Undercounting Controversies in South African Censuses

Jeremy Dickson Gumbo

Student No 395681

Supervisor

Professor Clifford Odimegwu

A Thesis submitted to the Faculty of Humanities in fulfilment of the requirements of the Degree of Doctor of Philosophy in Demography and Population Studies, University of The Witwatersrand, Johannesburg, South Africa

DECLARATION

I Jeremy Dickson Gumbo declare that this is solely my own work, and where otherwise appropriate referencing has been made.

I also declare that this Thesis is a recipient of The Vodacom's Wits University Top 8 International PhD Thesis Award for 2014.

Signed  **Date**: 26 March 2016

BIO SKETCH FOR JEREMY DICKSON GUMBO

Jeremy Dickson Gumbo, hereafter referred as Jeremy Gumbo, is a demographer employed by the Respiratory and Meningeal Pathogens Research Unit (RMPRU) at Chris Hani Baragwaneth Hospital where he leads two projects. The projects are the Child Health and Mortality Prevention Surveillance (CHAMPS) and the Healthcare Utilisation Survey (HUS). The CHAMPS tracks the causes of mortality among children under the age of 5 years in Soweto suburb of Johannesburg, South Africa. This surveillance study covers 25 000 households and its baseline population will be determined in January 2017. The follow up period is 5 years. The HUS survey investigates healthcare utilization also among children below the age of 5 years in the same area using participants from 500 randomly selected households.

Jeremy Gumbo is also a member of the teaching staff at the University of Witwatersrand's Department of Demography and Population Studies and the School of Public Health, a post he held since 2013. He has taught two courses at Undergraduate level namely Population Studies, and Population and Development. At Post Graduate level he has taught Masters and Honours Students in Demography and Population Studies (DPS) courses of Statistics for Social Sciences and Advanced Demographic Methods from the year 2013 to 2016. During the same period, Jeremy Gumbo also taught Masters Students in the courses of Intermediate Demography, and Introduction to Demographic Methods at Wits School of Public Health. Since 2013 he has supervised and graduated 12 Honours students in Demography and Population Studies programme. Currently [2016] he co-supervises three Masters Students with Prof Clifford Odimegwu in DPS.

Other positions of responsibility which he currently holds at the University of the Witwatersrand are Project Manager for the research on Labour migration, Social networks and HIV Risk in Gauteng, South Africa which started in November 2015. The project is jointly conducted by Wits' Department of Demography and Population Studies and the Department of Sociology of Illinois State University, United States of America. The data used in this project is collected from 5000 households sampled in Gauteng Province. Jeremy Gumbo has also been an Assistant Administrator at the Humanities Postgraduate Centre from March 2013 to March 2016. This is a research centre for the Faculty of Humanities.

Between March 2013 and June 2016 Jeremy Gumbo has been involved in 7 publications. The eighth paper which is drawn from the Thesis is on its way out. It received favourable comments from reviewers; suggested corrections have been addressed and the update version has been re-submitted to the *Southern African Journal of Demography*. The paper is titled Service Delivery Implications of Undercount in Sub Sahara Africa. Data from South Africa's last three censuses were used to estimate the effect of undercount on resource allocations, meant for service delivery in the region, using the PRICEWATERHOUSECOOPERS method. Findings were that undercount compromised fair resource allocations. Two other manuscripts from the Thesis have also been submitted to Journals. One is titled; "Coverage evaluation in South Africa's census 2011", and was submitted to *Southern African Journal of*

Demography in September 2015. The other one is titled; “Undercount patterns in South African censuses: Systematic and Consistent?” which was submitted to *African Population Studies (APS) Journal* in April 2016.

Manuscripts outside the Thesis that were accepted for publication include the following: **i)** “Power Dynamics and Men’s’ vulnerability to HIV infection in Sub Sahara Africa”, published in the *Gender and Behaviour Journal*. **ii)** “Multiple sexual partners and gendered differences in HIV testing among youths in South Africa”, also published in 2014 in the same journal. **iii)** “Is Swaziland census data suitable for fertility measurement?” *Genus* **iv)** “Estimation of Swaziland Fertility: What do the methods tell us?” The paper was accepted by the *Southern African Journal of Demography* this year. **v)** “Single motherhood and marasmus among under-fives in sub-Saharan Africa: a regional analysis of prevalence and correlates”, *South African Journal of Child Health*. **vi)** “Partner Age Difference and Fertility in Uganda”, *African Population Studies Journal*. **vii)** “Let’s leave together: Is cohabitation a precursor of marriage?”, *African Population Studies Journal*

With regards to research collaboration, Jeremy Gumbo has worked on the Alternative Formalities, Transnationalism and Xenophobia Report. The research was conducted under the University of the Witwatersrand’s School of Architecture. This project investigated the contribution of the informal sector to Johannesburg Metro City, following the eviction of informal traders by Johannesburg Metro City in 2013-14. Jeremy Gumbo worked on the chapter that investigated the demographic composition of Johannesburg City. He also worked on the Report on The State of Gauteng’s Population. The project was conducted by Wits University’s Demography and Population Studies Programme between mid-2015 and the first quarter of 2016.

On conference participation, Jeremy Gumbo presented at local, regional, and international conferences that included Population Association of Southern Africa (PASA), African Population Conference (APC), and Population Association of America (PAA). At PASA conferences he has presented and won The Best Oral Student Presenter’s Award in Bloemfontein, South Africa in 2011. The manuscript was drawn from his Masters Dissertation, and was titled: Is marriage a Viable Strategy for Reducing HIV Infection among Women in Zimbabwe? He also presented at the 2014 edition of the same conference a manuscript drawn from his PhD Thesis titled; Undercounting Controversies in South African Censuses.

At the 2011 edition of the APC in Burkina Faso he presented the same manuscript from his Masters Dissertation noted above. At the 2015 edition of the same conference, held in Johannesburg, South Africa, he presented two manuscripts from his PhD Thesis. These were titled; Service Delivery Implications of Undercount in Sub Sahara Africa; and Undercounting Controversies in South African Censuses. At PAA conference of 2013 he presented another manuscript from his Masters Dissertation titled; Risk of HIV infection among women in Zimbabwe: Does Living Arrangement Matter? The manuscript has seen been submitted to SAHARA journal after addressing final comments from reviewers. At PAA 2014 conference he presented preliminary findings from his manuscript Undercounting Controversies in South

African Censuses. At the PAA Conference for 2016 Jeremy Gumbo presented two manuscripts drawn from his Thesis. These were titled; Service Delivery Implications of Undercount in Sub Sahara Africa, and Coverage Evaluation of South Africa's 2011 Census. Jeremy Gumbo was also a discussant at the 2015 APC in 2015, for the session on Contraceptive Use and Unmet need.

Prior to joining Wits University in 2013 Jeremy Gumbo worked at Mott MacDonald International in the company's health research unit, as a Senior Research Associate. He handled the company's main research projects as well as overseeing progress in the other research projects which were handled by junior Research Associates. Some of the projects are; the United Nations Educational Scientific Cultural Organisation (UNESCO) project on Sexual and Reproductive Health of young adults in 21 countries from Eastern and Southern African (ESA), and the John Hopkins Health and Education South Africa (JHHESA).

With regards to his PhD studies Jeremy Gumbo enrolled with the Faculty of Humanities at University of Witwatersrand in 2013. The title of his Thesis is, Undercounting Controversies in South African Censuses. The context of the Thesis is that South Africa' censuses of 1996, 2001, and 2011 recorded high undercount estimates of 10.6%, 17%, and 14.6%, respectively. The Post Enumeration Survey (PES) was used for estimating and adjusting for the undercount. High undercount estimates were the reasons for controversies in these censuses as the accuracy of the PES was contested by various stakeholders. Studies around this discourse have however not investigated accuracy of the undercount estimates from the method by comparing them with respective undercount estimates from alternative methods. For, undercount estimates are direct outcomes of respective methods used in estimating the undercount. Hence, findings from such investigations should reduce controversies in these censuses. The main research question of Thesis was: how can controversies associated with South African censuses be reduced? The argument of the Thesis was that; findings from comparisons of respective PES and alternative methods' undercount estimates largely determine accuracy of the former.

Five complementary objectives that addressed the above research question were set for this study. Data was drawn from South Africa's three censuses noted above and was only available as adjusted; from both the 10% samples and full data from SUPERCROSS. The former were at national level, and the latter were at small areas level i.e. for the area covered by Agincourt Health and Demographic Surveillance System (HDSS). ArcGIS was used to extract respective census counts for area covered by Agincourt HDSS. Unadjusted counts were reconstructed from respective adjusted counts through multiplication of the latter by respective proportions of population enumerated. MORTPAK PROJCT programme was used for modelling constructed population estimates that matched respective censuses.

Objective one explored systematic patterns of undercount in the three censuses, and related them with those suggested from studies around the topic. Findings were there that the undercount patterns from the three censuses were generally systematic, and they were largely consistent with those from related studies. For objective two, findings were that despite some suggestions of some errors in these censuses, the counts were largely credible. From the tests for accuracy of adjusted and unadjusted counts which addressed objective three; as expected the findings indicated that adjusted counts closely matched respective counts from none census data compared to unadjusted counts. But in few instances this pattern was reversed. Findings which addressed main objective indicated that the respective overall undercount estimates for males, females and combined sexes from PES and Demographic Analyses closely matched. However, the two methods' respective undercount estimates indicated some discrepancies in the comparisons for age groups. The findings for the firth objective indicated that undercount compromised accuracy of demographics parameters and quality of service delivery.

Though the evidence from the study's findings was not sufficient enough to confirm accuracy of the PES in estimating and adjusting for the undercount, it was sufficient enough to suggest credibility of the method and its outcomes. This implied that the processes and outcomes of these censuses were reliable; and this should reduce the controversies in these censuses.

The message to Statistics South the custodians of census taking in the country is that; though the overall findings of the study suggested that the PES and its outcomes in the three censuses were reliable, there were also indications that the method may have ushered in some errors into the data. To improve quality of future censuses, attention should be paid to issues raised in prior studies, over the PES. In particular the sample size of the PES has been noted as too small and hence should have been one of the sources of error in these censuses.

DEDICATION

To my Lord Jesus Christ who gives me strength to pursue and reach out to my set goals.

ACKNOWLEDGEMENTS

Professor Clifford Odimegwu my mentor, Stella Gumbo, Boikhutso Maumane, Takalani Muloiwa, Nitiel Nyasha, Tapiwa Gerald, Natasha Tanatswa, Zvikombore, Simbarashe, Danzil Rufaro, and Elsie. Dr. Joram Gumbo, *mama* Naggie Olipha, *mai* Stella Makazvagwa, *mama* Portia, *mama* DJ Makoni. I'm also grateful to the following relatives who have worked closely with me during my studies; Farai Dube, Portia Gumbo-Dube, Victor Moyo, Stella Moyo, Tavona Gumbo, Lyton Shumba, Tembinkosi Magwaliba, and William Kandowe. All the friends I have made at Wits, students and staff in Demography and Population Studies, all PhD students I shared time with in the Carrels at Humanities Graduate Centre, Mapuleng Tsaori the Manager of the Gradcentre, Coleman Dube for updating me on GIS techniques, Jacob my friend at StatsSA Limpopo province, Pedzisai Ndagurwa, and Professor Eric Worby. The list cannot be complete without acknowledging my PhD Funders i.e. Hewlett Foundation, the Respiratory & Meningeal Pathogens Research Unit (RMPRU) at Chris Hani Hospital, and the Humanities Postgraduate Centre. I also acknowledge those institutions that gave me access to their data, i.e. Statistics South Africa and Agincourt HDSS.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Introduction

Census taking dates back to the era of the Roman Empire as can be drawn from the gospel of Luke Chapter 2, Verses 1-5. Under the Roman rule censuses were conducted to keep records for individuals that were eligible for conscription into the army. Later during the colonial era, censuses were conducted to capture individuals that were eligible to pay tax. Currently censuses are widely used in guiding efficient planning and fair resource allocation. Content error, which refers to recording inaccurate information on captured individuals, and coverage error, i.e. either undercounting or over counting of people in a census, presents challenges in achieving these goals. Coverage error is frequent in censuses, especially undercount, which is of interest in this study.

