Undercounting controversies in South African Censuses

Jeremy Gumbo:
Demography & Population Studies Programme, School of Social Sciences, Wits University
Email: Jeremy.d.gumbo@gmail.com
Clifford Odimegwu
Demography & Population Studies Programme, School of Social Sciences, Wits University
Email: Clifford.odimegwu@wits.ac.za

Short abstract

The main purpose for census taking is to obtain an accurate population count often used for directing policy formulation and resource allocation in a given country. South Africa's last three censuses have largely failed to achieve this, as high undercount rates have been consistently recorded. National undercount rates have been 10.6%, 17%, and 14.6% for censuses 1996, 2001 and 2011 respectively. Such high undercounting has triggered controversies especially around population counts arrived at by census authorities using Post Enumerative Survey (PES) when estimating and adjusting for undercount. In this study we applied various demographic techniques to access the accuracy of adjusted population counts in these censuses. These included; growth rate analysis, graphical cohort analysis, age ratios, and sex ratios. Findings from the various analyses suggest coverage errors in these censuses. However, such distortions could also result from content errors, and/or population changes through migration.

Extended Abstract

Background

Censuses are any country's biggest statistical collection process, and to underline their significance, conduction of censuses is often mandated in national constitutions. In South Africa for example, the Statistics Act No. 6 of 1999, makes it mandatory for Statistics South Africa (Statssa) to conduct a census after every 5 years despite the fact that 10 year censuses are the norm (Cronje and Budlenger, 2004). In most cases national governments show commitment to censuses by including them in national budgets. A census involves enormous resource input, and the expenditure of resources arising is believed to be worthwhile as data collected is expected to inform and direct policy makers on circumstances and needs of a population (Cronje and Budlenger, 2004; Steffey 1997).

Though today, the key objective in conducting a census would be to obtain both accurate population estimate and distributions; this has often remained elusive for most census authorities. This is so, because censuses are prone to a range of errors; whose fundamental distinction is between coverage and content errors (Keane *et., al*, 1995). The latter are errors that are recorded about characteristics of individuals who will have been captured in a census. Such errors include misclassifications by; age, education, sex, region, to mention some. Content errors affect the distribution of population recorded in a census with respect to characteristics like those listed above (Keane *et., al*, 1995). On the other hand coverage errors can either be a result of undercounting, or over counting. Coverage error also affects population distribution just like content errors. However, unlike content errors, coverage errors further affect population figures. This implies that irregular or distorted population distributions that can be noted from census data may not be pretty clear as to whether they are due to either content or coverage error, or even both. Therefore estimation of undercounting using techniques that rely on observing whether population distributions from investigated census data deviate from expected standard distributions may not provide reliable undercount estimates unless effect of content error has been taken care of. A fact which this proposed study shall consider.

Undercounting which is the focus of this proposed study is often prevalent in census taking than over counting. In the United States (US) censuses of 1990 and 2001 for example, undercount estimates have been 1.6% (Hogan, 1993) and 1.2% (Statistics South Africa, 2010) respectively. In China, the world's most populous nation, undercount in 2000 census was estimated at 1.8% (Anderson, 2004). Other examples of undercounts are; 2001 Napal census, 5.3%; and Canada 1996 census, 2.6% (Statistics South Africa, 2010). In countries like the US, undercounting has often resulted in unresolved contests mainly between census and local government authorities, which at times ended up spilling into courts of laws (Breiman, 1994).

As for South Africa, a trace of her censuses indicates that high undercounting has remained a consistent feature over time. Due to high undercounting, South African censuses have become subject to controversies particularly since 1996 census when Dorrington (1999) questioned the credibility of undercount estimates for white population. Since then more researchers, organisations and members of the public joined the debate questioning certain processes and outcomes in these censuses. For example in the 2011 census, a research by University of Cape Town's Centre for Actuarial Research (CAREe) suggested that PES adjustment over estimated young white women population. Yet other researchers argued that PES adjusted census counts favoured provinces like Western Cape and Gauteng at the expense of a province like Eastern Cape in 2011 census (Berkowtz, 2012). Even members of the public have also strongly expressed their reservations on the adjusted census counts, through various social media platforms.

Our Research question

In our study we identified that despite the magnitude of controversies surrounding South African censuses particularly on accuracy of the census counts arrived at using PES adjustments, there has been very little effort made in carrying out scientific studies to investigate these concerns. Based on this research gap, our study sought to answer the research question: Is there a difference in undercounting estimates between those from PES and those from Demographic Analysis?

