Problems and concerns with the 2022 South African census

Tom Moultrie and Rob Dorrington

Centre for Actuarial Research (CARe), UCT

Background: Why a census…?

- **The census matters**
  - A snapshot at a point in time
    - Demography; socio-economics
    - Key benchmark estimates of fertility and mortality
  - Largest 'survey'
    - Allows estimation at smaller spatial units
  - Basis for planning
    - Equitable share formulae
    - Infrastructural planning
  - Sampling frame for other surveys

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 "*The 2000 Census: Counting Under Adversity*" p1

Background: ...when the world moves on

‣ Many developed countries (the US being an important exception) are or are considering relying on linked administrative databases to replace regular censuses.

‣ This is not possible in most countries. The quality of such administrative data in developing countries is scanty, patchy, incomplete, and often unlinkable, making those countries more reliant on the census as a basis for planning etc.

Background: Censuses in South Africa

▶ Pre-1994, only two South African censuses are thought to have accurately enumerated the population; in 1936 and 1970

  ▶ Separate censuses conducted in the 'independent homelands' after 1975
  ▶ Ongoing conflict and lack of legitimacy compromised censuses in the 1980s

▶ Post-1994, censuses conducted in 1996, 2001, and 2011

  ▶ Undercounts of 10%+, 17%, and 14% respectively

Background: Census 2022

▶ Original plan was to hold the census on 10 October 2021, exactly 10 years after the 2011 census

▶ Repeated waves of Covid caused delays in planning and logistics

   ▶ Census date finally set as 2 February 2022; with data collection planned to be completed by the end of that month

      ▶ This was extended (first) to 20 March, then to 30 April (for CAWI), and - in the Western Cape - to 31 May 2022

   ▶ Forced move to electronic (CAPI/CAWI/CATI) data collection – with insufficient testing/training

Background: Census 2022 'highlights'

▶ Results of the South African census released on 10 October 2023
  ▶ 20 months after the census date (cf. 12 ½ months in 2011)

▶ Headline population estimate of 62m at the census date

▶ Undercount (nationally) of 31%
  ▶ 72% among Indian, and 62% among white, South Africans
  ▶ Highest of over 300 census undercounts on the UN database
  ▶ Previous 'record': Comoros 2017, with an undercount of 21%

Background: What we still don't know

▸ Appraisal of census data hampered by limited data availability
  ▸ Still no substantial data on migration; none on fertility or mortality
  ▸ Still no 10% public-use micro-sample

Analysis – Part 1

▸ The 'Balance Equation' is the most fundamental tool for evaluating a census:

$$P(t+n) = P(t) + B(t, t+n) - D(t, t+n) + NM(t, t+n)$$

▸ We have:

   ▸ a high degree of confidence in the population count in 2011

   ▸ a good idea of births and deaths over the intercensal period

▸ Assuming the earlier census and intercensal births and deaths are accurate, then any imbalance must be migration or census count

   ▸ Dominant migration tends to occur at particular ages, peaking in young adult ages (Rogers & Castro)

Analysis – Part 1

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%

Requires net immigration of 3.76m from 2011 to 2022 to 'balance' the reported populations

But only 460-620k immigrants identifiable in the census data based on reported place of residence, or place of birth

Notes, sources, and formulae:

(1)   Table 2.1 [1]
(2)   Own estimate (probably accurate to nearest 100 000)
(3)   Own estimate (probably accurate to the nearest 100 000)
(4)   (5) – ((1)+(2)–(3))
(5)   Table 2.1 [1]
(6)   Table 2.1 [1] The reported value of in-migration in Limpopo is mis-calculated based on the data shown.
(7)   Own estimate (subtracting estimated survivors of foreign-born identified in the 2011 census from those reported as foreign-born in the 2022 census).

Analysis – Part 1

▸ If we do the same exercise by sex…
Of the implied migration of 3.76m…

   ▸ 2.06m would have to be female; and 1.70m male

▸ If we repeat the exercise by population group…
Of the implied migration of 3.76m…