In countries that have a well-documented history of census taking like the United States of America, Canada, and China, there are indications that respective censuses recorded substantial numbers of people that were missed. Nigeria and South Africa are some of the countries in Africa where high undercounts have been recorded in censuses. The latter country, which is the focus of this study, recorded undercount estimates of 10.6%, 17%, and 14.6% in the last three censuses of 1996, 2001, and 2011 respectively. These high undercount estimates were the source of controversies that have been associated with the three censuses. The controversies centred on the accuracy of the Post-enumeration Survey (PES). Critiques argue that the PES has been inaccurate in estimating and adjusting the undercount in the respective censuses. For this reason, the accuracy of both the undercount estimates and adjusted counts drawn from this method has also been contested.

Some studies indicated that there has been an underestimation of children under five years and the white population, while overestimating the female population in censuses 1996 and

2001 respectively (Dorrington, 1999; Dorrington, 2002). Other studies have however suggested the migration theory to explain the deficit of white population in census 1996 (Centre for Development Enterprises, 1998). The increase in fertility suggested in census 2011 was questioned by other studies as indicating inaccuracy of these counts considering that the country's fertility trend was on the decline for many decades (Centre for Actuarial Research, 2012). A critique of these censuses from the public also commented on the suggested increase in fertility through social media by tweeting that 'The odd baby boom ...' (sarahemilyduff). However, other researchers are of the opinion that the counts of Census 2011 were a better estimation of reality than those from models now proven to be incorrect (De Wet, 2012).

Reports from the media also claim that two top officials from Statistics South Africa were dismissed after refusing to lower the undercount estimates for Census 2011 from the 18% which they had reportedly obtained, to 2% which the Statistician General expected (City Press, 2012). However, according to the Statistician General the two presented wrong results after committing methodological and computational errors (City Press, 2012); thereby indicating the two were dismissed for incompetence. Statistics Council member Professor Jacky Galpin of the University of Witwatersrand has also argued that these census counts were consistent with findings from the post-enumeration survey (PES).

Studies around this discourse of census controversies in South Africa have largely centred on contesting the accuracy of the PES and its adjustment outcomes. What these studies have not addressed is determining how the PES undercount estimates compare with respective undercount estimates obtained from alternative methods. Such findings would be vital in reducing controversies associated with these censuses. Since the controversies are centred on the accuracy of the PES and its adjustment outcomes, findings from investigations on the accuracy of the PES undercount estimates, complemented by findings from further investigations on the accuracy of adjusted counts from the three censuses should contribute to reducing the controversies around these censuses. This is because the findings would empirically establish the accuracy of the PES in determining undercount estimates and their adjusted counts.

The main research question for this study was: How can controversies associated with South African censuses be reduced? The other research questions were stated as: Is there a systematic pattern of undercount in South African censuses? What are the levels of content and coverage errors in South African censuses? Which is better to adjust or not to adjust for the undercount in South African censuses? Are the respective undercount estimates from the PES and Demographic Analyses (DA) different? What is the effect of undercount on demographic parameters and service delivery in South Africa?

The main objective was therefore to reduce controversies associated with South Africa's last three censuses. The specific objectives were to explore systematic patterns of undercount in South Africa's last three censuses, to estimate levels of both content and coverage errors in South Africa's last three censuses, to determine which census counts between adjusted and unadjusted are better estimates of South Africa's actual population counts, to compare the PES undercount estimates with alternative estimates from DA, and to estimate the effect of census undercount on demographic parameters and service delivery in South Africa.

<u>Data sources</u>

Data used in this study came from South Africa's censuses of 1996, 2001 and 2011. Non-census data was drawn from Agincourt Health and Demographic Surveillance System (HDSS), and constructed population estimates from MORTPAK's PROJCT programme. Census data is classified as adjusted if it has been corrected for both content and coverage error and unadjusted if not corrected for these errors. Adjusted data was obtained as either 10% samples from Statistics South Africa's website or as full data from SUPERCROSS. For public use Statistics South Africa (StatsSA), which is the custodian of the country's censuses, avail the adjusted data at national level down to Small Areas level (SAs). Data at Enumeration Areas level (EAs), which is the lowest level at which data is collected at group level, is not available for public use. This is largely for confidentiality reasons. However, StatsSA has the prerogative to provide this data upon request. For Census 1996, this study was given the permission to use the data at EAs level. The method of enumeration used in the three censuses is the De-Facto method. This is the enumeration of people according to the place where they are found.

Non-census data was obtained from Agincourt HDSS for dates corresponding with those of South Africa's three censuses. The surveillance site is located in the rural district of Bushbuckridge in Mpumalanga Province. The data is longitudinal, and information collect is on individuals' demographic and socioeconomic characteristics, and the updates are done on yearly basis. Temporary migrants are included among the enumerated population for the surveillance site despite the fact that they spend six or more months residing elsewhere. The baseline population for Agincourt HDSS was in 1992. The surveillance site is subdivided into villages. They were 20 at baseline, increasing to 22 at the time of Census 2001 and to 28 by the time of Census 2011. The response rate has been very high. For instance, in 2011 only 2 out of 15 000 households did not participate.

<u>Data processing</u>

Unadjusted counts were necessary for this study for both analyses at national and SAs level. Since StatsSA does not avail this data for public use the counts were reconstructed by multiplying the proportion of population enumerated with the respective adjusted counts. The proportions of population enumerated were invented from respective adjustment factors. At national level unadjusted counts were reconstructed from adjusted counts drawn from the 10% samples. At SAs level unadjusted counts for the area covered by Agincourt HDSS were reconstructed from adjusted counts drawn from SUPERCROSS, for area covered by the HDSS.

Non-census counts at national level were constructed using MORTPAK's PROJCT programme. This software was designed by the United Nations Population Development for various uses, including the construction of population estimates for respective countries using the cohort component analyses.

For census counts at SAs level, ArcGIS software was used to extract the counts for the area covered by Agincourt HDSS by overlaying small area boundaries on corresponding village boundaries for Agincourt HDSS. The overlays matched SAs (in the case of the 2001 and 2011 censuses) and EAs (for Census1996 there was no data at SA level) with coinciding villages from Agincourt HDSS. This enabled the extraction of adjusted census counts from SUPERCOSS for respective censuses for the area covered by Agincourt HDSS. Counts from the surveillance site were managed using STATA software. To enable comparisons, counts for temporary migrants in the surveillance site were excluded from the analyses. The reason was that the methodology employed in collecting census data did not include this subpopulation yet the HDSS methodology included them.

Data analysis methods

In achieving objective 1, the PES undercount estimates from the country's three censuses were compared with those suggested in related studies to check for consistence between the estimates from the two sources. For objective 2, content errors were estimated using the United Nations Age Sex Joint Score, Whipple, Meyers and Bachi indices. In testing for coverage error, adjusted counts for area covered by Agincourt HDSS were compared to respective counts from Agincourt HDSS. Other tests for coverage errors included growth rate analyses, age ratios analyses, sex ratios analyses, and graphical cohort analyses. In achieving objective 3; at national level adjusted and unadjusted counts were compared against respective constructed population estimates from MORTPAK's PROJCT programme. At small areas level, respective adjusted and unadjusted counts for the area covered by Agincourt HDSS were compared with respective counts from Agincourt HDSS data. The latter was treated as a gold standard because it is collected at a small areas level and is updated regularly.

For objective 4, overall national undercount estimates for males, females, both sexes combined, and for age groups obtained from PES were compared with respective undercount estimates computed using DA. Other methods for estimating undercount, like the Bayesian model, were beyond the scope of this study. For the fifth objective, demographic parameters from adjusted counts were compared to those from respective unadjusted counts. The counts were drawn from 10% samples, and the analyses were at national level. In testing for the effect of undercount on service delivery, the PRICEWATERHOUSECOOPERS method was used to compute and compare fund allocations from the Basic Services Grant (BSGF) for respective provinces. Furthermore, parliamentary seat allocations from adjusted and unadjusted counts were also computed and compared. Allocations from unadjusted counts indicate the undercount effect since they are not corrected for the error.

Results

Objective 1's main finding was that PES undercounts estimates for South Africa's last three censuses were consistent with those from related studies. For objective 2, tests conducted for content errors indicated that the three censuses had an insignificant amount with regard to this error. However, such findings have to be interpreted with caution as these methods also have limitations. In the test for coverage error, findings from comparison of counts and growth rate analyses suggested less coverage errors in the censuses. This was contrasted by findings from the other three tests which suggested significant coverage errors. It was also possible that the latter findings may have reflected the distortive effect of factors like migration and not necessarily coverage error. For objective 3, most of the findings from comparisons at both national and SAs levels largely confirmed that adjusted counts closely matched respective non-census counts relative to unadjusted counts. There were however some cases where findings suggested the opposite. For objective 4, the majority of the compared undercount estimates from PES relative to respective estimates from DA closely matched each other. Inconsistencies were noted from some of the compared undercount estimates for age groups. Finally, findings for the fifth objective indicated that failure to adjust for the undercount would have compromised accuracy of demographic parameters by either inflating or deflating respective rates. Also undercount would have led to unfair service delivery as some provinces would have been allocated resources unfairly, which were meant for those provinces that had been deprived by the undercount, had adjustments for the error not been done.

Conclusion

The findings should contribute in reducing controversies associated with South African censuses as they indicated that the PES and its adjustment outcomes were credible. None of the findings of the objectives were sufficient enough to confirm absolute accuracy of the method and its outcomes. However, as stated above, there were strong indications that the PES, its undercount estimates and adjusted counts were credible. This implied that the method and its outcomes are reliable.

Indications that the PES and its adjustment outcomes are credible were drawn from findings that undercount patterns drawn from PES and those suggested from related studies were largely similar. Such consistency suggests the credibility of the PES in estimating and adjusting the undercount in these censuses. There were also indications of less content errors from tests conducted, largely suggesting the method's credibility in correcting for the error. The main findings from tests for coverage error also complemented the above findings, though they were contradicted by findings from three other tests for the error. Findings from objective 3 also suggested the credibility of the PES because in most cases adjusted counts i.e. counts corrected for the undercount error closely matched respective counts from non-census data compared to respective unadjusted counts which had the undercount error. Findings for the main objective also indicated that most of the respective undercount estimates from the PES and DA were largely consistent. Differences were only observed in some cases in the comparisons of undercount estimates by age groups.

By indicating that undercount would have compromised the accuracy of demographic parameters and quality of service delivery if no adjustments were carried out, objective 5's findings should also contribute in reducing controversies around these censuses as the findings indicated that the PES adjustment outcomes improved the accuracy of demographic parameters and fair service delivery among provinces.

CHAPTER 1: INTRODUCTION

1.1 Background

Censuses are any country's biggest statistical collection process, and to underline their significance, conduction of censuses is often mandated in national constitutions (Cronje and Budlenger, 2004). In South Africa, the Statistics Act No. 6 of 1999 makes it mandatory for Statistics South Africa (StatsSA) to conduct a census after every five years despite the fact that a ten-year inter-census period is generally recommended (Cronje and Budlenger, 2004). In most cases national governments show commitment to censuses by including them under national budgets (Onsembe and Ntozi, 2006). A census involves enormous resource input, and the expenditure of resources arising is believed to be worthy as data collected is used to direct policy makers on circumstances and needs of a population (Cronje and Budlenger, 2004; Steffey, 1997).

Reasons for taking censuses have evolved over time. Among the earliest censuses, e.g. during the Roman Empire, the main purpose was for identifying individuals that were eligible for tax collection. The same reason has also been associated with census taking during the colonial era in Africa. As Onsembe and Ntozi (2006) have noted, census taking in colonial Africa was primarily for estimating people that were eligible for hut tax. In other cases, for example in the British empire in the 1800s, censuses were carried out in countries like Sierra Leone to investigate on alleged re-emergence of kidnappings for the purpose of slave trade (Onsembe and Ntozi, 2006). Early censuses in countries like the United States of America (USA) are believed to have been largely driven by the motive of military recruitment. Under such circumstances there could have been deliberate under coverage, e.g. of women and children, as such individuals were often not needed in the army. Today the key objective of census taking is to obtain an accurate estimate of the entire population for the purpose of guiding planners in formulating policies of good governance that ensure justice in service provision (Census in South Africa, 2012; Onsembe and Ntozi, 2006). Censuses provide socioeconomic and demographic information of individuals in a defined geographical area which respective authorities can use as reference material for evidence-based development instituting (Onsembe and Ntozi, 2006).