Data

The study used data came from censuses 1996, 2001, 2011. Statistics South Africa is mandated by an Act of Parliament to carry out censuses in South Africa. This data is collected from all individuals residing in the country; and includes individuals' demographic and socioeconomic characteristics. Since 100% accuracy in census enumeration is not often achievable mainly due to undercounting, adjustments are often done to correct for this. The adjusted census count becomes the official census result that census authorities in South Africa publish for public use. A 10% sample of the data is available for each census from Statistics South Africa's website, and can also be obtained from IPUMS-International's website.

The default process of data collection in South African censuses is through enumerators' visiting each household in South Africa on the census day, where they hand over a questionnaire to the head of household to complete. The form is completed in the presence of the enumerator and immediately handed back upon completion. In exceptional cases the household head was allowed on request to remain with the form and feel it later and send it e.g. via post.

Methods

Various demographic techniques applied to assess accuracy of PES adjusted census counts included firstly, growth rate analysis. History confirms that population growth rates (r) for any given country should range between 0 and 3.5 particularly if natural increase is the main explanation of population change (Moultrie et al, 2013). Overlap or approach to these limits by the respective country should be investigated. Secondly, is the graphical cohort analysis, where the lines marking the distribution of population by age in each census should follow same trend, should not cross the other census' line, and lines for latest censuses should be at the bottom (Keane et al, 2005). Thirdly, age ratios expected across age groups should be 1, based on the assumption that population change between age groups should approximately be linear, and hence the formula for computing age ratios becomes:

$$ {}_nAR_x = \frac{2 * {}_nN_x}{{}_nN_{x-5} + {}_nN_{x+5}} * 100 $$

Where ${}_nAR_x$ = age ratio for age group x; $2 * {}_nN_x$ = population count for a particular age group x whose age ratio is being estimated in the equation; ${}_nN_{x-5}$ = population count for age group coming immediately before the age group x; and ${}_nN_{x+5}$ = population count for age group coming immediately after the age group x

And the key assumptions underpinning this formula are that: Census counts for the two age groups that immediately come before and after the investigated age group's count are assumed to be both correct. Secondly fertility, mortality and migration are all assumed to be constant.

Fourthly, the sex ratios which are largely considered to be one of the most robust demographic analyses for estimating undercounting (Robinson J G et al, 1993) were estimated based on the formula:

$$ {}_nSR_x = \frac{{}_nN_{x\ males}}{{}_nN_{x\ females}} * 100 $$

Where ${}_nSR_x$ represents the enumerated population of sex i (i = male or female) between ages x and x + n; ${}_nN_{x\ males}$ represents enumerated males in age group x; and ${}_nN_{x\ females}$ represents enumerated females in age group x

Preliminary Results in Tables and Graphs

Growth rate Analysis

Table 1: Males

Age group (Years)	1996 Census	2001 Census	Growth Rate (r)	2001 Census	2011 census	Growth rate (r)
0-4	2215745	2214369	-0.00012	2214369	2867584.9	0.02585
5-9	2340363	2423906	0.00701	2423906	2425181	0
10-14	2309587	2510361	0.01667	2510361	2344275	-0.0069
15-19	2048208	2454284	0.03617	2454284	2498572	0.00179
20-24	1914992	2100064	0.01845	2100064	2694646	0.02493
25-29	1661866	1893200	0.02607	1893200	2542681.7	0.0295
30-34	1452053	1596760	0.019	1596760	2036206	0.02431
35-39	1278644	1438418	0.02355	1438418	1709346.5	0.01726
40-44	1026535	1230423.1	0.03623	1230423.1	1402328	0.01308
45-49	809067.789	962657.87	0.03476	962657.87	1195740	0.02168
50-54	599708.912	770704.03	0.05017	770704.03	1011349	0.02717
55-59	482331.661	551102.11		551102.11	811949.96	0.03875
60-64	351752.415	447549.3	0.04817	447549.3	612363.96	0.03135
65-69	307073.316	305168.98	-0.00124	305168.98	401548.2	0.02745
70-74	195351.08	230192.45	0.03282	230192.45	297144.509	0.02553
75-79	141217.881	136967.29	-0.00611	136967.29	163690.73	0.01782
80-84	61926.161	91981.021	0.07913	91981.021	100128.35	0.00849
85-89	29129.458	30519.573	0.00932	30519.573	43720.4802	0.03594
90+	13943.54	17077.66	0.04055	17077.66	30334.111	0.05745
Total	19239495	21405705.2	0.02134	21405705.2	25188790.9	0.01627