   ▸ 3.57m African; -0.04m Coloured; 0.34m Indian; -0.13m White

      ▸ As a percentage of 2011 population: 9%; -1%; 26%; -3%

Analysis – Part 2

▸ A second investigation is to compare the results with previous censuses (taking mortality and migration into account) and population projections
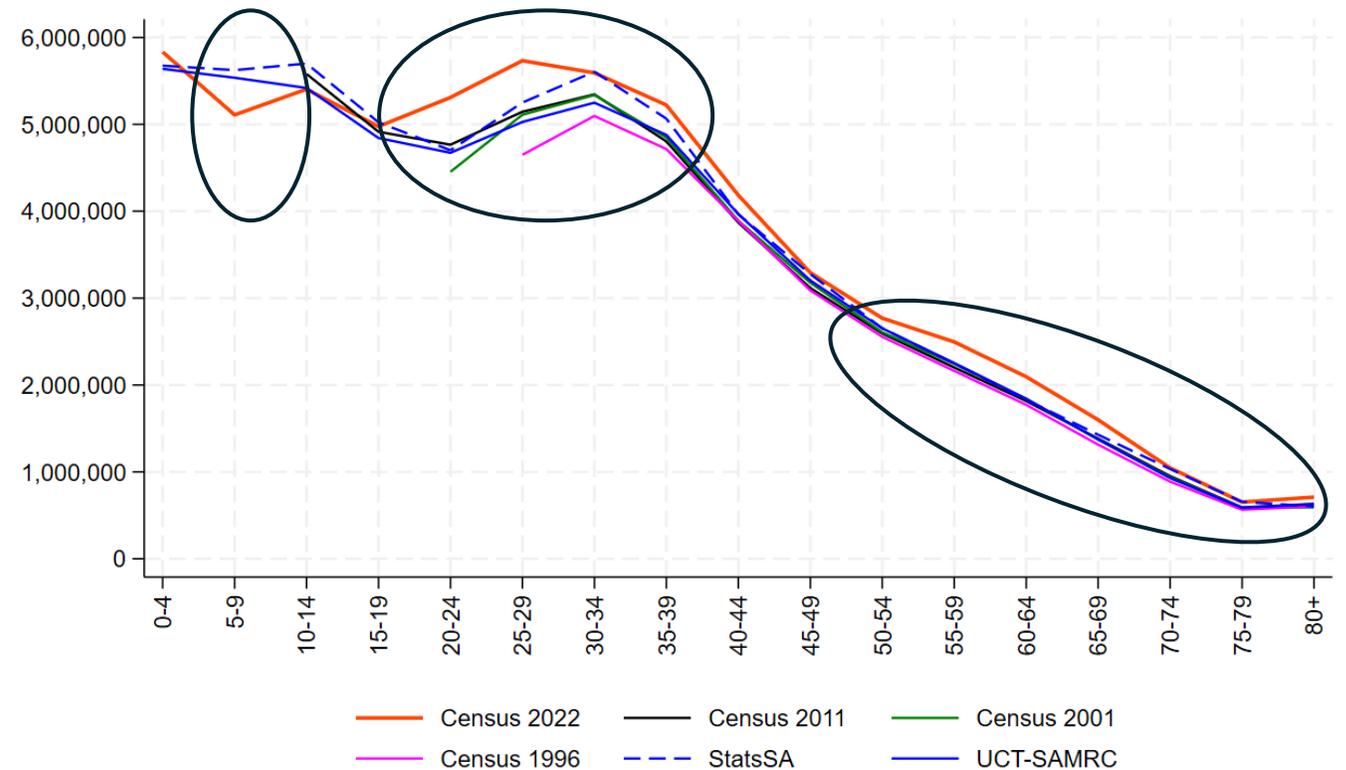
African
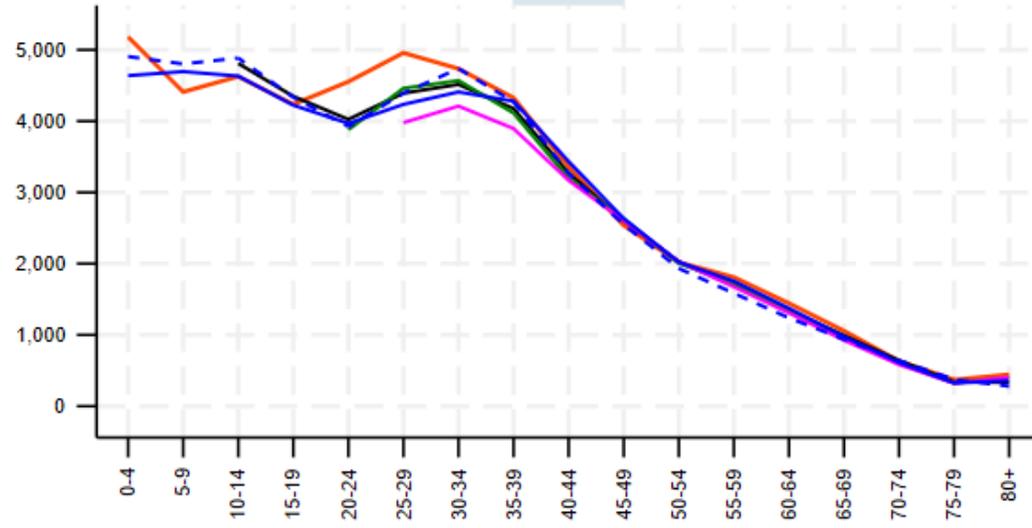
Coloured
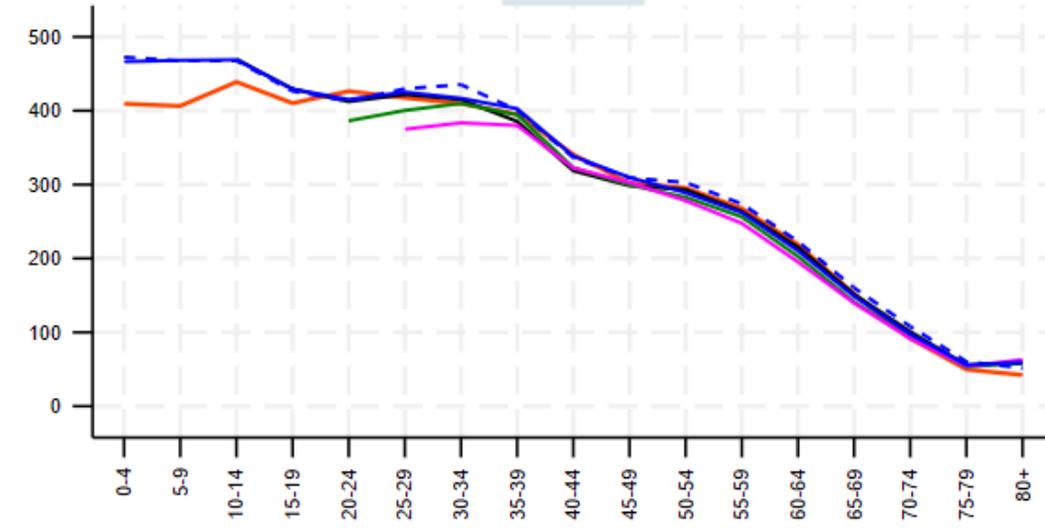
Indian
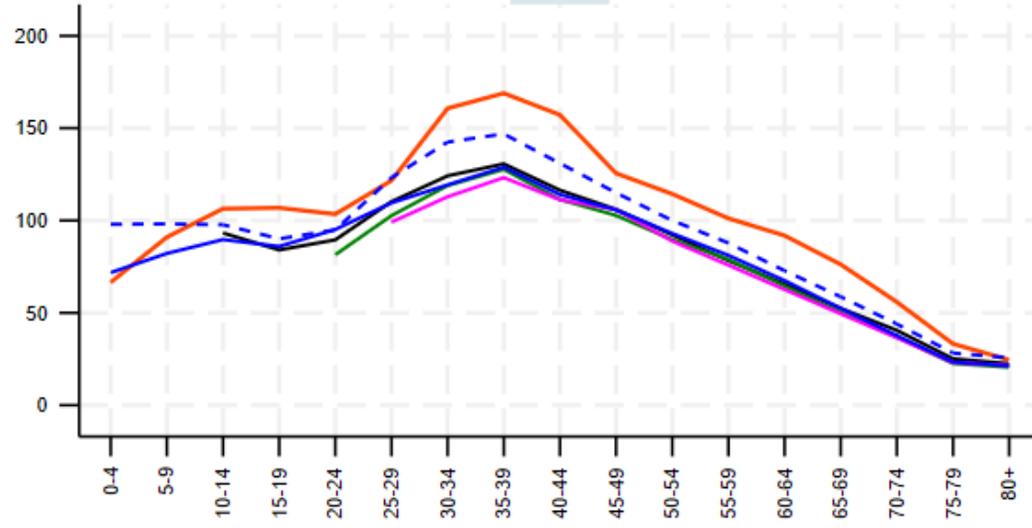
White
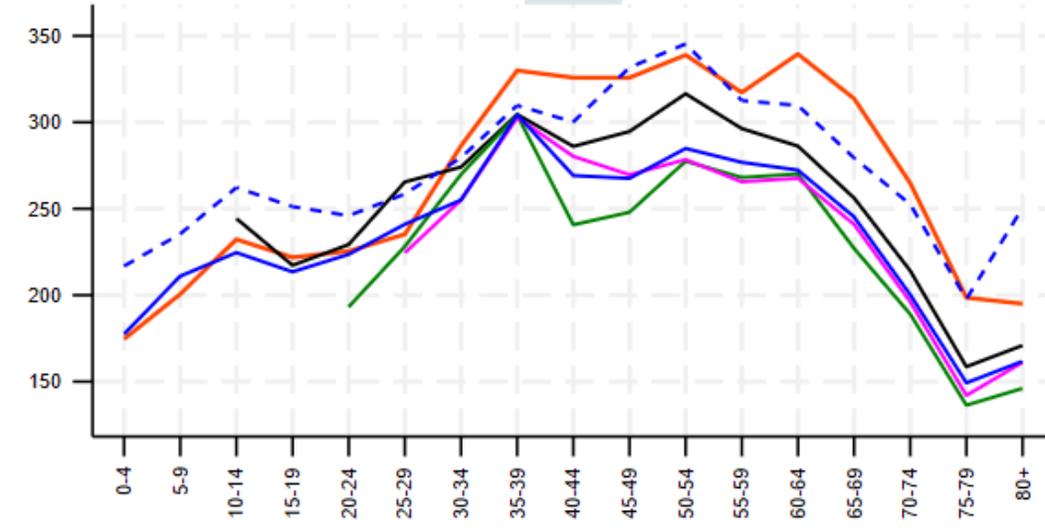
Census 2022   Census 2011   Census 2001