Though the primary objective in conducting censuses has always been to obtain an accurate population count for the targeted population, this has remained elusive for most census authorities. Censuses are susceptible 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 on characteristics of individuals who will have been captured in a census. Such errors include misclassifications by age, education, sex, and region, to mention a few. Content errors affect the distribution of population obtained from a census with respect to such characteristics (Keane et al., 1995). The other form of error is coverage, and this can either be undercounting or over counting. Coverage error also affects population distribution, just like content error. However, unlike content error, coverage error further affects overall population count. Irregular or distorted population distributions that can be noted from census data may not be clear as to whether they are due to content or coverage error, or even both (Keane et al., 1985). Therefore the estimation of the accuracy of census counts using techniques that rely on observing whether population distributions from investigated counts deviate from the expected standard distributions may not be valid unless the effect of content error has been controlled. It is partly for this reason that this study incorporated the investigation of content error among the set objectives.

Undercount, which is the focus of this study, is often more prevalent in census taking than over counting. For example, the United States of America's censuses of 1990 and 2001 recorded undercount rates of 1.6% (Hogan, 1993) and 1.2% (Statistics South Africa, 2010) respectively. In China, the world's most populous nation, the undercount rate for the 2000 Census was estimated at 1.8% (Anderson, 2004). Other examples of documented census undercount observed elsewhere across the globe are Mozambique's 1996 Census, 6%; Nepal's 2001 Census, 5.3%; and Canada's 1996 Census, 2.6% (Statistics South Africa, 2010). In countries like the United States of America (USA) undercount has often resulted in unresolved contests mainly between census and local government authorities, and the disputes often spilled into courts of laws (Breiman, 1994). Brown et al. (1999) noted that censuses in the USA have been subject to controversy since the country's 1980 Census.

Reasons for census controversies among countries have been different. For instance in the US, controversies around Census 1980 centred on the question of adjustment for the undercount or not (Schirm, 1991; Brown et al., 1999; Steffey, 1997). Yet for Census 2000 controversies were about sampling approaches to be used in designing the PES (Brown et al., 1999). In Chinese censuses, questions have often been raised over the suspicious 'missing girl child' (Liang and Ma, 2004). There have been suggestions that deliberate under reporting of the girl child for cultural reasons exist in the Asian nation where the preference for a boy child is higher. For this reason some scholars argue for the use of sources like administrative records as alternative sources of data for population estimation (Vincent, 1992). Yet some argue that such records are equally flawed to give an accurate population estimate. How census accuracy can be improved or alternative sources can be used to supplement census or provide alternative sources of population information has also become an issue of controversy (US Bureau of Census, 1993b).

In South Africa, the history of census taking can be traced back to over one and a half centuries ago, as suggestions exist that the first census was conducted in 1865 (Cape Town, Government Printer, 1866). The latest census was taken in 2011. During this period, there have been some discrepancies on how successive censuses have been conducted. Firstly, in terms of national inclusiveness, censuses conducted prior to 1970 were inclusive (Courbage, 2000) as have been those after 1991 (Cronje and Budlenger, 2004). However, censuses conducted between 1970 and 1991 excluded enumeration of people residing in 'homelands', i.e. where the majority of the black population resided (Courbage, 2000). For this reason it is likely that these censuses had the highest undercount records in the country's history of census taking. As Courbage (2000) further noted, the 1980, 1986, and 1991 censuses may have excluded about 20%, 21% and 17% of the black populations respectively. In these censuses, despite many black people being excluded, the undercount was still high at about 9.7%, 9% and 12.8% respectively (Central Statistical Service, 1991). Across races, black people always had the highest undercount in those censuses, 22.5%, 20.4% and 16.8% respectively, as further observed by the same source. Such a high undercount among black people should be expected, as enumeration in some homelands was conducted through estimations from aerial photographs, as authorities considered these areas too dangerous to visit for enumeration (Cronje and Budlender, 2004).

The de facto method has always been preferred in South African censuses above the de jury method. In Census 1991 and those prior to this, the reference point has been the night between 7th and 8th of March (Central Statistical Service, 1991). The country's last three censuses still used the de facto method, but the reference point changed to the night of the 9th of October. A new feature introduced in the last three censuses was an option of leaving a questionnaire for the household to complete and send by post (Cronje and Budlender, 2004). In prior censuses, only face-to-face interviews were used. There has also been gradual integration of modern techniques in census taking processes, especially since Census 2001. For example, there has been an increase in the use of the Geographical Information System

(GIS), video training, and barcodes for questionnaires (Cronje and Budlender, 2004). Such technological advancements were expected to improve census outcomes.

Testing and adjusting the undercount has remained a permanent feature in South African censuses. Censuses before 1980 tested for undercount, but did not adjust for the undercount. The first census to test and adjust for undercount was in 1980 (Central Statistical Service, 1991). As for the 1985 Census, after determining the undercount it was the enumerated count that was published. In the 1991 Census, both unadjusted and adjusted census results were published (Central Statistical Service, 1991). The last three censuses only officialised adjusted data for public use. StatsSA is regulated to withhold unadjusted data, and it can allow access to limited fractions of this raw data to individuals or organisations upon request (StatsSA, 2011).

This study was confined to the country's last three censuses. The censuses recorded high undercount estimates of 10.6%, 17% and 14.6% for censuses 1996, 2001, and 2011 respectively (StatsSA, 2011). The high undercount estimates became the source of controversy in these censuses. Critiques questioned the accuracy of the PES in estimating and adjusting the undercount in South Africa's last three censuses (Dorrington, 1999). In particular they argued that the sample size used when conducting the PES in South Africa was too small, and as such introduces further bias in the censuses (Moultrie and Dorrington, 2012).

This was compounded by the fact that in South Africa this survey is conducted three or more weeks after census enumeration. Hence, the method's underlying assumption that the population characteristics prevailing at the time of census enumeration remained the same by the time of conducting the PES does not seem to hold for South Africa as the country's population is believed to be highly dynamic. For such reasons, the accuracy of undercount estimates and adjusted counts in the PES became equally contestable (Dorrington, 1999; Dorrington et al., 2004; Moultrie and Dorrington, 2004). This is because they are linked to the PES in that the undercount estimates are directly determined by the PES, and the adjusted counts are computed using respective undercount estimates.

1.2 Problem statement

South Africa's undercount estimates are high in relation to other developing countries. Undercount estimates for the majority of developing countries' censuses range between 2% and 5% (Berkowitz, 2012). This is contrary to South Africa's estimates which have always been above single digit. At provincial level, the undercount estimates were not homogenous. For instance, in the 1996 Census the Northern Cape recorded the highest undercount of 15.6%, followed by KwaZulu-Natal with 12.8% (StatsSA, 2011). In the 2001 Census, KwaZulu-Natal had the highest undercount of 22.5%, followed by Gauteng at 18.7% and the Free State at 17.6% (StatsSA, 2011). However, in Census 2011, the Western Cape recorded the highest undercount of 18%, whereas Limpopo had the lowest at 10% (Schultz, 2013). Undercount also varied among the country's four main races of black, white, coloured and Asian people in these censuses. In the 1996 Census, black people had the highest undercount of 10.8%, whereas in the 2001 and 2011 censuses white people had the highest undercounts of 22% and 15.8% respectively (StatsSA, 2011).

High undercount estimates led to unresolved contests around the accuracy of the PES in estimating and adjusting the undercount in the three censuses. The main problem has been that none of the related studies investigated the accuracy of the PES undercount estimates, e.g. by comparing them with those from other methods. Failure to investigate the accuracy of the undercount estimates in these censuses has partly led to persistence of these census controversies. Studies conducted have mainly questioned the accuracy of the adjusted counts in the PES (Dorrington, 2002; Moultrie and Dorrington, 2012; Centre for Actuarial Research, 2012). Others studies also question the accuracy of the PES, suggesting that the sample sizes used to conduct these surveys have been small, thereby introducing further errors into the data (Moultrie and Dorrington, 2012). Findings from such studies have suggested that adjusted counts in these censuses reflected internal inconsistencies and incompleteness (Moultrie and Timaeus, 2002).

Other researchers have raised concern over the method's high undercount estimates that were recorded in these censuses, which some argue could be an indication that the error may be higher (Gernertzky, 2012; Schultz, 2013). The high undercount estimates are inconsistent with the high funding that these censuses received (Gernertzky, 2012; Schultz, 2013). For such reasons others have questioned the reliability of these censuses. For instance, they argued that the PES-adjusted counts favoured certain provinces, like the Western Cape, in Census 2011 (Berkowtz, 2012). However, other researchers are of the opinion that these counts are a fair estimate of the country's population (Centre for Development Enterprises, 1998; De Wet, 2012).

A less recognised effect of census undercount is its potential to compromise the quality of national surveys where the sampling frame is based on census counts (Raley, 2002). In South Africa, such surveys include the Demographic Health Surveys (DHS) which were last conducted in 2003. Other notable surveys include the National HIV Communication Surveys (NCS), and the Nelson Mandela/HSRC Study of HIV/AIDS. Census undercount does not only affect the quality of census data, but also that of national surveys as inaccurate census data is a source of inappropriate sampling frames for respective surveys. The effect is particularly significant in cases where the emphasis is on ensuring that surveys are representative of the population they are drawn from. A member of the public using the pseudonym ‘Anecdotal’ critiqued the census and tweeted that South African censuses are highly erroneous and are merely politically contrived, a fact that should lead to bias in the sample designing of national surveys. Such a view is credible due to the fact that surveys like the South Africa Demographic and Health Surveys (DHS) for 1999 and 2003 produced contested outcomes, partly explaining the temporary suspension of these surveys.

In the same context, if for e.g. it is suggested that the 2001 Census adjustments underestimated the white population (Dorrington, 2002), this would imply that the national surveys following this census will not be truly representative of race. Erroneous censuses and surveys do not only lead to inaccurate estimates of population distributions, but also to inaccurate estimates of events such as infant and child mortality, total fertility, and in-moving and out-moving migrants, and this can create serious planning problems (Steffey, 1997). Furthermore, South Africa is believed to house the biggest HIV population in the world, of about 5, 6 million people (UNICEF, 2012), and the pandemic is believed to be more prevalent among young people, who are expected to be highly sexually active. Undercount of such a subpopulation group, e.g. as claimed by some researchers in the case of Census 1996, is likely to lead to an underestimation of the HIV burden in South Africa.

In South Africa, as observed elsewhere across the world, undercount is often higher among needy subpopulation groups (Pillai and Barton, 1992) such as children aged 0-4 years. Further research has noted that undercount among this group tends to be higher among minorities, where 50% of children are likely to be missed compared to 14% of the elderly population (O’ Hare, 2009). Children are among the most vulnerable members in any population, especially in terms of risk with regard to ill health. Accurate population counts for this group are therefore important, as an undercount may result in inadequate service provision being channelled towards this needy group (Pillai and Barton, 1992). National governments often roll out programmes aimed at improving the welfare of children, e.g. the Children Health Insurance programme in the USA and immunisation programmes in South Africa. In the USA, the former programme received $5, 7 billion based on the 2000 Census estimates (O'Hare, 2009). Supposing that there was no undercount correction, this will imply that the missed children will have to share from an allocation which did not include them. The main problem in such cases is that the whole subpopulation group would be affected, as the budgeted funds become inadequate. As Schirm (1991) pointed out, all the members of the subpopulation group suffer when a significant number of their fellow members are missed in a census. Moreover, if for example the high estimate of the elderly population in the 2011 South African Census is an overestimation, as some researchers have already suggested, this might imply resource misdirection to a non-existent elderly population, at the expense of undercounted and needy children.

Furthermore, the extent and patterns of an undercount are also expected to have implications for demographic parameters and service delivery (Johnson, 2012). Undercount may lead to erroneous estimates of fertility, mortality and migration being obtained from census data. For instance, undercount of children in Census 1996, as suggested by research (Dorrington 1999), may result in low fertility estimates. In the case of mortality, undercount of population groups may result in inaccurate mortality patterns and rates for respective population groups. For instance, this may compromise mortality rate ratios between compared subpopulation groups. South Africa's last three censuses recorded higher undercount estimates overall among the black population in comparison to other races (StatsSA, 2011). Research findings in census data that is largely affected by such coverage error are likely to produce inaccurate findings which could lead to wrong decisions and policies.