Table 2: Females

Age group (Years)	1996 Census	2001 Census	Growth Rate (r)	2001 Census	2011 census	Growth rate (r)
0-4	2223343	2215008	-0.00075	2215008	2817867	0.02407
5-9	2332033	2425994	0.0079	2425994	2394570	-0.0013
10-14	2359138	2541811	0.01492	2541811	2250611	-0.01217
15-19	2135672	2527782	0.03371	2527782	2504905	-0.00091
20-24	2067653	2189344	0.01144	2189344	2679896	0.02022
25-29	1790412	2034172	0.02553	2034172	2516635	0.02128
30-34	1617576	1741231	0.01473	1741231	1992804	0.0135
35-39	1375399	1635554	0.03465	1635554	1758420	0.00724
40-44	1105325	1376879	0.04394	1376879	1546291	0.0116
45-49	863268.102	1125861	0.05312	1125861	1424543	0.02353
50-54	665172.108	870990.936	0.05392	870990.936	1206940	0.03262
55-59	586841.047	650859.782	0.02071	650859.782	985458.39	0.04148
60-64	537440.517	622622.77	0.02942	622622.77	773404.01	0.02169
65-69	455615.111	483069.23	0.0117	483069.23	556256.456	0.01411
70-74	286534.542	396651.2	0.06504	396651.2	453343.42	0.01336
75-79	238685.045	231978.27	-0.0057	231978.27	317675.03	0.03144
80-84	116341.363	179941.73	0.08722	179941.73	222072.27	0.02104
85-89	61987.192	65320.607	0.01048	65320.607	102683.16	0.04523
90+	32145.978	47908.833	0.0798	47908.833	77394.255	0.04796
Total	20850582.4	23362979.1	0.02275	23362979.1	26581769.3	0.01291

Cohort Graphical Analysis

Fig 1: Males.

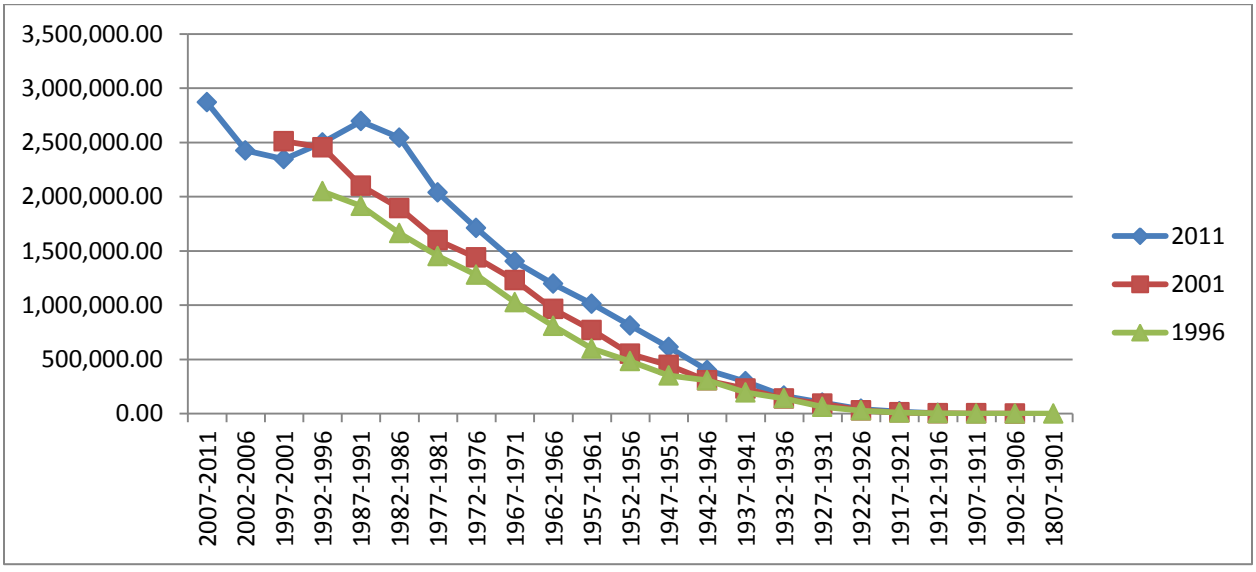
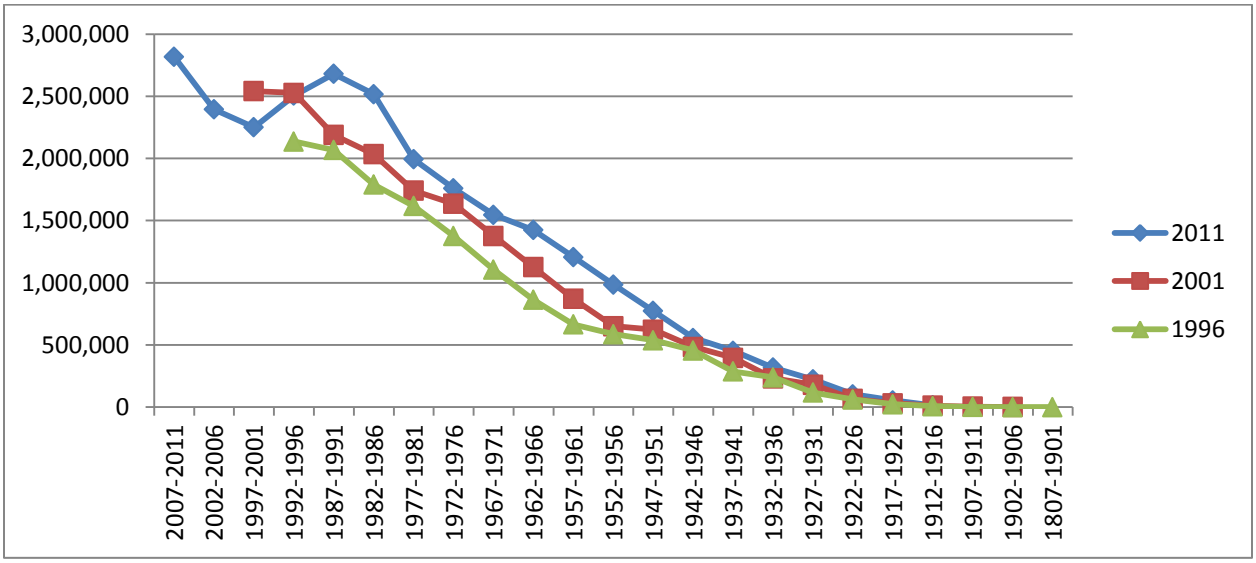


