

**AUTHORS:**

Nthabiseng Mkhatha<sup>1</sup>  
George Djolov<sup>1</sup>  
Diego Iturralde<sup>1</sup>

**AFFILIATION:**

<sup>1</sup>Statistics South Africa, Pretoria,  
South Africa

**CORRESPONDENCE TO:**

Diego Iturralde

**EMAIL:**

Diegol@statssa.gov.za

**HOW TO CITE:**

Mkhatha N, Djolov G, Iturralde D.  
Response to Moultrie and Dorrington  
(2024): 'Problems and concerns  
with the 2022 South African census'.  
S Afr J Sci. 2024;120(11/12), Art.  
#20465. <https://doi.org/10.17159/sajs.2024/20465>

**ARTICLE INCLUDES:**

- Peer review
- Supplementary material

**KEYWORDS:**

census, South Africa, population,  
housing, digital census

**PUBLISHED:**

04 December 2024

# Response to Moultrie and Dorrington (2024): 'Problems and concerns with the 2022 South African census'

Statistics South Africa welcomes the opportunity to comment on the Commentary entitled 'Problems and concerns with the 2022 South African census'.<sup>1</sup> The population and housing census (henceforth referred to as 'Census') takes place in South Africa every 10 years and represents a rich source of statistical information that is designed to guide planning and policy development as well as to guide sampling design for the next inter-censal period. Contrary to previous censuses in South Africa, the 2022 Census<sup>2</sup> was South Africa's first digital census. A multi-modal approach was taken by collecting data using Computer Assisted Web Interviewing (CAWI), Computer Assisted Telephone Interviewing (CATI) as well as in person with a digital data collection instrument, Computer Assisted Personal Interviewing (CAPI), and data collection hinged on a geographic digital frame. Whilst all censuses were de facto, including Census 2022, the 2022 Census enumeration period extended over a 4-month period from February to May 2022.

Stats SA, and by extension the Census, was not spared from the impact of the global COVID-19 pandemic. The planning of this Census was severely disrupted by the pandemic, which impacted not only the data collection phase from February 2022 but also the preparatory phases when training and the pilot census were meant to take place. During this preparatory period, the much-publicised KwaZulu-Natal and Gauteng social unrest also took place during the winter of 2021, which restricted the training and pilot related data collection that could take place at that time.

The pilot was eventually executed in only seven provinces. Training reverted from face-to-face to online mode as a result of the COVID-19 restrictions, and, as such, implications for the quality of training were inevitable. Indeed, the training for the pilot had to be suspended in December 2020, when, despite precautionary measures, delegates at a training venue tested positive for COVID-19, and the pilot, due to start in February 2021, was postponed. From a human resource perspective, the uptake of positions for census data collection and supervision was lower than anticipated. As previously indicated, the timelines for CAWI and CAPI were extended due to poor uptake from communities with fieldworkers. Measures to improve uptake were undertaken; these included reallocation of fieldworkers, remote monitoring of data collection, etc.

The digitisation of the census data collection allowed for the enhancement of quality of the processing of Census 2022. Digitisation enabled data validation to be built into the data collection instruments, resulting in minimal data editing and imputations. As part of the census undertaking, a Post-Enumeration Survey (PES) is conducted, which enables not only the quality of the census data to be assessed but also to estimate over or undercount and to adjust accordingly. Through this process, content error and coverage error are assessed. Thus, the PES is a statistically grounded basis for adjusting census counts appropriately. The PES estimated a *final* net undercount of persons of 31.06% and a net household undercount of 30.49%, which was announced at the release of the Census results in October 2023.

Whilst Stats SA forms part of the global population and housing census programme by adhering to the United Nations Principles and Recommendations of Population and Housing Census, a statutory body is in place to ensure that all processes have been followed correctly. The South African Statistics Council plays an important part in the delivery of a census. The South African Statistics Council is appointed by the Minister in the Presidency to exercise oversight over Stats SA and provide counsel to the Minister and the Statistician-General. This oversight involves ensuring that all processes pertaining to the Census have been meticulously followed and requesting updates on progress from the leadership of Stats SA. Facilitating engagement with the data, an independent consultant appointed by the South African Statistics Council critically reviews the implementation of plans and outcomes of operations, rigorously interrogates the data, and ultimately delivers their professional opinion and concerns. This comprehensive process enables the South African Statistics Council to present its informed opinion on the Census. In the specific case in question, the South African Statistics Council expressed its satisfaction with the quality of the data, leading to the launch of the census results in October 2023.