The quality of service delivery is largely a function of resources allocated to subpopulation groups based on estimated population counts. For instance, in the United States of America parliamentary seats and state funds are allocated on the basis of population estimates from censuses (Schirm, 1991). In South Africa about R350 billion is shared among provinces, partly or solely based on census counts for respective provinces (City Press, 2013). Research has observed that provincial population estimates from Census 2011 were highly defective as they did not correspond with those from mid-year population estimates. For example, 2011 mid-year population estimates for Gauteng and the Western Cape were lower than estimates from Census 2011 (Berkowitz, 2012). On the other hand, provinces like the Eastern Cape had lower population estimates from the same census compared to those predicted by the mid-year population estimates. Higher population estimates for provinces like the Western Cape and Gauteng in Census 2011 meant higher fund allocations for the two provinces, and a concomitant budget cut for the Eastern Cape (Berkowitz, 2012).

An example of the United States of America demonstrates how undercount can affect service delivery. Fund distributions based on 2000 census counts found that undercount could have resulted in reallocation of $4.2 million out of a total of $5.7 billion (Johnson, 2012). Such reallocation of funds is likely to negatively affect service delivery among states that are losing the funds because of the undercount, yet unfairly benefitting others. For example, a report prepared by PRICEWATERHOUSECOOPERS (2000) for eight major federal funding grants worth $145 billion came up with projected findings that the 2000 Census undercount would have cost 31 states and the District of Columbia $4.1 billion between 2002 and 2012 (Johnson, 2012). On average, if each of the 31 states and the single district were receiving equal funding, each one is expected to have lost $125 million of federal grant in the projected period. Such substantial amounts of funds would significantly contribute to improving service delivery among the respective states and district. Equally the same, less parliamentary seat allocations expected to result from undercount should also affect service delivery. More parliamentary seat allocations per area often mean better service delivery as funds for development are often channelled through parliamentary constituencies. For example, in Zimbabwe, each parliamentary constituent is allocated $50 thousand each year for development purposes under the Constituent Development Fund. Such grants could be higher in countries with better economies like South Africa.

1.3 Purpose statement

The purpose of conducting this study is to compare respective PES and DA undercount estimates from South Africa's last three censuses. This is because of the controversies that have trailed census enumeration in South Africa since 1996. The controversies centred on the accuracy of the PES in estimating and adjusting the undercount in the three censuses. Among related studies conducted around this discourse of census controversies in South Africa none have investigated this subject.

This study argues that the comparison of undercount estimates from the PES with respective undercount estimates from alternative methods is vital in determining the accuracy of the PES. Findings from this investigation complemented by those from related objectives that were also set for this study should help in determining the extent of PES accuracy in estimating and adjusting for the undercount.

1.4 Research questions

The main research question is:

How can controversies associated with South Africa's last three censuses be reduced? The specific research questions are:

i. Is there a systematic undercount pattern in South African censuses?
ii. What are the levels of content and coverage errors in South African censuses?
iii. Which is better: to adjust or not to adjust the undercount in South African censuses?

iv. Is there a difference between respective undercount estimates from the PES and Demographic Analyses (DA)?

v. What is the demographic and service delivery effect of census undercount in South Africa?

1.5 General objective

The main objective of the study is to reduce controversies associated with South Africa's last three censuses.

1.5.1 Specific objectives

The specific objectives are as follows:

i. To explore systematic patterns of undercount in South Africa's last three censuses.

ii. To estimate levels of both content and coverage errors in South Africa's last three censuses.

iii. To determine which census counts between adjusted and unadjusted are better estimates of South Africa's actual population counts.

iv. To compare the PES undercount estimates with alternative estimates from DA.

v. To estimate the effect of census undercount on demographic parameters and service delivery in South Africa.

1.6 Definitions

Adjusted data: Census data corrected for missed individuals during enumeration

Adjusted census count: Population counts from adjusted data

De facto census: Enumeration of individuals according to place they were captured

De jury census: Enumeration of individuals according to place of usual residence

Reconstructed counts: Estimates of unadjusted census data rebuilt from adjusted data using mathematical methods

Constructed population estimates: National population estimates that were constructed using MORTPAK projections

Service delivery: Provision of services as expected by a population, by authorities. This is mostly measured in terms of excellence and efficiency.

Undercount: Recording of less people during census enumeration relative to the actual estimate

1.7 Justification

To the knowledge of this researcher there is no study around the discourse of census controversies in South Africa that compared undercount estimates from the PES with respective undercount estimates from alternative methods. Controversies associated with South African censuses are largely centred round the undercount estimates as the undercount estimates are determined by the PES; hence if they are inaccurate that implies the PES is also an inaccurate method. On the other hand, it also means the adjusted counts are inaccurate since these counts are computed using respective undercount estimates. Therefore the key justification of this study is that it should be the first one to conduct this vital comparison. Furthermore, findings go a long way in determining the quality of the PES undercount estimates, thereby significantly contributing to the reduction of controversies associated with South African censuses.

The other justification of this study is that it determined which census counts between adjusted and unadjusted are better estimates of South Africa's actual population. Again this is an investigation that related studies have not focused on. Findings from this investigation guide census authorities in making informed decisions in future on which counts to declare as official. Furthermore, they inform users of census data on the caution they may need to take, where necessary. Findings from this investigation have direct benefit to the general population as better census counts are determined for use, e.g. when planning relative to benefits that may be expected from mere contestations over accuracy of counts. As noted by Motho Batho who commented via Twitter and was quoted in the Ma*il & Guardian* (2013), pointing fingers won't solve problems; rather what is needed is to find solutions. Furthermore, to the knowledge of this researcher, the issue on which is better, to adjust or not to adjust in South Africa's censuses is an unexplored matter. This study has the potential to provide the first contribution in this regard.

Investigations on this issue are also important for census authorities in South Africa since there are indications that they have been indecisive about which census counts should be official. As noted in the early sections of this study, census authorities in South Africa have been inconsistent when deciding on which census counts to declare as official (Central Statistical Service, 1991). For instance, they opted for unadjusted in all censuses prior to 1980, then for both adjusted and unadjusted in Census 1991, and for adjusted in censuses 1980, 1996, 2001, and 2011 (Central Statistical Service, 1991; StatsSA Census, 2011). Such inconsistencies indicate large uncertainty of authorities about which census data should be treated as official. This predicament is particularly demonstrated in the case of Census 1991 where both data sets were officialised. This study provided empirical evidence that informed census authorities in the country about which census counts were better.

A further justification of the study is that it investigated the effect of undercount on demographic parameters and service delivery. This issue has not been examined in South Africa. Hence again, this study is likely to be the first one to make a contribution with regard to this other issue in South Africa. Researches from other countries, especially the USA, confirmed the negative implications of undercount on some of these outcomes (Steffey, 1997). For instance, health planners depend largely on estimates of demographic parameters from census data to map out health policies, as well as rolling out healthcare services. Knowledge, e.g. of prevailing fertility, mortality and migration rates in the country especially from census data, is vital for achieving this. In this regard the study is significant because it highlighted how undercount can compromise the quality of demographic parameters.

With regard to the effect of undercount of service delivery, the study is important because it provided empirical evidence on how the error leads to inaccurate funds and parliamentary allocations among the country's provinces. This is important because census authorities are informed on the potential that undercount has, e.g. in creating despondency in a population. As noted earlier, the error does not occur homogenously across subpopulation groups, and prejudiced groups may become hostile. By alerting census authorities about the heterogeneous effect of undercount of subpopulation groups, census stakeholders are informed of the need to effectively strategize regarding the eradication of the error, in other words to avert such negative responses from affected subpopulation groups. By successfully strategising against undercount, the entire population benefits fairly from censuses.

An unresolved issue in this undercount discourse, as already noted, is the accuracy of adjusted counts. This study is therefore important in that it further pursued the issue and hence also contributed to the body of literature around the subject. The study conducted the investigation on the subject using data that has been rarely used in investigating the issue. Agincourt HDSS data is at small areas level, and is regularly updated since it is longitudinally collected. Hence the data should have good coverage of participants. For this reason, the analysis plan for investigating accuracy of adjusted counts by comparing them with respective counts from Agincourt HDSS is sound.

A further justification of the study is that it explored undercount patterns in South African censuses. This is important particularly for census authorities as the findings highlight which subpopulation groups are susceptible to undercount. This information is vital as it guides census authorities when they strategize against undercount. Undercount does not occur heterogeneously across subpopulation groups, and this implies that strategies to combat the error have to be specific to particular subpopulation groups that are consistently being missed. For example, the reasons for undercount among migrants are not expected to be the same as those for undercount among young male adults. Hence strategies to reduce the undercount among these different groups ought to be different.

1.8 Thesis structure

The thesis has 11 chapters.

Chapter 1

The chapter starts with a background that describes the context of the study. This is followed by the problem statement which articulates issues of concern from set research questions and the respective objectives. This is followed by the purpose statement which articulates the thesis argument, and then the specification of research questions and their respective objectives. This is followed by the section on the definition of terms and delimitations. The Justification section which describes why it was necessary to do this study follows. The last section of the chapter outlines the thesis structure.

Chapter 2

This chapter covers the review of literature from related studies. The first section of the chapter gives a background of the census-taking process, followed by a review of literature on key aspect of the topic, i.e. undercount, and the census controversies in South Africa. The rest of the sections cover literature reviews of specific objectives. The last section of the chapter discusses deficiencies in literature around the discourse of census controversies in South Africa.

Chapter 3

This chapter is on data sources, processing, and analyses. It starts with describing the types of data used in this study, followed by the various data processing that was done, i.e. reconstruction of unadjusted counts, construction of population estimates from MORTPAK, and extraction of adjusted census counts for area covered by Agincourt using ArcGIS. This is followed by methods of analysis used to achieve each objective.

Chapter 4

This chapter describes results for objective 2 on estimating levels of content error. The error was estimated using various indices that are known to be sensitive to the presence of this type of error.

Chapter 5

The chapter describes results for the second part of objective 2, i.e. the estimation of the level of coverage error. The chapter starts by describing how census counts for area covered by Agincourt HDSS were mapped using ArcGIS. This is followed by results from the comparison of these counts with respective counts from Agincourt HDSS. This is followed by other results from growth rate analyses, age ratio analyses, sex ratio analyses, and graphical cohort analyses respectively.

Chapter 6

This chapter reports the results that address objective 3, by comparing census counts i.e. adjusted and unadjusted with counts from non-census data. The comparisons are done at both national and small areas level.

Chapter 7

This chapter presents the main findings of this study. They answer objective 4. Recorded findings are taken from the comparison of undercount estimates from PES and those from DA. These findings largely determine how accurate the PES has been as a method for undercount estimation and adjustment.

Chapter 8

This chapter reported part of the results for objective 5, i.e. on the effect of undercount on demographic parameters.

Chapter 9

The chapter concludes the reporting of the remaining results for objective 5 and covers the effect of undercount on service delivery.

Chapter 10

This chapter presents a discussion of the results. It starts with an introduction which outlines the structure of the chapter, and the issues that were considered in the discussion, e.g. reviewed literature, discourse around topic, and limitations. Then results of each chapter are discussed separately in subsequent respective sections starting with those for the first objective and lastly those for the last objective. Lastly, there is a discussion of findings relative to the discourse around the topic.

Chapter 11

This chapter rounds up the research with a conclusion section as well as recommendations. The conclusion emphasised how findings provided answers to research questions. In the recommendations it was suggested how censuses can be improved.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The process of census taking involves preparation of questionnaires, personnel selection and training, and statistics compiling, checking, and analysis of data (Steffey, 1997). Each stage could introduce errors in a census, and this should explain why censuses are rarely accurate. Yet in essence an accurate estimate of the size of a population is usually the prime demographic outcome that any government is often interested in from a census. Errors in censuses impede accurate census counts; hence census officials often use different methods of enumeration that they believe suit their respective contexts. For example, some countries deliver enumeration questionnaires by post and respondents also send them back by post after completion. Yet in other instances respondents complete and submit their questionnaires online. Countries using such methods include the United Kingdom, Wales and Scotland (Office of National Statistics UK, 2011).