Fig 1: females



Expected Age Ratios versus those obtained

Table 3: Comparison of PES adjusted census' age ratios against those expected (males).

Age group	1996 census	Age ratio		2001 census	Age ratio		2011 census	Age ratio	
	Count	PES	Expec	Count	PES	Expec	Count	PES	Expec
0-4	2215745			2214369			2867585		
5-9	2340363	103.4	1	2423906	102.6	1	2425181	93.1	1
10-14	2309587	105.3	1	2510361	102.9	1	2344275	95.2	1
15-19	2048208	97.0	1	2454284	106.5	1	2498572	99.2	1
20-24	1914992	103.2	1	2100064	96.6	1	2694646	106.9	1
25-29	1661866	98.7	1	1893200	102.4	1	2542682	107.5	1
30-34	1452053	98.8	1	1596760	95.9	1	2036206	95.8	1
35-39	1278644	103.2	1	1438418	101.8	1	1709347	99.4	1
40-44	1026535	98.3	1	1230423	102.5	1	1402328	96.5	1
45-49	809068	99.5	1	962658	96.2	1	1195740	99.1	1
50-54	599709	92.9	1	770704	101.8	1	1011349	100.7	1
55-59	482332	101.4	1	551102	90.5	1	811950	100.0	1
60-64	351752	89.1	1	447549	104.5	1	612364	100.9	1
65-69	307073	112.3	1	305169	90.1	1	401548	88.3	1
Missing	252292			0			0		0

Counts= PES adjusted census counts; PES= adjusted census counts' obtained age ratios; Expec= expected age ratios from undistorted census count

Table 4: Comparison of PES adjusted census' age ratios against those expected (females).