Census 1996   StatsSA   UCT-SAMRC
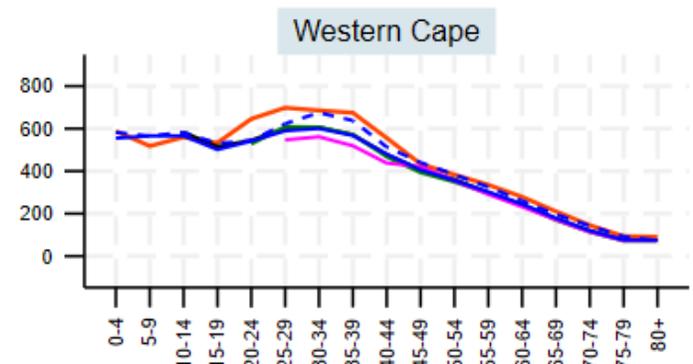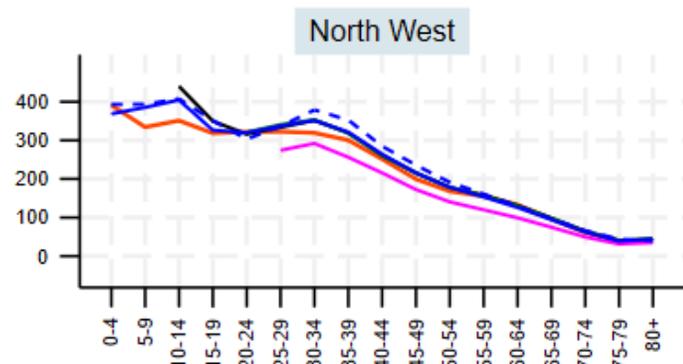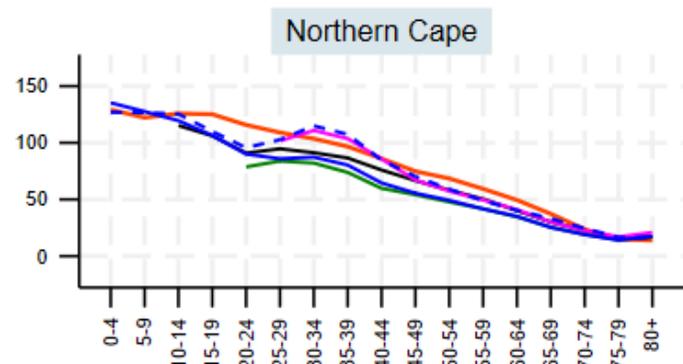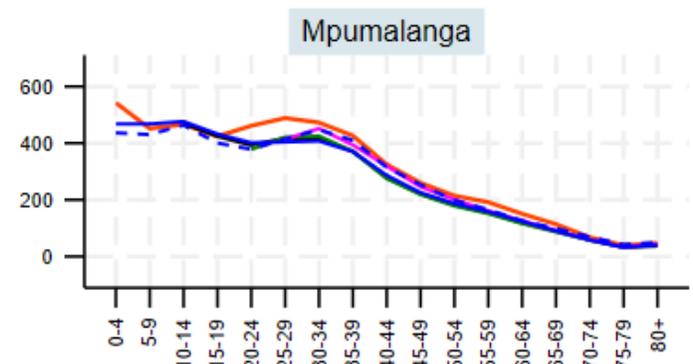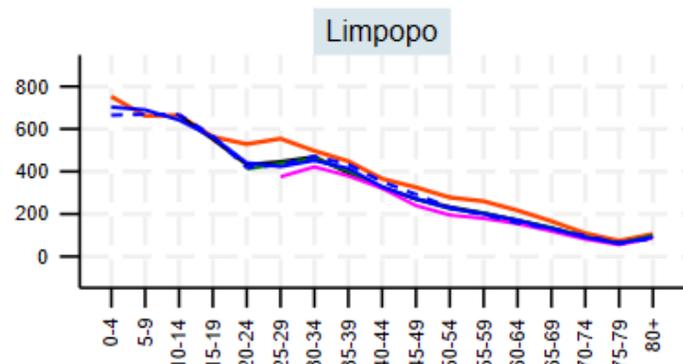