In other countries, particularly those from Eastern Europe, enumerators carry out census interviews and complete questionnaires on behalf of participants. France runs rolling censuses, which involves basing census results on moving averages worked out over a five-year cycle which has been updated every year using other non-census data (Office of National Statistics UK, 2011). South Africa uses different ways of both delivering and collecting the questionnaires. Respondents have three choices: completing and submitting questionnaires online, receiving and returning completed questionnaires through the post, or a face-to-face interview with a census enumerator. The different methods of census enumeration in different countries partly indicate the respective census authorities' desire to apply methods that are suitable for their country's circumstances, and particularly for the purpose of reducing coverage errors.

2.2 Undercount

Very few countries in sub-Saharan Africa have conducted censuses consistently. (Onsembe and Ntozi, 2006). The main explanation for this is their lack of resources, specifically financing. In countries that have conducted censuses consistently, like Ghana, South Africa, Zambia, Zimbabwe, and Nigeria, only South Africa has completed the undercount measurement and correction consistently in their censuses. Undercount measurement in Nigeria dates back to the 1952 Census where they obtained a 10% undercount estimate (Lalasz, 2006). With regard to censuses conducted later, undercount estimates went above 20%, which partly led to contestation, protestation and at times nullification of the entire census, as further noted in the same source. Zambia's first attempt to measure the undercount in the 1990 Census was a failure and the findings were never released (Banda et al., 2013). The Zambian census authorities sought assistance for technical support in the 2010 Census from the US Census Bureau. This time the focus shifted from implementing the PES for the purpose of determining the undercount to capacity building through training on software packages and developing statistical knowledge (Banda et al., 2013) for future censuses. Ghana introduced the PES for the purpose of measuring the undercount in 2010, but the results were not used to make adjustments. Zimbabwe incorporated the PES in Census 2011, as did Tanzania, to test for the undercount, but the majority of sub-Saharan African countries have rarely tested for the undercount.

In countries that have regularly tested for the undercount, indications have been that this type of error is persistent despite improvements in census-taking methods (Anderson, 2004). For example, concern over census undercount in the US dates back to the early 20th century, though census authorities then generally assumed that everyone in the country had been counted (Steffey, 1997). However, over time there has been an increase in awareness of the negative implications of undercount among the public in America. This is partly due to improvements in census marketing, improvement in census-taking procedures and techniques for measuring the undercount and increasing political salience with regard to population statistics (Steffey, 1997). Public awareness about censuses, particularly their concern about the undercount, can be ascribed to improved marketing strategies that have been incorporated in modern census-taking processes. Such strategies are designed to make the census brand appealing to members of the public. This partly involves census authorities articulating the benefits of an accurate census count to the public. A population that is aware of the significance of a census is likely to question census results that are assumed to be inaccurate. Equally, the same development of better procedures for measuring the undercount makes the public highly expectant of better census estimates. Increased political salience with regard to population statistics means that members of the public can interpret the advantages and disadvantages of the undercount, especially with regard to issues like resource allocation and service delivery.

In sub-Saharan Africa, South Africa is among the leading countries that have modernised their census taking over the past decades. This is a measure that is expected to minimise coverage errors like undercount. Head counts of household heads at administrative offices became a thing of the past in South Africa more than a century ago. Modern census procedures which include the use of cartographic maps, geographic information systems (GIS), modern technology in data manipulation, and increased census funding, have been incorporated in South African censuses (Onsembe and Ntozi, 2006). According to the same source, the country is among the few African countries where cartography sections have become permanent units of national statistics offices. This is beside the fact that South Africa is the only country in SSA that has the capacity to afford census expenditures from its own national coffers. GIS and cartographic mapping improve speed, efficiency and quality of demarcation of census enumeration areas. Maybe partly for this reason South African censuses have always been carried out as scheduled. The use of modern technology in data manipulation, as in South African censuses, is expected to reduce human error, improve speed and efficiency in data processing. However, others are of the opinion that technological advancement in census taking can also contribute to substantial bias in census outcomes. For instance, enumerators may take long to master the efficient use of electronic devices. At times, devices like iPads are stolen, they crash or become damaged before data has been downloaded and saved.

When estimating the undercount comparison of counted people against the actual number of people in a population is recommended, but the latter is a parameter that can never be known with certainty (Anderson, 2004). For this reason various techniques that estimate the undercount have been devised. For example, from 1880 until 1950 the undercount in US censuses has been estimated using the DA technique (Steffey, 1997). This technique may involve the use of non-census data, e.g. vital records of births and deaths and comparing them with the census data to check for similarities or differences, for instance in the case of population distribution by age, sex, and race.

One of the weaknesses of this method of evaluating the undercount is that the technique cannot provide the assurance that variations of population distribution are primarily a result of errors in a census as opposed to errors in the standard measurement (Keane et al., 1985) as it is possible that the variations could actually be the result of errors in non-census data that are being used to evaluate a census. As researchers have noted, very few independent sources contain adequate and consistent data (Robinson et al., 1993). Also since the technique uses aggregate data from census and non-census sources, estimates do not identify the separate effects of e.g. omissions, duplications or erroneous inclusions and reporting errors like by age, sex and race (Robinson et al., 1993). Variations that could be noted between population distributions based on a census and those based on independent sources may be a result of reporting bias in a census. In such a case the variation in population distribution may be wrongly interpreted to suggest evidence of a coverage error, yet in actual fact this is the result of content error in the census. Individuals will have been captured in a census, but their characteristics like age, sex or race will have been recorded erroneously.

The USA's census authorities introduced the use of the PES as an alternative means of estimating the census undercount at individual level in 1950 (Steffey, 1997). This method was first used in the 1996 Census in South Africa and it has been used in the successive two censuses. The technique involves drawing a representative sample from the census and re-interviewing individuals from sampled households. The climax of the process is the matching of the census and PES outcomes, where individuals and households identified by the PES but not in the census are classified as having been undercounted. Just like other DA techniques, the PES has its own limitations and these have been highlighted in the introduction.

2.3 Methods of measuring the undercount

2.3.1 Post-enumeration Survey

The South African census conducted since 1996 determined the undercount using the Post-enumeration Survey (PES). This is a survey carried out after a census on the same population (Keane et al., 1985). Ideally the PES has to be carried out soon after completion of the census. This is primarily done to ensure reduced chances of change in the population's characteristics between the census date and PES date. The underlying rationale is that responses in the PES should perfectly match those in the census, a scenario that would suggest that the census was accurate (Keane et al., 1985). When the matching is below 100%, this suggests a census undercount, and matching estimates above 100% suggest over counting. The PES can make an effective evaluation of a census if it is conducted using the same procedures and conditions of the census (Keane et al., 1985). These include the same response-obtaining methods, response-recording methods, and questionnaires used. The formula used to compute the undercount is stated below.

$$\textbf{Undercount rate} = \frac{\textbf{Omission}}{\textbf{PES population}} * \textbf{100}$$

However, matching outcomes rarely yield 100% resolved cases as final undercount estimates in South African censuses are not solely based on matching results. For example in the 1996 Census, 78% of cases were successfully matched, i.e. resolved, meaning a substantial 22% cases were unresolved (StatsSA, 1998). Decisions on whether unresolved cases should be treated as counted or not cannot be made merely based on respondents' answers as to whether or not they have been counted because of the possibility of errors like recall bias. Each person sampled for the PES had to be allocated a score of probability of having been counted that lies between 0 and 1. The latter would mean the person was really matched, yet the former would mean the person was clearly missed (StatsSA, 1998). Concerning unresolved cases, computation of the probability of having been enumerated was done using a multivariate technique known as Chi-square Interaction Detection (CHAID). The estimation of the probability to be attached to an unresolved case was entered using the characteristics of resolved cases (StatsSA, 1998) to create a binary-dependant variable that categorises individuals as either having been counted or not counted. Explanatory variables predicting the probability of having been counted as opposed to having been missed included the following: whether or not the respondent identifies the person as having been counted, type of enumeration area, race, gender, age group, and household (StatsSA, 1998).

The PES's major strength in census evaluation is that it is a direct method, and research has identified that direct methods often entail less labour than indirect ones (Keane et al., 1985). In most cases indirect methods are used when direct methods are not feasible. However, the PES has its limitations. For instance, there is an assumption that population characteristics remain the same during both the time of conducting the census and the PES. Populations are known to be highly dynamic, especially in countries like South Africa where there is a lot of migration, both national and international. Furthermore, the PES is not likely to produce any different finding than a census, especially among the hard-to-find people, as someone that is hard to find in a census is likely to be hard to find in a PES (Freedman and Navidi, 1986). The same researchers have further noted that the independence of being counted in a census versus that of being counted in a PES is an assumption that has remained open to serious questioning.

2.3.2 Bayesian model

An alternative measurement of the undercount at micro level that developed after the PES is the Bayesian model. This method attempts to overcome the use of census data on migration, which is a major challenge with regard to the PES. The method is founded on the belief that most reliable counts of people are births and deaths, and attempts to measure undercount without relying on data on annual migration (Redfern, 2001). Immigrants are drawn from local censuses. However, emigrants are based on population figures of people born locally, but are captured as residents in other countries' foreign censuses. The strength of this method is that it estimates the undercount at micro level like the PES (Redfern, 2001).

The method also applies to countries that have reliable statistics on births and deaths (Redfern, 2001). This is the challenge in the South Africa censuses which did not record the mortality data for censuses 1996 and 2001, and even in the 2011 Census the mortality data was incomplete. Worse still the method draws emigrant counts from censuses in other countries and these come from unadjusted data. However, counts of emigrants obtained from other countries are only for adjusted data that is obtained from respective countries' statistical agencies, and or from IPUMS. Moreover, countries that provide emigrant counts are often too many, for example in the case of the UK the model used 224 countries when estimating the undercount for Census 2000. In the case of South Africa, there are more countries receiving South Africans since this is a developing country, in comparison to the UK. For such reasons this method was beyond the scope of this study.

2.3.3 Demographic analysis

Undercount at macro level can be estimated through demographic analysis; for instance census counts at national, provincial or even district level can be compared with those from non-census data (Redfern, 2001). This might involve comparisons of sex, age, and/or race distribution between the different data sets to see if they match. Besides direct comparison of census data with non-census data, the population outcomes from the former can also be compared to theoretical distributions. The underlining fact will be that due to biological and cultural factors national population distribution follows well–known and fairly predictable patterns (Keane et al., 1985). For example, the population distribution of males compared to that of females at younger ages can be expected to show a high sex ratio, and this should decrease with an increase in age. Any departures from such theoretical population distribution may be interpreted to signify the possibility of error, which could partly be a result of undercount. Besides this approach, demographic analysis can involve comparisons of population outcomes or distributions between different census data sets. This might entail the comparison of population outcomes from previous censuses (Keane et al., 1985).

2.3 Controversies in South African censuses

South African censuses have been consistently characterised by high undercount. In the last three censuses the undercount has been at 10.7%, 17%, and 14.6% in the 1996, 2001, and 2011 censuses respectively (StatsSA 2011 Census). Such high undercount estimates triggered controversies around these censuses as researchers, organisations and the public contested the processes and outcomes of these censuses. They questioned the accuracy of the PES as method for undercount testing and adjusting in the country's censuses (Dorrington, 1999; Dorrington and Moultrie, 2012). For this reason both the undercount estimates obtained using this method and the adjusted counts became equally controversial. The main problem with undercount is that although remedies are improvised to correct it, none have proven to be completely satisfactory (Anderson, 2004). This section reviews some of the contestations that have been recorded in these censuses.

Some studies have noted that adjusted counts from these censuses underestimated children of 0 to 4 years as well as the white population while exaggerating counts for adult females (Dorrington, 1999; Dorrington, 2002). Substantiations from these studies were that Census 1996 underestimated about 800 000 white people, 1, 1 million men, and overestimated about 687 000 women aged 1five-years and above, and Census 2001 is believed to have underestimated about 400 000 white people. The researchers used population estimates based on population projections dating back from the 1970 Census and concluded that there were apparent deficiencies in the PES adjusted counts in these censuses. Other studies raised the hidden migration theory which suggested that between 400 000 and 600 000 white people might have left the country, implying that his population was not underestimated (Centre for Development Enterprises, 1998).

