2020, the organisation's head made only passing reference to the pandemic, noting that the Trial Run would not be conducted in person, "as we have challenges of Covid-19" (Statistics South Africa 2020).

It is significant, too, that the press releases in this period emphasised the role that remote data collection<sup>26</sup> would play in the conduct of the census.

However, it would seem - given the timing of announcements relating to the census - that the National Treasury decision to force the census to be conducted before the end of the 2021/22 financial year emerged fairly late in the census exercise. Until May 2021, when Stats SA launched the Pilot Census (Statistics South Africa 2021f), all Stats SA communications regarding the census indicated that it was to be held some time in 2021. However, the timing of the Pilot Census made it highly implausible that the census would be conducted in the 2021 calendar year. Registration to participate in the Pilot Census was to run from mid-May to the end of June 2021, with data collection happening between 3 and 28 July 2021 and subsequently extended to 31 August 2021. Given those dates (and the time required to process and evaluate the data, and rectify any identified shortcomings), it was implausible that the full census would be conducted in 2021. Importantly, that release made no mention even of the year that the full census would be conducted, referring instead to "the upcoming census".

A further call for individuals and households to register for the Pilot Census was issued on 24 June 2021 (Statistics South Africa 2021i), 6 days before the closing date for registration. This noted (firstly) that late registrants would not have the option of CATI or CAWI interviews, but would be visited by a Stats SA fieldworker; and (secondly) informed that "the date for the main census will be announced in due course", implying that the census date had yet to be fixed. At this time, South Africa was entering its third wave of infections and deaths from Covid – a wave that lasted from June to late September 2021. As a consequence, all CAPI data collection was suspended on 9 July 2021, and the collection period extended to 4 August 2021 (Statistics South Africa 2021c). Barely a month later, on 11 August 2021 and still in the middle of the third wave of infections, Stats SA announced that face-to-face data collection for the Pilot Census would resume from 13 – 31 August 2021 in seven provinces, with fieldwork in KwaZulu-Natal and the Western Cape to resume at an unspecified later date (Statistics South Africa 2021g).

It was only on 25 October 2021, exactly 100 days before the census date, that the first public indication was given that the census was to be conducted on the night of 2-3 February 2022 (Statistics South Africa 2021h). Thus, it is most probable that the National Treasury ruling was finalised some time between May and October 2021, and probably closer to October than to May. If this was indeed the case, Stats SA probably had less than six months to finalise preparations for the census.

On 14 December 2021, 50 days from the enumeration night, the 2022 Census Project Director was quoted in a media release (Statistics South Africa 2021b) as saying "Census 2022 has been the most challenging population count that we have encountered in the history of our census-taking post-democracy. We introduced the new methods of data collection, tested them in the midst of alternating COVID -19 alert levels to which we had to align by adjusting our methodologies from time to time. The journey started in 2018 and we will arrive at our destination in the coming 50 Days" (emphasis added).

This is hardly the description of an orderly preparation for a census, and the rushed timeframe described almost certainly contributed to the problems with the census.

#### Recruitment and training of field staff

The same press release that announced the census date (Statistics South Africa 2021h) is also noteworthy for the fact that it reveals that recruitment of 165,000 temporary field staff was to begin on 25 October 2021 and would continue until 31 December 2021 (a mere 33 days before the census).

<sup>26</sup> Using CAWI and CATI methods. In contrast, CAPI is digital collection of data by fieldworkers using portable devices in a face-to-face interview.

On 29 December 2021, there was a further announcement to the effect that recruitment of fieldworkers, which was scheduled to end two days later, was to be extended to 31 January 2022 (that is, only 2 days before the census day) "to give the 2021 matriculants ... a chance to apply for the available work opportunities". Doing so, it was claimed, would "not affect census data collection, and fieldwork will commence on Census Night, 2 February 2022, as scheduled" (Statistics South Africa 2021e).

A press release dated 25 January 2022, 8 days before the census, recorded that "National and Provincial training of permanent staff concluded by 13 January 2022 while virtual district training across the country commenced last week 18 January" (Statistics South Africa 2022).

The Omicron wave of SARS-CoV-2 saw the proportion testing positive for the virus peaking in late December 2021/early January 2022 – in the middle of the fieldworker recruitment and training exercise. It is unclear how the training occurred, but the experiences of those who had to teach during the pandemic calls into question the efficacy of remote teaching and learning, especially when training materials were not designed with that in mind. In our own experience, student participation and attentiveness were significantly attenuated by the shift to remote teaching.

Furthermore, it defies understanding that even with a cascading model of training (trainers training the trainers), how it would have been possible to effectively train and perform adequate quality assurance on 165 000 newly-recruited fieldworkers in such a short span of time.

Accordingly, systematic failures of training of the field staff cannot be ruled out as a major cause of the problems identified with the census.

#### Public communication

Public communication and media outreach and coverage was poor. One hundred days before the census, on 25 October 2021, Stats SA announced that they would "roll-out a public awareness campaign" (Statistics South Africa 2021h). The media strategy adopted by Stats SA to ensure national awareness of the census was limited. Certainly, it was possible (not regularly reading a newspaper, or listening to the radio, or only watching satellite TV) to entirely miss the build-up to the census. As can be seen from the media releases from Stats SA, outreach and partnership building with the mainstream media appeared to be desultory.