Another issue that should be acknowledged was that the Census was not very positively received by various communities, which is reflected in the contact rates that were monitored during enumeration. COVID-19 impacted training and the pilot census, as indicated previously. Leading into the enumeration phase of the Census, COVID-19 also had an indirect effect through the economic repercussions of all the COVID lockdowns that had previously followed. All of this was very divisive during the lockdown period across the political and societal divide.

## Issues with the South African Census of 2022

### *Balancing equation*

It needs to be placed on record that Stats SA has always been transparent in methods and operations translating to generated official statistics. While the article<sup>1</sup> cites two census evaluation methods, the authors seem to suggest that the balancing equation method they used provides better population estimates when compared to Census 2022. Such an assertion is misleading. Stats SA has equally and consistently used the Post-Enumeration Survey (PES)<sup>3</sup> as part of the census data evaluation. PES is used to determine the degree of coverage errors and improve

the quality of census results through the application of adjustment factors generated from the PES. The PES, undertaken after the conclusion of the Census enumeration, also allows for examination of characteristics of persons that may have been missed during enumeration, as per the guidelines in the UN Principles and Recommendations (Rev3, par 3, p.221<sup>4</sup>).

The authors of the Commentary must be aware that assumptions on the three demographic processes (fertility, migration and mortality) underlying the balancing equation method vary based on available information at the time of estimation. Such assumptions are the basis for variations in population estimates by different researchers. It is surprising for the authors to seem to suggest that modelled data based on theoretical assumptions is more accurate compared to a census count, adjusted for under coverage. The authors seem to suggest that either their modelled population estimates and Stats SA's mid-year population estimates (MYPE) for the year 2022 are the closest reliable estimates compared to the Census 2022 population count. Important to note is that the MYPE population has its own assumptions. Traditionally, modelled data are updated when a new census data point becomes available. Although there has been upward and higher than usual undercoverage, as highlighted in the PES report in Census 2022, the results are based on the actual count, and transcend the MYPE and any other population estimates. Furthermore, the authors' assessment overlooked the residual (e), which denotes the error of closure. For census evaluation, it is essential to include the residual (e) to balance the equation precisely. In this context, "e" signifies the "error of closure", encompassing the balance of errors in the data on births, deaths, net migration, and the coverage of the two censuses.

Regarding the analysis on the balancing equation, the error in Limpopo is noted from Table 1; however, a further correction of the value for item 6 is that the value for immigration by place of residence should be 1 145 835, which makes the reported/implied migration 30.5%. It is impossible to comment on the reported immigration by place of birth as the assumptions and computation of estimated survivors of foreign born individuals in 2011 are not shared.

It is apparent that the authors undermine Stats SA's technical capabilities in producing reliable, accurate statistics, be it MYPE or censuses. Statements in the article such as "Stats SA methodology for producing their population projections is somewhat opaque" illustrate inherent undermining behaviour that is uncalled-for in an academic and research space.

### **Comparison with previous results**

When comparing the current census to previous ones, it must also be noted that previous censuses have their own considerations and biases, which might distort some comparisons. It is also reasonable to assume that, particularly between 1996 and 2011, the quality of the censuses improved, which makes a comparison between 1996 or 2001 with 2022 biased by different quality measures. The use of population estimates, whether from Stats SA or the SAMRC-UCT collaboration, is noted as an independent data source against which comparisons are made. It must be noted that the different assumptions between the SAMRC-UCT and Stats SA models, and by extension the model derived by any other independent researcher, are likely to be different. One must nonetheless acknowledge that such estimates prior to the 2022 Census do not take into account any changing patterns that the Census might reveal, particularly at lower levels.