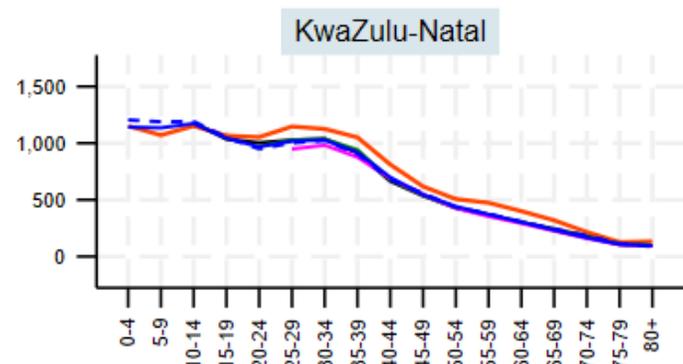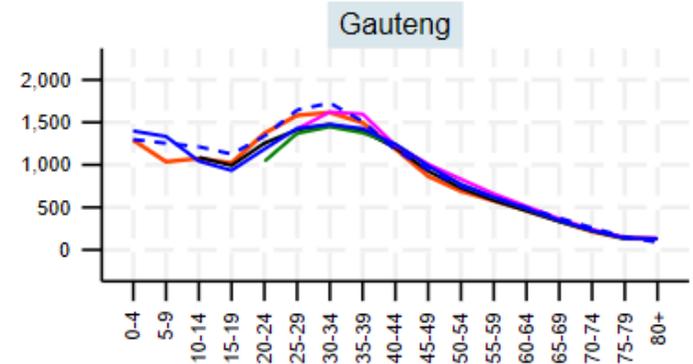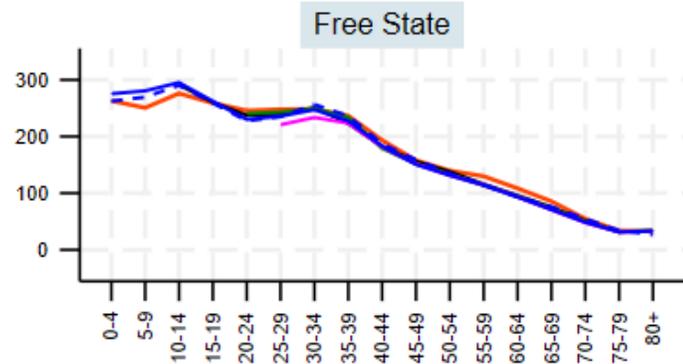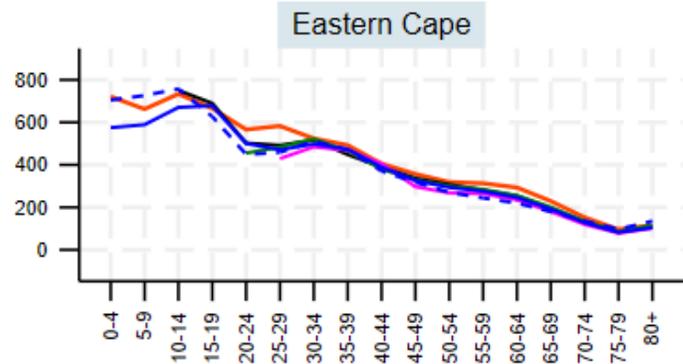
Census 2022	Census 2011	Census 2001
Census 1996	StatsSA	UCT-SAMRC