Other studies expressed concern about the high undercount estimates recorded in these censuses (Gernertzky, 2012; Schultz, 2013). In particular they have questioned the contrast between the massive resource input and the high undercount estimates recorded. South African News Agency (2012) noted that although the undercount dropped from 17% to 14.5% in the 2001 and 2011 censuses respectively, it is still too high. The Statistician-General also admitted that he had hoped for less than 2% undercount in the 2011 Census, and admits that 14.5% undercount is a concern (Schultz, 2013). Others have questioned the credibility of the 2011 Census results after noting that the undercount translated to 1 in every 7 people having been missed in the census, which slightly varied from the 2001 undercount where 1 in every 6 people were missed (Berkowitz, 2012). With regard to Census 2011, other research suggested that the results were hurriedly published and had inaccuracies (Moultrie and Dorrington, 2012). For instance, the increase in fertility trend suggested by Census 2011 was seen as inconsistent with the country's fertility decline observed over many decades.

Some studies also questioned the suggested increase of young white women which they felt had no trace from previous censuses (Centre for Actuarial Research, 2012). A member of the public also tweeted about these issues, referring to 'The odd baby boom and the strange influx of young white women' (sarahemilyduff). However, Professor Eric Ujo insisted that the 2011 Census counts were a better estimation of reality than models now proven to be incorrect (De Wet, 2012). These views were complemented by Professor Michael Garenne who noted that the fertility increase suggested by Census 2011 was consistent with findings from recent fertility estimates of the country.

Other studies also suggested that provincial population distributions from midyear population estimates and those suggested in Census 2011 were inconsistent. For instance, they indicated that in Census 2011, the Western Cape contributed 11.5% of the total population and Gauteng 23.5%. This was contrary to midyear population estimates of 10.5% and 22.4% for the respective provinces (Berkowitz, 2012).

The public also had further remarks on these censuses on social media platforms:

'… My family was not counted and I thought it wouldn't make much difference ...' (Zwelakhe Sithole); 'my family too was not counted' (Boingotlo Molale); 'this is not a surprise to us … in our region at least 35% did not fill in the census. Most of the rural areas were left uncounted, and as we are a holiday region, 80% of homes and seaside cottages are vacant. On top of that, almost 600 homes belong to people who only come here for 6 months of the year from abroad' (Beverly Young); 'Counting in the previous census was very sloppy with many people living behind security gates not being reached ...' (Niki V); 'Census: A number of questions.' (Simaxis), 'In the Addo area very few members of the white community were enumerated or as in my case the forms were never collected …' (Cynic1964); 'I don't believe this census and the data manipulation or lack of it. South Africa is one of the countries where it is difficult to do a census or even do a survey. The other thing is did they count illegal migrants children as South Africans? To me this a political contrived census….' (EvansS Mazi).

The media has also added its voice to these controversies. *City Press* (2012) claimed that two top officials with Statistics South Africa (StatsSA) were dismissed for refusing to lower the 18% undercount estimates they reportedly obtained for Census 2011. The source further claims that the Statistician-General had expected a 2% undercount, but the officials stood firm with their findings, leading to their dismissal. However, the Statistician-General responded by stating that the two presented wrong results after committing methodological and computational errors (City Press, 2012).

Others have questioned the Census 2011 figures for the elderly population, which are estimated at 15 000 (Mybroadband, 2013). Among them, 350 are aged between 115 and 120 years. In the USA, people in the same age range were estimated at 50 000 in 2012. This implies South Africa has a higher proportion of 0.025% of the population aged 11five-years and above compared to the USA's 0.015%. AnimateX (2012) tweeted the following: 'This is extraordinary and Guinness Book of Records should be excited to hear this.' However, Statistics Council member Professor Jacky Galpin of the University of Witwatersrand also insisted that these census counts were consistent with findings from the Post-enumeration Survey.

Politicians have also been drawn into the controversies as 'red flags' were raised over some of the outcomes of Census 2011, and questions were asked in parliament (TNA Reporter, 2012). The reporter further noted that Tim Harris, the Democratic Alliance's shadow minister of Finance, wanted the parliament to review the census results with regard to their credibility. They were suspicious of the 14.6% undercount estimate reported by StatsSA as some claimed that the undercount was even higher than 17% in the 2001 Census.

2.4 Undercount patterns

The undercount patterns observed from both the censuses around the globe and South Africa's last three censuses are described in this section. The first subsection describes the former and the last subsection describes the latter.

2.4.1 Censuses around the globe

The PES has often been used by census authorities for estimating the and countries like Australia, Canada, China, Ghana, New Zealand, United Kingdom, USA, South Africa, Zambia, and Zimbabwe continue to use this method. Other countries have used the Demographic Analysis. This technique estimates the undercount at macro level, and it mainly involves the comparison of census counts against counts from vital records (Keane et al., 1985). The Bayesian model was introduced at the beginning of this millennium in the United Kingdom.

Despite the different methods used to estimate the undercount, there are indications of systematic patterns of undercount that have been observed. Awareness of the extent and patterns of undercount is important for stakeholders, particularly policy makers and census authorities (Jonson, 2012), as obtaining such information may assist them to devise strategies that reduce undercount. This information is important to policy makers as they are equipped with information that guides them in making informed decisions.

Research has identified that children aged five years and below are consistently missed in censuses (Anderson, 2004). As further noted by Anderson (2004), the undercount of children aged four years and below is common in Australian and Chinese censuses. This pattern was pronounced in censuses 1995 and 2000 in China and in Australia in censuses 1976, 1996 and 2001 (Anderson, 2004). US research has also confirmed that children under five have constantly been missed in the country's censuses (Steffey, 1997). In South Africa's last three censuses there have also been reports of high undercount of children under five (Dorrington, 2002). Even the counts that have been adjusted for undercount are believed to have underestimated children in this age group (Dorrington, 1999; Dorrington, 2002)

Young adults, especially males, are also highly undercounted in censuses (Anderson, 2004). Research in China observed an increase in undercounted individuals among young males in censuses 1990 and 1995 (Anderson, 2004). In South Africa, particularly in censuses 1996 and 2001, the undercount for young male adults has been very high (Dorrington, 1999 & 2002). The majority of young people are difficult to capture in censuses mainly because they are highly mobile and often change their location, a behaviour that increases their chances of being missed in censuses. It has been indicated that there is a correlation between undercount of children below five years and young adults. As noted by the United States Bureau of Census (1953:2) as quoted by Anderson (2004), young adults were among the largest undercounted group in the 1950 Census in the US and it has been observed that among the 82% of missed infants the entire young family had also been also missed. This is probably because mobility behaviours of young adults often include their young children, hence increasing the chances of both subpopulation groups being concurrently missed.

Another hard-to-find sub–population group noted in various censuses is minorities. Though results indicate that census taking in America has improved since 1880, the differential in undercount between black minorities and whites majorities has persisted (Steffey, 1997).

Minority groups that also included Hispanics were undercounted in the1990 census in the US (Steffey, 1997). In South Africa's last three censuses, the outcomes have not been consistent with regard to the undercount of minority groups. In Census 1996, black people, who are the majority, had a higher undercount than white people (StatsSA, 2011). However, the last two censuses recorded the highest undercount among white people relative to black people (StatsSA, 2011). Possible explanations as to why minorities are often undercounted may not be that apparent. However, it is possible that in most instances majority groups have a greater sense of belonging to a system than minorities and therefore they are more likely to participate enthusiastically in national events like census enumeration than minority groups.

Another hard-to-find group is migrants, and these can either be of domestic or international origin (Liang and Ma, 2004). Researchers in China have identified an increase in a group of domestic migrants which they have termed the 'floating population' (Liang and Ma, 2004), and according to these researchers, the term refers to the ever-increasing number of domestic migrants without local household registration status. Such a characteristics associated with the 'floating population' is likely to lead to a high undercount of this subpopulation group (Liang and Ma, 2004). This is largely because they don't have official residence where enumerators can locate them. International migrants such as those of Mexican origin have often been undercounted in US censuses. In South Africa, it is also possible that immigrants, particularly those who are undocumented, are highly undercounted. High mobility and fear of victimisation among migrants creates problems for census takers wishing to locate them for enumeration (Liang and Ma, 2004).

2.4.2 South Africa's last three censuses

Undercount estimates from South African censuses are summarised in tables and graphs below. Census 2011 indicates that minority groups i.e. white people, coloured people, and Indians respectively had higher undercount estimates relative to black people, who are the majority. Males also had an overall higher undercount rate relative to that of females. According to age groups, children under five had the highest undercount estimates. Young adults aged 20 to 29 had the highest undercount estimates, followed by adults aged 30 to 44. The elderly aged 6five-years and above had the lowest undercount estimates [Table 2.1].

Table 2.1: Undercount estimates by race, sex, and age group for census 2011

Category	Net undercount rate	Standard error (+ or -)	95% confidence interval limits Lower	Upper
All persons	14,6	0,32	14,34	14,86
Population group				
Black African	9,9	0,138	9,58	10,12
Coloured	12,8	0,604	11,66	14,03
Indian or Asian	11,5	0,528	10,50	12,57
White	15,6	0,367	14,86	16,30
Other	23,2	1,165	20,88	25,45
Sex				
Male	15,9	0,135	15,63	16,16
Female	13,4	0,133	13,11	13,63
Age group				
Under 5 years	15,1	0,156	14,81	15,42
5-9 years	11,4	0,173	11,10	11,78
10-14 years	11,1	0,150	10,79	11,37
15-19 years	12,8	0,146	12,47	13,04
20-29 years	18,1	0,147	17,85	18,43
30-44 years	16,9	0,139	16,60	17,14
45-64 years	12,5	0,165	12,16	12,80
65+ years	9,8	0,193	9,46	10,22

*Subject to rounding error
*The undercount rate for undetermined population group is excluded from this table

Source: StatsSA, 2011

Undercount for race by sex indicated that males were missed more than femlaes. This was confirmed from comparisons for all races. The undercount rate for black people by age groups indicated that children under five years and adults aged 20 to 29 and 30 to 44 were mostly missed relative to other age groups. The elderly had the lowest undercount estimates. The same patterns were observed among colourde people and Indians/Asians. White age groups between 10 and 44 years had higher undercount estimates compared to children aged below five years. The elderly black population group had the lowest undercount rate compared to the elderly population groups of other races. [Table 2.2]

Table 2.2 Undercount estimates for races by age group and sex for census 2011

		Net undercount rate	Standard error (+ or -)	95% confidence interval limits Lower	Upper
Population group by sex					
Black African	Male	10,5	0,142	10,21	10,76
	Female	8,6	0,140	8,34	8,88
Coloured	Male	13,0	0,609	11,77	14,16
	Female	11,9	0,602	10,68	13,04
Indian or Asian	Male	11,9	0,591	10,77	13,08
	Female	10,3	0,500	9,34	11,30
White	Male	15,6	0,370	14,84	16,29
	Female	14,5	0,368	13,79	15,23
Other	Male	24,0	1,002	22,07	26,00
	Female	19,8	1,562	16,76	22,88
Population group by age group					
Black African	Under 5 years	11,5	0,150	11,25	11,83
	5-9 years	8,2	0,191	7,83	8,58
	10-14 years	7,7	0,162	7,36	8,00
	15-19 years	9,4	0,157	9,07	9,69
	20-29 years	14,1	0,156	13,77	14,38
	30-44 years	12,8	0,147	12,46	13,04
	45-64 years	8,5	0,162	8,16	8,80
	65 or more	5,6	0,179	5,26	5,96
Coloured	Under 5 years	14,8	0,615	13,63	16,04
	5-9 years	14,1	0,633	12,89	15,38
	10-14 years	14,5	0,629	13,22	15,69
	15-19 years	13,9	0,648	12,63	15,17
	20-29 years	14,4	0,632	13,17	15,65
	30-44 years	14,6	0,622	13,39	15,83
	45-64 years	12,4	0,652	11,08	13,64
	65 or more	11,2	0,741	9,75	12,66
Indian or Asian	Under 5 years	14,4	0,723	12,98	15,82
	5-9 years	12,7	0,737	11,23	14,12
	10-14 years	11,5	0,759	10,06	13,04
	15-19 years	8,8	0,633	7,59	10,07
	20-29 years	16,4	0,553	15,33	17,50
	30-44 years	14,3	0,675	13,01	15,66
	45-64 years	8,8	0,597	7,63	9,97
	65 or more	10,0	0,594	8,88	11,20
White	Under 5 years	16,8	0,547	15,69	17,83
	5-9 years	16,4	0,487	15,41	17,32
	10-14 years	18,1	0,452	17,19	18,97
	15-19 years	18,3	0,396	17,52	19,07
	20-29 years	20,0	0,499	18,99	20,94
	30-44 years	18,1	0,398	17,37	18,93
	45-64 years	15,8	0,403	15,04	16,62
	65 or more	13,3	0,527	12,26	14,33