### **Issues at sub-provincial level**

Finally, with regard to the issues at the sub-provincial level, although the evidence is convincing, care should be taken to compare with IEC data, which only capture those over 18 and those interested in voting at the local government level. It is also common, albeit anecdotally, that people who have recently moved to a province are often likely to not re-register with the IEC at their newer place of residence and are happy to vote only in general elections. The differences in growth between the IEC data and the Census are nonetheless noted. It must also be noted that satellite imagery should be used with caution, because in built-up areas it is expected that the population may grow in areas where there has been little to no physical growth.

In parting, Stats SA has recently published the 2024 mid-year population estimates by implementing the cohort-component method and using empirical data sources to reconstruct the population in the inter-censal period 2011–2022. The MYPE provides a series of population estimates that are not only robust but also align themselves to the 2001, 2011 and 2022 Censuses. It must further be stressed that we should not lose sight that 70% of the population was counted and that these data are very useful and valuable.

### **Post-Enumeration Survey**

This section seeks to address the following two points:

1. Preliminary undercount rate vs final undercount rate
2. Computation of standard errors for the adjusted population estimates

#### **1. Preliminary undercount rate vs final undercount rate**

While Stats SA encourages the public at large and academia, including renowned researchers, to use census data in their areas of interest, at the same time, it cautions all stakeholders to use data responsibly. Using data responsibly includes highlighting data limitations on Census 2022 and any other products the organisation produces that can inform future improvements and enhancement of methods, processes and products. The introductory remarks of the authors quoting preliminary undercount figures for white and Indian/Asian populations instead of actual figures which were readily available points to the misuse of statistics and an attempt to undermine the credibility of Stats SA.

The article does not make a distinction between the preliminary and final undercount rates. The preliminary undercount rate (29.6%) is derived directly from the PES estimate of the population, while the final undercount rate (31.1%) is calculated based on the adjusted and content-edited census data. Therefore, any mention of the official undercount rate should make reference to the final undercount rate.

The article further references undercount rates of 72% and 62% of the Indian/Asian and white population groups, respectively. It should be stressed that these are preliminary undercount rates and should not be mistaken for the final undercount rates. The final undercount rates for these population groups were estimated at 42.10% for the Indian/Asian and 24.86% for the white population groups (see Table 15 in the PES report<sup>5</sup>).

#### **2. Computation of standard errors for the adjusted population estimate**

The article makes an assertion that there could be a computational error in the derivation of the standard errors (SE) for the adjusted population estimate in the PES 2022. The basis for this assertion seems to be the expectation that a final undercount rate of 31.1% should inherently result in a higher SE for the adjusted population estimate in Census 2022 compared to Census 2011, given the lower undercount rate of 14.6% in Census 2011.

As it was the case in Census 2011, the adjusted population estimate in Census 2022 is a product of the actual census counts and adjustment factors within homogeneous classes obtained by post-stratification, as detailed in the released PES report. Table 1 shows estimates of the undercount rate and their SE, and estimates of the population and their SE at national and provincial levels for 2011 and 2022. The calculation of SE for the population estimate in PES 2022 has been verified and confirmed as reported. The SE is a measure of how much the estimate would vary if the survey could be repeated over and over using a different sample of units of the same size every time, i.e. it gives an absolute amount of variability in the estimate. The SE for population estimate is dependent on the variation associated with the estimate and not on the magnitude of the undercount. A higher undercount rate does not necessarily translate to a higher SE of the population estimate. In theory, a smaller sample size would result in high standard errors. However, the undercount rates are not sample sizes but estimates of the missed population.

The coefficients of variation (CVs) of the population estimates for 2011 and 2022 are within the prescribed limits of the South African Statistical Quality Assessment Framework. The CVs measure the variability of each estimate

**Table 1:** Estimates of the undercount rate (UC) and their standard errors (UC SE), and estimates of the population and their standard errors (SE), for South Africa and its provinces, 2011 and 2022