On 1 December 2021, two months from the commencement of fieldwork for the census, the public was informed (by press release) that registration for enumeration was open (in an untried approach, the implication was that the public was requested to pre-register for enumeration, and had to select up-front how they wished to be enumerated (by CAPI, CATI, or CAWI)) (Statistics South Africa 2021d).

Two weeks later, on 14 December 2021 – and fifty days from the scheduled start of the census – the impact of the Omicron wave was noted. Members of the public were urged to register online for "remote data collection" as a "safety measure against Omicron" (Statistics South Africa 2021b). The same release also revealed that fieldwork was expected to be conducted "from 3 to 28 February" 2022. Again, the timing of the fieldwork, leaving exactly a month before the end of the financial year, was the consequence of having to ensure that the census was substantially completed in that financial year.

#### Logistics

The announcement of the census date on 25 October 2021 also revealed that mapping of enumerator areas was yet to be completed: "We will ...dissect the country into enumeration areas" (Statistics South Africa 2021h). The use of a future tense, particularly, suggested that the mapping and demarcation of enumerator areas was, at that date, still incomplete.

On 2 December 2021, an urgent call was put out for providers of a required fleet of some 6,000 vehicles for a period "between January and July 2022" in order to conduct the census (Statistics South Africa 2021a). By the time fieldwork had commenced, the organisation was still seeking 500 vehicles for the same period (Statistics South Africa 2022a). Eight days before the start of enumeration, on 25 January 2022, Stats SA reported that "[d]istribution of working tools i.e. tablets and vehicles continues in all areas" (Statistics South Africa 2022). In other words, it is not clear that – even at that very late stage –all field staff were fully equipped and ready to enter the field. Anecdotal evidence bears this out, with informants reporting that hardware (digital devices) was not standardised and did not all operate in the same way.

#### **Fieldwork operations**

After fieldwork began, it was clear that there were manifold problems with data collection using the new data collection methodologies. Some of these may be attributable to the desultory media campaign already mentioned. Thus, by 22 February (a week before the planned ending of the fieldwork exercise, it was reported that only a quarter of the 400 000 households (out of a total number of 17 million households) that had registered for CAWI enumeration had actually completed the exercise (Statistics South Africa 2022h). The fact that enumeration by CAWI required pre-registration, as well as the limited media reach into the populations most likely to prefer CAWI enumeration, no doubt contributed to this very low registration and response rate.

But even CAPI (i.e. face-to-face) enumeration faced problems – a joint statement from the chair of the Statistics Council and the Statistician-General on 22 February 2022 alluded to issues with fieldwork as well as technical glitches with the uploading of data to the servers: "we take full responsibility for Census 2022 from what we can control – enumeration, systems, quality control and so on. Where these faltered, we have repaired them. When the platform could not handle the volume of responses, within 72 hours we had transferred it to the cloud. We hear you and we are doing all we can to help you" (Statistics South Africa 2022e). By this date, it was revealed that less than 1.5 million households (barely 10 per cent of the total estimated number of households) had been interviewed using CAPI. The same announcement informed that CATI and CAPI data collection would be extended to 20 March 2022. By this date, the enumeration had reached only 12 million households, having completed more than 70% of estimated households in only three provinces (Limpopo, Mpumalanga and the Free State). The Western Cape lagged far behind with only 30% estimated completeness by the (revised) closing date for fieldwork (Statistics South Africa 2022k).

Again, such statements suggest that the fieldwork had not proceeded smoothly, pointing to possible deficiencies in stress-testing systems, as well as operational management of the fieldwork exercise.

The failure to timeously complete the enumeration in the initially-planned-for timeframes resulted in still further extensions to the fieldwork exercise. In announcing the start of the 'mop-up' phase of the census enumeration on 29 March 2022, Stats SA said that this would last to 14 April 2022, and that while CAPI and CATI enumeration would continue, online registration would not re-open (Statistics South Africa 2022k). This decision was then reversed on 20 April 2022 (after the previously declared closing date for the 'mop-up' operation), when CAWI enumeration was re-opened for a period of 10 days, to the end of April 2022 (Statistics South Africa 2022d).

While enumeration in KwaZulu-Natal was affected by bad weather, problems in the Western Cape persisted with only around 60% of households having been counted in that province by 3 May 2022, compared to a reported 88% in other provinces (Statistics South Africa 2022m), resulting in successive extensions to the fieldwork in that province, including allowing CAWI enumeration, first until 14 May 2022, and then until 31 May 2022 (Statistics South Africa 2022c, f).

#### **PES** operations

The repeated delays and extensions of the main fieldwork exercise saw the PES operations being delayed too. PES operations eventually began in June 2022, and this exercise was further extended, with operations first expected to end on 26 July 2022, but finally running until 6 August 2022 (Statistics South Africa 2022i, j).

PO Box 19070 7505 Tygerberg, South Africa Enquiries: Tel: +27 21 938 0911 Email: info@mrc.ac.za

#### www.samrc.ac.za