Source: StatsSA, 2011

The white population group had the highest undercount in Census 2001, followed by the black population. The lowest undercount was among the coloured population. With regard to to sex, the overall undercount estimate of females was less relative to that of males. Among the age groups, the highest undercount was of young adults aged 20 to 29 years, followed by those aged 30 to 44 years. Children aged below five years are often the most undercounted and they had the third highest undercount, together with adults aged 45 to 64 years. As in Census 2011, the elderly again had the lowest undercount. [Table 2.3]

Table 2.3 Undercount estimates for race, sex and age groups for census 2001

	Net undercount rate	Absolute error (+ or -)	95% Confidence interval limits*	
			Lower	Upper
All persons	17,6	1,1	16,6	18,7
Population group				
Black African	17,4	1,2	16,2	18,6
Coloured	15,3	1,4	13,9	16,7
Indian or Asian	16,7	2,1	14,6	18,9
White	23,3	2,6	20,7	25,9
Sex				
Male	18,6	1,0	17,6	19,7
Female	16,9	1,1	15,8	18,0
Age group				
Under 5 years	16,8	1,2	15,6	18,1
5-9 years	16,2	1,4	14,8	17,6
10-14 years	16,0	1,5	14,5	17,5
15-19 years	16,6	1,1	15,4	17,7
20-29 years	20,6	1,0	19,7	21,6
30-44 years	19,6	1,0	18,6	20,6
45-64 years	16,8	1,0	15,8	17,8
65+ years	14,6	1,2	13,5	15,8

Source: StatsSA, 2001

Undercount by age groups in races indicates that among the coloured population the highest estimates were of young adults aged 20 to 29 years, followed by adults aged 30 to 44 years. There were no estimates for children aged below five years. Children aged five to nine years had the fourth highest undercount after those aged 15 to 19 years. Indians/Asians were the most undercounted group in adults aged 30 to 44 years, followed by those aged 20 to 29 years and 45 to 64 years respectively. Children below five years had the fourth worst undercount. The last undercount was for children aged five to nine years followed by those aged 10 to14 years.

The lowest undercount estimates comapred to other age groups was in white children under five years and elderly people aged 6five-years and above. The middle-aged groups, as was the case with the other two races, had the highest undercount rates in those aged 20 to 29 years. There was no information on the black population. Undercounts estimates for males by age groups were consistently compared to respective estimates of their female counterparts. [Table 2.4]

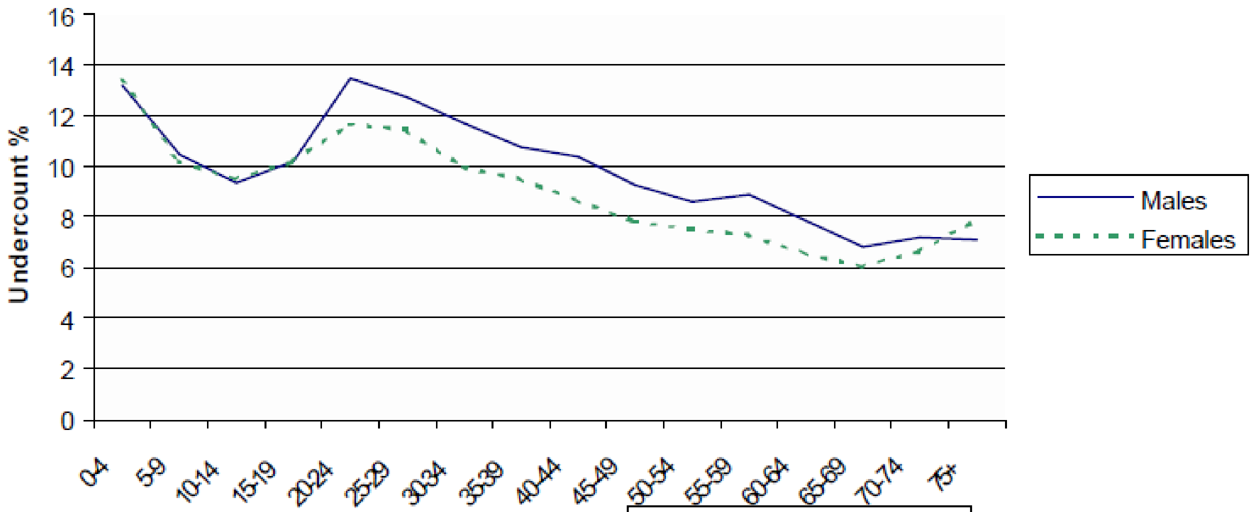
Table 2.4 Undercount estimates for race and sex by age group for census 2001

		Net undercount rate	Absolute error (+ or -)	95% Confidence interval limits*	
				Lower	Upper
Coloured	5-9 years	15,2	1,4	13,8	16,7
Coloured	10-14 years	14,9	1,6	13,3	16,4
Coloured	15-19 years	15,3	1,3	13,9	16,6
Coloured	20-29 years	16,9	1,0	15,8	17,9
Coloured	30-44 years	16,0	1,4	14,6	17,4
Coloured	45-64 years	13,8	1,9	12,0	15,7
Coloured	65+ years	13,3	3,2	10,2	16,5
Indian/Asian	Under 5 years	15,5	3,5	12,0	18,9
Indian/Asian	5-9 years	14,6	3,2	11,5	17,8
Indian/Asian	10-14 years	14,7	1,4	13,3	16,2
Indian/Asian	15-19 years	15,4	2,1	13,3	17,5
Indian/Asian	20-29 years	17,8	1,6	16,2	19,5
Indian/Asian	30-44 years	18,9	2,2	16,7	21,1
Indian/Asian	45-64 years	16,8	2,7	14,2	19,5
Indian/Asian	65+ years	15,2	4,8	10,4	20,0
White	Under 5 years	21,3	3,6	17,7	24,9
White	5-9 years	25,1	5,0	20,1	30,1
White	10-14 years	25,0	4,1	20,9	29,1
White	15-19 years	24,8	3,3	21,5	28,1
White	20-29 years	24,1	1,7	22,4	25,8
White	30-44 years	26,9	2,5	24,4	29,3
White	45-64 years	21,5	2,9	18,7	24,4
White	65+ years	18,1	2,7	15,4	20,8
Sex by age					
Male	Under 5 years	17,0	1,3	15,7	18,3
Male	5-9 years	16,4	1,5	14,9	17,8
Male	10-14 years	16,2	1,5	14,7	17,7
Male	15-19 years	16,7	1,1	15,6	17,8
Male	20-29 years	22,3	1,0	21,3	23,3
Male	30-44 years	21,5	1,0	20,6	22,5
Male	45-64 years	18,2	0,9	17,2	19,1
Male	65+ years	15,9	1,1	14,8	17,0
Female	Under 5 years	16,9	1,1	15,7	18,0
Female	5-9 years	16,2	1,4	14,8	17,6
Female	10-14 years	16,0	1,4	14,5	17,4
Female	15-19 years	16,6	1,2	15,4	17,9
Female	20-29 years	19,3	1,0	18,3	20,3
Female	30-44 years	18,1	1,1	17,0	19,2
Female	45-64 years	15,7	1,1	14,6	16,8
Female	65+ years	14,0	1,3	12,7	15,3

Source: StatsSA, 2001

As for Census 1996, the overall undercount estimates for age groups by sex indicated that children below the age of five years had the worst undercount estimates together with young male adults aged 20 to 24 years. There was a general decline in undercount estimates as age increased, for both sexes after the latter age group. The males' undercount estimates remained consistently higher than that of females between 25 and 74 years. From age 7five-years and above it was indicated that females had higher undercount estimates relative to males. [Fig 2.1]

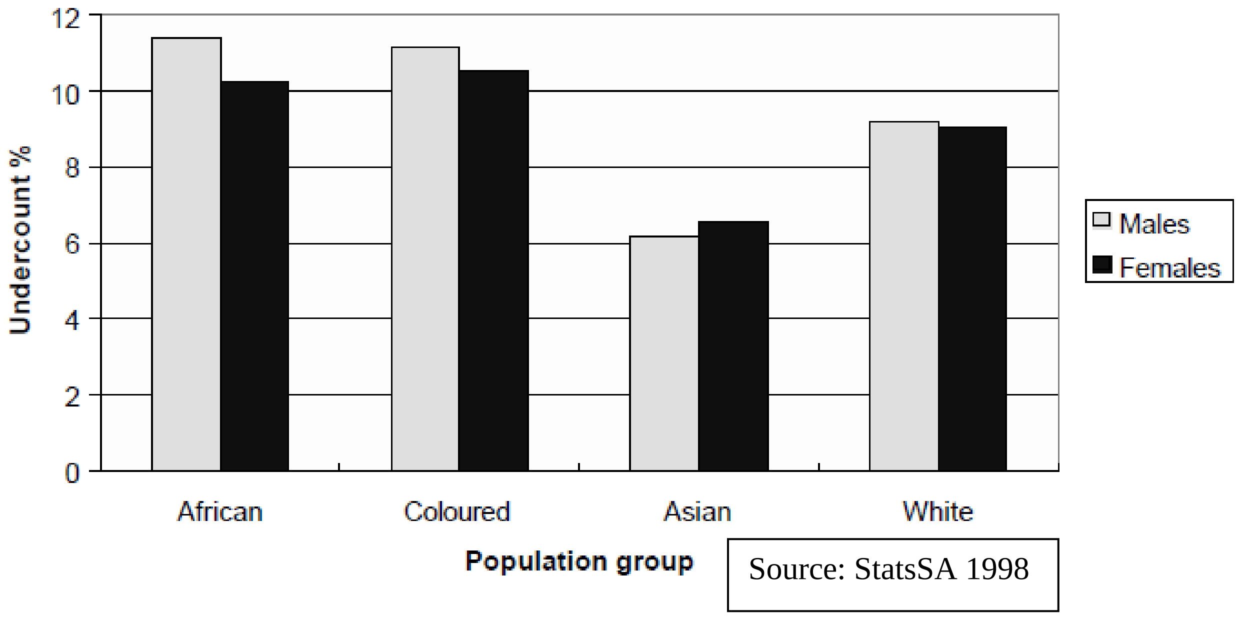
Fig 2.1 Undercount estimates by age group and sex for census 1996



Source: StatsSA, 1998

By race, Census 1996 indicated the highest undercount among the black and coloured populations. Males had a higher undercount relative to females in both races. The lowest undercount was in the Indian race, and females had a higher undercount relative to males. in the white population group, just as in the black and coloured population groups, males were missed more than females. [Fig 2.2]

Fig 2.2 Undercount estimates by race for census 1996



2.5 Estimation of content error

When evaluating coverage errors in censuses, the control of the levels of content errors in the respective censuses is important. Various demographic methods that were used to investigate coverage in this study rarely separate the effect of the latter errors from content errors. In fact it is often difficult to determine the extent of the one error over the other (Ewbank, 1981). As stated by Keane et al. (1985), both share a common characteristic of distorting the population distribution by census characteristics. Yet it is these distortions that are interpreted as being indicative of coverage errors in a census. Therefore a better estimation of coverage errors using demographic methods needs to be complemented by first estimating the levels of content in the data. This way one can account for the extent to which content errors are distorted, e.g. expected population distributions relative to the effect of coverage error.

Research has identified various sources of content error in censuses which include; respondents, enumerators, questionnaire, and processing errors (Keane et al., 1985, Statistics Canada, 2006). Responses from captured individuals in censuses are often not entirely accurate (Statistics Canada, 2006). Misreporting, though more common when reporting for age (Dahiru and Dikko, 2013); can also be common when reporting on other characteristics like sex, race and marital status. Misreporting can be a result of various factors. At times respondents may suffer from recall bias, for example because of a time lapse. Yet at times certain individuals may decide not to provide the correct information. This often occurs when respondents feel insecure or unclear about the information-gathering exercise being conducted.