Age group	1996 census	Age ratio		2001 census	Age ratio		2011 census	Age ratio	
	Count	PES	Expec	Count	PES	Expec	Count	PES	Expec
0-4	2223343			2215008			2817867		1
5-9	2332033	101.8	1	2425994	102.0	1	2394570	94.5	1
10-14	2359138	105.6	1	2541811	102.6	1	2250611	91.9	1
15-19	2135672	96.5	1	2527782	106.9	1	2504905	101.6	1
20-24	2067653	105.3	1	2189344	96.0	1	2679896	106.7	1
25-29	1790412	97.2	1	2034172	103.5	1	2516635	107.7	1
30-34	1617576	102.2	1	1741231	94.9	1	1992804	93.2	1
35-39	1375399	101.0	1	1635554	104.9	1	1758420	99.4	1
40-44	1105325	98.7	1	1376879	99.7	1	1546291	97.2	1
45-49	863268	97.5	1	1125861	100.2	1	1424543	103.5	1
50-54	665172	91.7	1	870991	98.0	1	1206940	100.2	1
55-59	586841	97.6	1	650860	87.2	1	985458	99.5	1
60-64	537440	103.1	1	622623	109.8	1	773404	100.3	1
65-69	455615	110.6	1	483069	94.8	1	556257	90.7	1
Missing	236530			0			0		

Counts= PES adjusted census counts; PES= adjusted census counts' obtained age ratios; Expec= expected age ratios from undistorted census count

Sex ratio deviations from expected

Table 5: Comparison of PES adjusted census counts' sex ratios relative to those expected

Age group	Census 1996			Census 2001			Census 2011		
	Males	Females	Sex R	Males	Females	Sex R	Males	Females	Sex R
0-4	2215745	2223343	99.7	2214369	2215008	100.0	2867585	2817867	101.8
5-9	2340363	2332033	100.4	2423906	2425994	99.9	2425181	2394570	101.3
10-14	2309587	2359138	97.9	2510361	2541811	98.8	2344275	2250611	104.2
15-19	2048208	2135672	95.9	2454284	2527782	97.1	2498572	2504905	99.7
20-24	1914992	2067653	92.6	2100064	2189344	95.9	2694646	2679896	100.6
25-29	1661866	1790412	92.8	1893200	2034172	93.1	2542682	2516635	101.0
30-34	1452053	1617576	89.8	1596760	1741231	91.7	2036206	1992804	102.2
35-39	1278644	1375399	93.0	1438418	1635554	87.9	1709347	1758420	97.2
40-44	1026535	1105325	92.9	1230423	1376879	89.4	1402328	1546291	90.7
45-49	809068	863268.	93.7	962658	1125861	85.5	1195740	1424543	83.9
50-54	599709	665172	90.2	770704	870991	88.5	1011349	1206940	83.8
55-59	482332	586841	82.2	551102	650860	84.7	811950	985458	82.4
60-64	351753	537441	65.4	447549	622623	71.9	612364	773404	79.2
65-69	307073	455615	67.4	305169	483069	63.2	401548	556257	72.2
70-74	195351	286535	68.2	230193	396651	58.0	297145	453343	65.5
75-79	141218	238685	59.2	136967	231978	59.0	163691	317675	51.5
Total	**19239495**	**20850582**	**92.7**	**21405705**	**23362979**	**92.2**	**25188790**	**26581769**	**95.5**

Sex R= Sex ratio i.e. males per 100 females

References

Berkowitz, P (2012) "The (incomplete) (probably inaccurate) sum of us." Daily Maverick

Breiman, L (1994) "The census adjustment: Undercount or bad data?" *Statistical science*, Vol.9, No, 4, p 458-537.

Cronje, M and Budlender, D. (2004) "Comparing census 1996 a with 2001: An operational persipective" Southern African Journal of Demography, 9 (1), p 67-90

Dorrington, R. (1999) "To count or to model that is not the question: Some possible deficiencies with 1996 census results." University of Cape Town

Hogan, H (1993), "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of American Statistical Association*, 88, p 1047-1060.

Robinson, **J. G.**, Ahmed, B., Das Gupta, P and Woodrow, K. (1993) "Estimation of Population coverage in the 1990 United States based on Demographic analysis".

Keane, J, G, Balridge, M., Brown, C, J, and Jones, S. (1985) “Evaluating censuses of population and housing.” Resource support Services in agreement with US Agency for International Development

(Statistics South Africa 2010). “Census undercount and Strategies: Version No 6.” Statistics South Africa.

Steffey, D, l. (1997) “A Review of census undercount issue.” Department of mathematical and computer sciences, san Diego State University

Moultrie, T., Dorrington, R., Hill, A., Hill, K., Timaeus, I and Zaba, B. (2013) “Tools for Demographic Estimation.” UNFPA