Analysis – Small-area estimates

▸ We examine the intercensal growth of the census populations by district council ...

| Code	District Name		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%

Analysis – Small-area estimates

▸ …and also compare the 2022 census results with the most-recent district-level mid-year population estimates produced by Stats SA (recast to the census date)…

Code	Name		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%
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%

A quick d(r)ive into the Central Karoo

▶ Beaufort West LM (WC053) accounts for approximately 70% of the population of the Central Karoo (DC5)

  ▶ Population increased from 50 000 to 73 000 between the censuses (47%; 3.5% p.a.)

Beaufort West Local Municipality: Population by age, 2011 and 2022
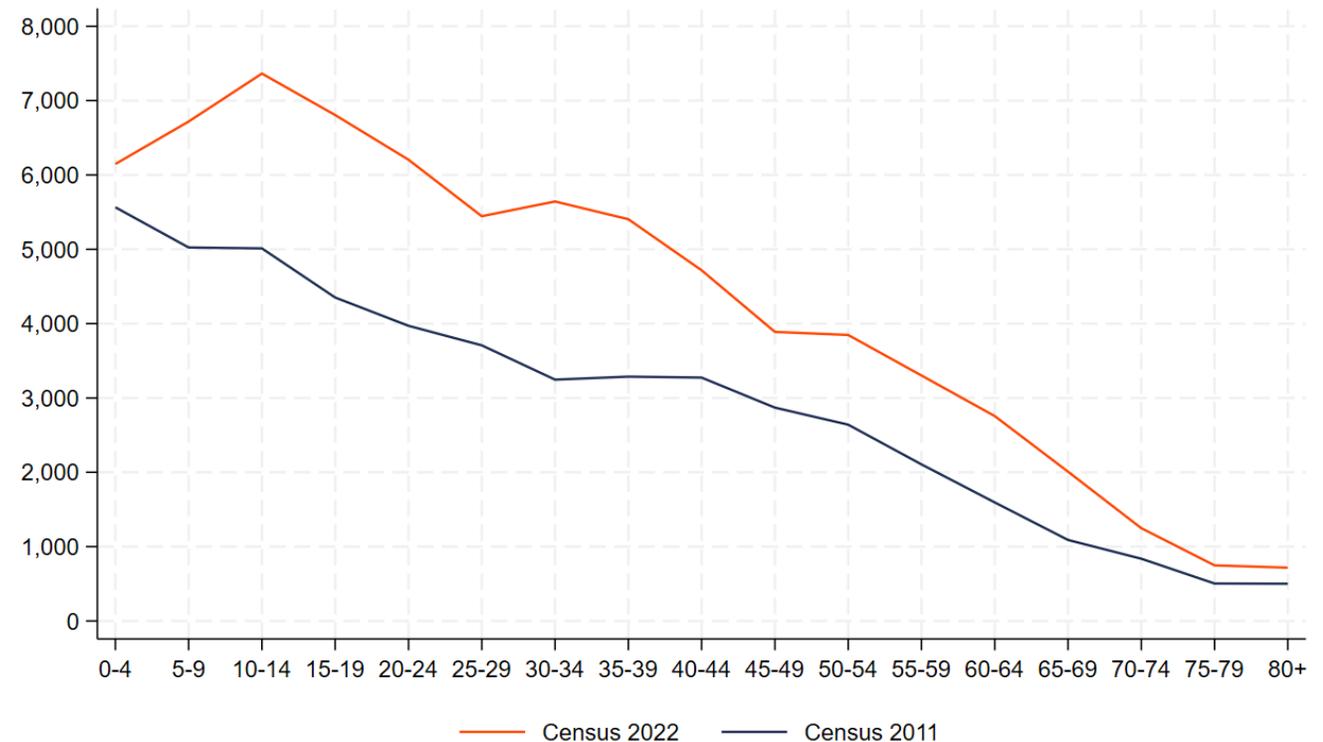
A quick d(r)ive into the Central Karoo

▸ ... but Beaufort West town accounts for roughly 70% of WC053, and hence just under half of DC5

▸ Examining GoogleEarth maps of Beaufort West town from January 2010 to January 2022

  ▸ Only the areas outlined in red were developed over that period

Analysis – Comparison with LGE data

▶ We compare, at a municipal level, the census populations aged 18 and over, and the voter registration data from the 2011 and 2021 LGE:

		Beaufort West WC053	Mossel Bay WC043	Local Municipality 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	Census	54.9%	68.7%	71.4%	102.4%	-23.6%
2011:2021/2	RV	12.9%	22.7%	32.4%	4.6%	17.4%

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

What might have caused this?

A tale of two post-enumeration surveys

▸ South Africa follows international best-practice in conducting a post-enumeration survey (PES) after the census, to

 ▸ Estimate the number of persons missed in the census

 ▸ Identify coverage and content errors in the census

 ▸ Correct the census enumeration

▸ The UN recommends the PES happens very soon after the census date (to limit discrepancies from births, deaths etc)

 ▸ But the PES was delayed (because the census fieldwork was extended)

A tale of two post-enumeration surveys

▸ The PES is quite small – sampling around 0.6% of (roughly) 130 000 Enumerator Areas (of roughly 450 people)

  ▸ Small sample size limits granularity of adjustments

    ▸ Province, type of EA, population group, sex, broad age group

▸ With such a large undercount, the adjustment factors do a lot more 'heavy lifting'

  ▸ It seems that the PES design did not take the possibility of such an extreme undercount into consideration

A tale of two post-enumeration surveys

▸ The 2011 and 2022 censuses sampled similar proportions of enumerator areas in the PES

  ▸ When a first estimate of the undercount is derived directly from the PES exercise, the standard errors in 2022 are broadly similar but significantly higher (as expected) …

	2011		2022	
	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)

A tale of two post-enumeration surveys

▶ … but (oddly!) the standard errors of the final population estimate in 2022 are significantly lower