	2011		2022		2011			2022		
	UC (%)	UC SE (%)	UC (%)	UC SE (%)	Population	SE	CV (%)	Population	SE	CV (%)
<b>Western Cape</b>	18.5	0.542	35.56	1.56	5 822 734	40 830	0.70	7 430 000	70 000	0.98
<b>Eastern Cape</b>	12.9	0.196	30.56	0.59	6 562 053	125 810	1.92	7 230 000	60 000	0.86
<b>Northern Cape</b>	13.4	0.318	29.11	2.27	1 145 861	82 466	7.20	1 350 000	30 000	2.20
<b>Free State</b>	10.1	0.362	18.52	2.62	2 745 590	117 567	4.28	2 960 000	40 000	1.35
<b>KwaZulu-Natal</b>	16.7	0.379	31.19	1.08	10 267 300	109 994	1.07	12 390 000	100 000	0.77
<b>North West</b>	14.9	0.532	17.92	5.39	3 509 953	166 754	4.75	3 800 000	40 000	1.14
<b>Gauteng</b>	14.7	0.174	30.79	2.07	12 272 263	106 023	0.86	15 120 000	90 000	0.61
<b>Mpumalanga</b>	15.5	0.473	34.00	4.51	4 039 939	219 299	5.43	5 160 000	60 000	1.13
<b>Limpopo</b>	10.0	0.135	23.77	1.02	5 404 868	251 244	4.65	6 570 000	60 000	0.91
<b>South Africa</b>	<b>14.6</b>	<b>0.132</b>	<b>29.6</b>	<b>0.82</b>	<b>51 770 560</b>	<b>997 560</b>	<b>1.93</b>	<b>62 030 000</b>	<b>120 000</b>	<b>0.19</b>

Note: The standard error (SE) is a measure of how much the estimate would vary if the survey could be repeated over and over using a different sample of units of the same size every time, i.e. it gives an absolute amount of variability in the estimate.

The coefficient of variation (CV) or relative standard error (RSE) is the ratio of the standard error of the survey estimate to the value of the survey estimate expressed as a percentage. It gives the relative amount of variability instead of the absolute amount of variability in an estimate. It allows for comparing domains whose estimates and standard errors differ in magnitude.

and, therefore, the CVs for coverage error rates and population estimates cannot be compared with each other as they refer to different constructs.

Finally, the language used in the article, including in its title ‘Problems and concerns with the 2022 South African census’, as well as the introduction, conclusion sections, and the various subtitles referred to as “issues” in the journal article, is characterised by a sensational and malicious tone. It appears that the article’s intent was not solely to highlight data limitations, as is common practice, but to cast doubt on the credibility of Stats SA, a renowned national statistical office. The use of such dramatising language in a journal article is concerning, as it may indicate an underlying agenda to discredit the Census 2022 results. The article suggests that the census was conducted without sufficient technical expertise, thereby calling into question the validity of its operations, methods applications, and data evaluation procedures. This insinuation not only undermines the integrity of Stats SA but also raises doubts about the accuracy of the Census 2022 results.

## References

1. Moultrie TA, Dorrington RE. Problems and concerns with the 2022 South African census. *S Afr J Sci.* 2024;120(7/8), Art. #18585. <https://dx.doi.org/10.17159/sajs.2024/18585>
2. Statistics South Africa. Census 2022 statistical release. Pretoria: Statistics South Africa; 2023. Report P0301.4. Available from: [https://census.statssa.gov.za/assets/documents/2022/P03014\\_Census\\_2022\\_Statistical\\_Release.pdf](https://census.statssa.gov.za/assets/documents/2022/P03014_Census_2022_Statistical_Release.pdf)
3. Statistics South Africa. Post-enumeration survey statistical release. Pretoria: Statistics South Africa; 2023. Report P0301.5. Available from: <https://census.statssa.gov.za/assets/documents/2022/P030152022.pdf>
4. United Nations. Handbook on the management of population and housing censuses, revision 2. New York: United Nations; 2021. Available from: [https://unstats.un.org/unsd/publication/series/series\\_f83rev2en.pdf](https://unstats.un.org/unsd/publication/series/series_f83rev2en.pdf)

### Note from the Editor-in-Chief:

We are grateful to Stats SA for their response. As with all such items for our Journal, we provided the authors with comments from expert readers; the authors did not change their submission subsequent to receiving these comments, and the response has been published as submitted, in the interest of open debate in science and its real-world applications.