- ▸ CI (2011): 51.77m±1.96m
- ▸ CI (2022): 62.03m ±0.235m

	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)

A tale of two post-enumeration surveys

▸ … and with no correlation between the two series of coefficients of variation (CVs) at a provincial level…

▸ Our concern is that, given the undercount, the 2022 PES was too small to accurately adjust the census enumeration

▸ Leading to the odd small-area and small-population estimates (*inter alia*)
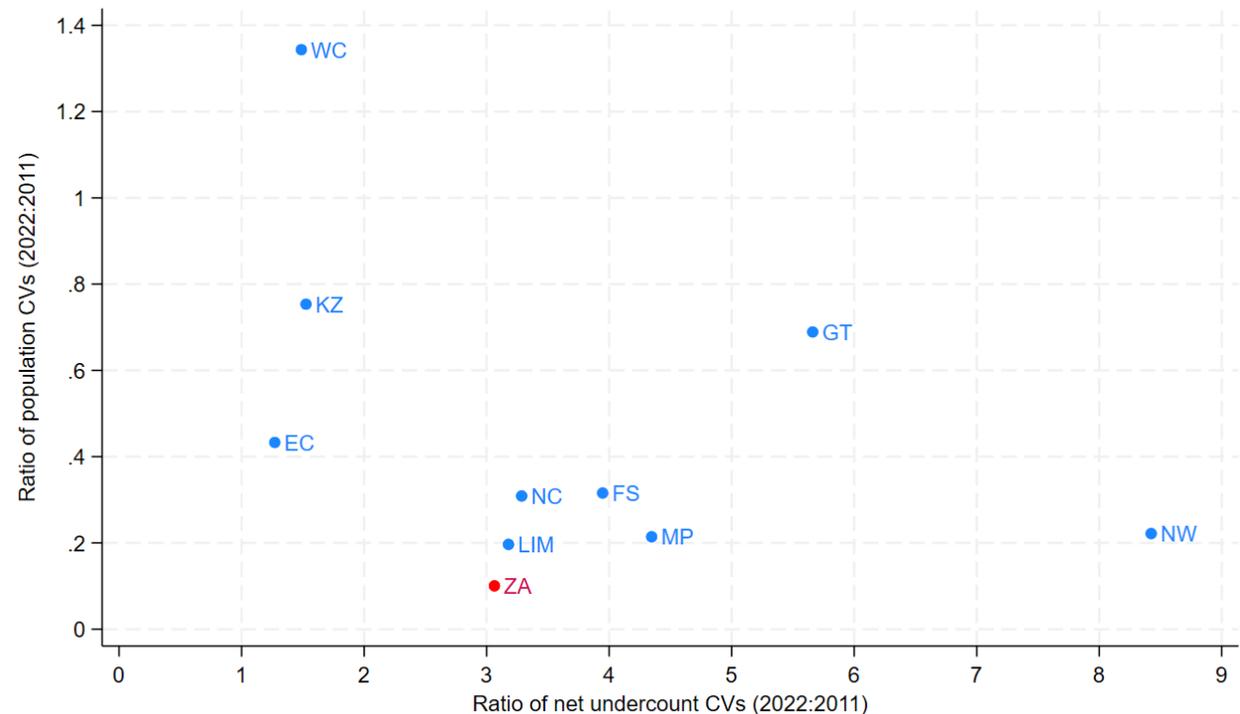
What (else) happened?

- Covid impacts on planning and logistics
  - Recruitment and training of field staff
  - Stress-testing of logistics and IT systems
- Reliance on rushed implementation of CAWI
  - Subsequent need to mount a sizeable fieldwork operation when not intended
- National Treasury
  - Prevented further delay beyond end of the 2021/2 fiscal year
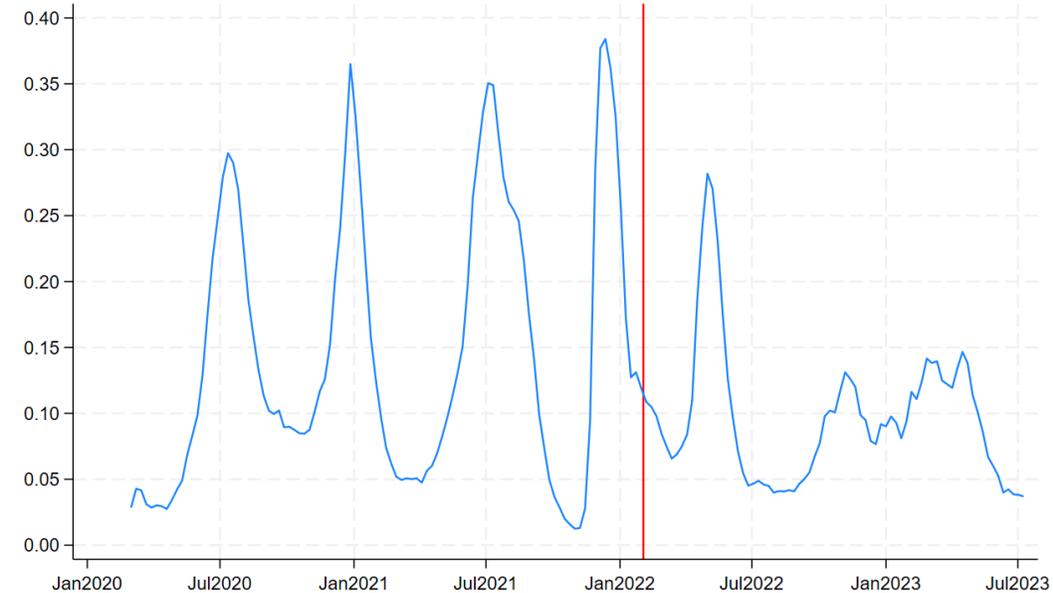
Conclusions

▸ The population estimated in the 2022 census is probably some 1m higher than it is likely to have been

▸ Estimates of small-area, and small(er) populations highly unreliable

▸ Undercount of children 5-9, especially at age 5

▸ Overcount of adults 50+ (highly unlikely to be migrants)

▸ Provincial numbers poor in EC, KZ and ?NC

  ▸ Less so in MP and LM; perhaps better in FS, GT, NW, and WC

Implications

- Numbers of people not fit for purpose for fiscal allocation, national, provincial, local government planning and resource allocation
  - Impact on Equitable Share Formula?
  - Impact on sampling frames for other studies

- Proportional measures derived from the census may be more robust to the extent that errors in numerator and denominator cancel each other out
  - But <u>not</u> rates and ratios where numerator from non-census source

Implications (2)

▶ For the SAMRC:
  ▶ Implications for sub-national health and disease monitoring
    ▶ Sampling, weighting, and interpretation of National Burden of Disease (NBD) study
  ▶ Denominators for populations-at-risk, and hence rates

▶ Uncertain if the data on births, survival of births, and deaths reported by households are at all useable
    ▶ These are key 'pegs in the ground' for checking vital registration, and particularly at sub-national levels

Next steps

▸ Release of this analysis through the SAMRC and *SAJS*

  ▸ hopefully, soon!

▸ Need to produce alternative population estimates, at least down to a district-level, to better inform evidence-based resource allocation and planning

  ▸ Future work is greatly constrained by Stats SA's failure to release more data, especially relating to fertility and mortality

  ▸ A work in progress…

Plus ça change … (or is it déjà vu?)

Crack of Dawn

'Ag, no man! … Some professors are saying the census is nonsensus.'
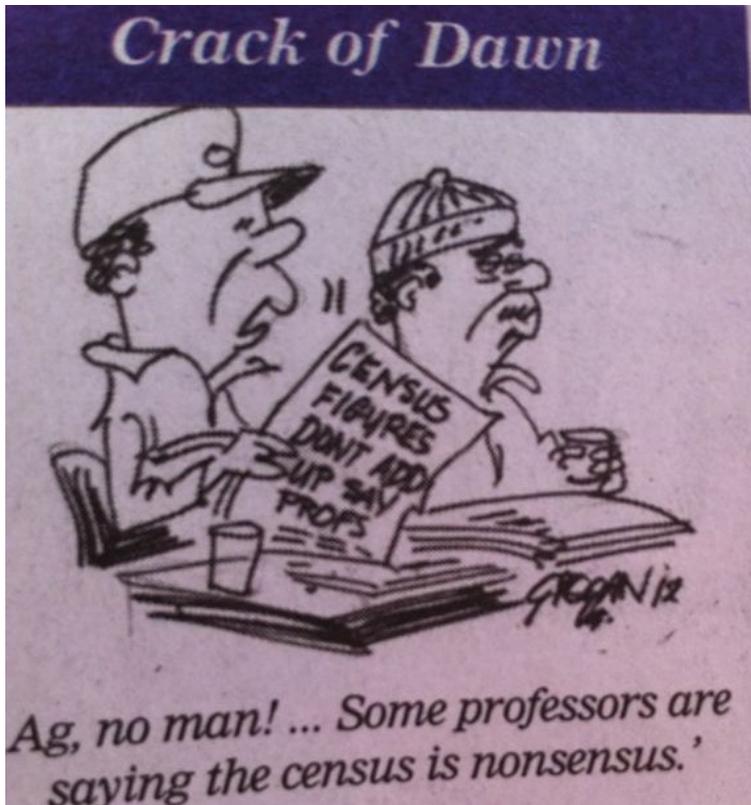
Tony Grogan, *Cape Times*, November 2012

Thank you!