Enumerators are also known to be another key source of content error in censuses. In other instances, enumerators may simply fail to articulate a question in case of respondents asking for assistance. For example, there are cases when respondents may not be clear about what is wanted, and enumerators may interpret the questions inaccurately. There are other instances where enumerators can influence respondents to give certain answers or even make up information for reluctant respondents (Keane et al., 1985). In case of the former, this can happen when a respondent requests question clarification and the enumerators has his/her own preconceived answer, e.g. based on political or racial motives. Whereas in the latter scenario, an enumerator may find environments at certain homesteads unwelcoming. For example, some enumerators may find it difficult to access certain homesteads because of vicious dogs; in other cases simply because household owners were not welcoming them or the navigation of certain areas were difficult. Enumerators faced with such situations may end up cooking information to just please their superiors.

Questionnaires are also known to introduce errors in censuses by the nature of the questions asked. A case example is the migration question P-12 for South Africa's Census 2001. The question reads: Five years ago (at the time of Census '96) was (the person) living in this place (i.e. this suburb, ward, village, farm, informal settlement)? The prescribed answers were: Yes, No, and Born after October 1996. The next question, i.e. P-12a, will exclude all other respondents except those who answered ‘No’, when probing to establish the migration status of respondents. Consequently this leads to inaccurate recording of migrants as the two questions may result in missing out return migrants. This also leads to distortions in the number of migrations occurring during the census period. Researchers have noted that in actual fact some questions may be confusing to both respondents and enumerators (Keane et al., 1985), and this might be one typical example.

Errors are also introduced in the processing stage. Census taking involves the collection of huge volumes of data, and errors emanating from coding and computations cannot be ruled out. Large volumes of gathered data often results in the personnel synthesizing it, and they are overwhelmed by work and also forced to work overtime in order to meet timelines (Kean et al., 1985). This further increases the likelihood of human error during the processing stage.

Content errors can either be randomly or systematically ushered into a census. In the former case, error occurs systematically in one direction. This often results in the reported statistics being far away from the true statistics, i.e. by being either lower or higher (Canadian Statistics, 2006). On the other hand, random error occurs in either direction, consequently offsetting each other and this may result in the reported statistics being close to the true statistics.

However, it is clear that content errors can lead to distortion of population distributions that may have been expected from data that is free from coverage errors. Therefore, when investigating coverage errors, e.g. of undercount using demographic techniques which often test for any distortions from expected trends, it is important to be clear about the extent to which content error could be contributing towards the observed distortions.

2.5.1 Accuracy of reporting on age and sex

A common measurement that has been developed to estimate levels of irregularities in reporting for both age and sex within collected data sets is the UN joint age-sex index. The index is expected to be less than 20 when error in reporting for age and sex is minimal and 20 to 40 when the data is inaccurate. If the index is above 40 then age and sex reporting in the respective data is estimated to be highly inaccurate. The index quantifies the overall accuracy of age-sex data arranged in five-year age groups (Chuks, 1996). The index is dependent on the summation of age ratio scores for both sexes to that of sex ratio score which has been multiplied by three.

2.5.2 Digit preference or avoidance in age reporting

Research has observed systematic digit preferences and avoidance in age reporting among various data sources (Swanson and Siegel, 2004). Methods that investigate the levels of content error often rely on the testing accuracy of age reporting. This is probably because age is an important variable that largely affect all other individuals' characteristics. One of these tests is the Whipple's Index, which measures preferences in age reporting for ages that end with digits 0 and 5. One of the limitations of this index is that it is based on an assumption which may not be often met. The measurement is based on the assumption of linearity in five-year age range (Swanson and Siegel, 2004). The measurement of preference of ages ending with other terminal digits other than 0 and 5 is therefore not feasible when using this method. A Whipple index of 500 would reflect age heaping at only one terminal digit of either 0 or 5, whereas less than 105 suggests accurate data with no preference digits ending with 0 or 5.

Other indices that measure age preference and/or avoidance that have been developed partly as an improvement of the Whipple Index are the Meyers and Bachi indices. The two indices are also compromised by the fact that they are based on the same assumption of linearity. However, the difference is that both the Meyers and Bachi indices measure either digit preference or avoidance for all ages ending with digits 0 to 9. The two are different from each other in that the Meyer's index is based on the sum of absolute deviations of values, whether positive or negative, whereas the Bachi Index is based on positive deviations only (Chuks, 1996).

2.6 Adjusted census counts

South Africa's censuses that were conducted since 1996 have been evaluated for coverage error using the Post-enumeration Survey (PES). This is a survey carried out after a census on the same population (Keane et al., 1985). Ideally the PES has to be carried out soon after completion of the census. This is primarily done to ensure reduced chances of change in the population's characteristics between census date and PES date. When responses from the PES perfectly match those of the census, this suggests that the census was accurate. When the matching is below 100%, this suggests a census undercount, and when the matching is above 100% this suggests over counting. PES can make an effective evaluation of a census if it is conducted under the same procedures and conditions of the census (Keane et al., 1985). These include: response-obtaining methods, response-recording methods, and questionnaires used.

However, matching outcomes rarely yield 100% resolved cases. Therefore the final undercount estimates in South African censuses have not solely been based on matched results. For example, in the 1996 census 78% of cases were successfully matched, and 22% cases were unresolved (StatsSA, 1998). The decision on whether unresolved cases should be treated as counted or not cannot be made merely based on a respondent's answer as to whether or not he/she had been counted; this is due to the possibility of errors like recall bias.

Each person sampled for the PES had to be allocated a score of probability of having been counted that lies between 0 and 1. The latter would mean the person was really matched, yet the former would mean the person was clearly missed (StatsSA, 1998). Computation of probability of having been enumerated among unresolved cases was done using a multivariate technique known as Chi-square Interaction Detection (CHAID). Estimation of probability to be attached to an unresolved case was imputed using characteristics of resolved cases (StatsSA, 1998) to create binary-dependant variables that categorise individuals as either having been counted or not. Explanatory variables predicting the probability of having been counted as opposed to having been missed included: whether or not the respondent identifies the person as having been counted, type of enumeration area, race, gender, age group, and household (StatsSA, 1998).

The PES's major strength in census evaluation is that it is a direct method, and these often have less error than indirect methods. In most cases the indirect methods are used when direct methods are not feasible. However, the PES has its own limitations. For instance, there is an assumption that the population characteristics remain the same during the time of census and the time when the PES is conducted. Populations are known to be highly dynamic, especially in countries like South Africa where there is a lot of migration both internally and internationally. Other researchers have suggested that a PES is not likely to produce different findings from a census among the hard-to-find people as someone who was hard to find in a census is likely to remain hard to find in a PES as well (Freedman and Navidi, 1986). Furthermore, as noted by the same researchers, the assumed independence of the PES from the census is an assumption that has remained open to questioning.

2.7 Estimating of coverage error

Accuracy of census counts at macro level can be analysed through demographic techniques; for instance census counts at national, provincial or even district level can be compared with those from non-census data (Redfern, 2001). This often involves comparisons of distributions of counts by age, sex, and race from compared data sources, to see if they match. Non-census data whose counts can be compared to census counts include vital records. However, in the case of South Africa such data may not necessarily be independent from census data, since the custodian of both types of data is the same, i.e. StatsSA. Though the compilation of vital records in South Africa is carried out in collaboration with other government departments, it is StatsSA that finally has the mandate to carry out data management and publication. Yet in the case of censuses, StatsSA has a sole mandate of managing the entire process from start to end.

Independent data that StatsSA has no official involvement in the collection processes includes the Health and Sociodemographic Surveillance Site data. South Africa has three surveillance sites: Agincourt in Mpumalanga, Dikgale in Limpopo, and Africa Centre in KwaZulu-Natal. These surveillance sites collect sociodemographic information on all individuals within the entire area covered. Regular updates of the data in these sites are expected to guarantee quality of data especially with regards to population size. Such sources of data present an opportunity to effect direct tests for accuracy of census counts, through comparison. Census counts can also be compared with respective counts from population projections. For instance the United Nations' population division devised the Mortpark model for projecting populations. This model takes into consideration the variation in population dynamics among the world's sub regions. However, like other population projection models, Mortpark also has limitations as it is based on mortality, fertility and migration assumptions.

Besides direct comparison of census data with non-census data, the population outcomes from the former can also be compared to theoretical distributions. The underlining fact will be that due to biological and cultural factors population distributions follow well-known and fairly predictable patterns (Keane, 1985). For example, population distribution of males compared to that of females at younger ages can be expected to show a high sex ratio, and this should decrease with increasing age. Any departures from such theoretical population distribution may be interpreted to signify the possibility of error, due to e.g. undercount. Besides the above approaches, analysis can also involve the comparison of population outcomes from successive censuses. This might involve checking for consistence in population distributions between the previous census and later censuses (Keane, 1985).

2.7.1 Growth rate analysis

The analysis of growth rates is a technique developed by demographers for partly examining the accuracy of census counts. Findings that have been observed from prior researches have indicated that in the absence of coverage errors, the cut-off point for growth rates in any given population is between 0 and 3.5 (Moultrie, 2004). The strength of this technique in investigating the accuracy of census counts is that it does not suffer from the confounding effect of content errors as the findings of such a growth rate analysis technique is determined only by counts in the investigated censuses. The technique's main limitation is that it cannot separate instances of genuine growth rate that are beyond stipulated limits, e.g. because of extreme emigration or immigration, from instances when the growth rates that are beyond the stipulated range are due to coverage error.

2.7.2 Graphical cohort analysis

The technique is applicable when examining two or more population distributions (Keane et al., 1985). The underlying principle is that due to mortality effect, each age cohort from the previous census should experience reduced counts in future censuses. Assuming that other distorting factors are ruled out, the lines describing population distribution by age for each census should follow the same trend, and should not cross each other. Furthermore, the line describing population distribution for the latest census should lie at the bottom, whereas the line describing the population distribution of the earliest census should be at the top. Departures from these characteristics suggest coverage errors, and in most cases undercount. However, deviations from the expected distributions may be due to other distortive factors like high volumes of migration (Keane et al., 1985). For instance, high immigrations may counter the mortality effect on the population size for respective age cohort in the succeeding census.

2.7.3 Age ratios analysis

The analysis of age ratios is based on the fact that the size of a particular enumerated cohort should approximate the average size of the immediately preceding and subsequent cohorts (Keane et al., 1985). The feasibility of this is based on the assumptions of constant mortality, fertility, and minimal distorting effect of migration. If this holds, then the age ratios for particular age cohorts should estimate 1, or 100 if multiplied by a constant of 100 (Keane et al., 1985). When age ratios for specified age groups indicate significant deviations from 1, this is interpreted as indicating coverage error.

However, limitations of age ratios in investigating accuracy of census counts include the distortive effect of other factors, like sharp changes in fertility and mortality trends, and high volumes of migrations (Keane et al., 1985; Moultrie, 2004). A country like South Africa is believed to receive high volumes of migrants particularly from Sub-Saharan Africa. Furthermore, the mortality effect of major killer diseases like AIDS should have been high in South Africa. However, increase in access of anti-retro viral drugs by majority of affected people in the country by 2008, should have led to sharp mortality declines. According to Keane et al. (1985) sharp swings in either mortality or fertility will distort age ratios.

2.7.4 Sex ratios analysis

Sex ratios are expected to follow known patterns, e.g. sex ratio at birth often should suggest more males than females. A low sex ratio at birth could therefore suggest undercount of males at birth. However, though high sex ratios like 105 are normal in developing countries, some researchers have found out that in Sub-Saharan Africa the sex ratio tends to be closer to 100 (Garenne, 2004). The research investigated countries from Eastern and Southern Africa, and demystifies the general perception that sex ratio at birth is always high. Across the rest of the age groups, since mortality for males is higher than that for females, the sex ratios should reflect this mortality differential (Moultrie, 2013). This implies a general declining trend in sex ratio as age increases. The overall expected sex ratio should be slightly less than 100 (Moultrie, 2013).

There are certain circumstances that are expected to inflate sex ratio across age groups, and consequently lead to an overall high sex ratio. These include high prevalence of female infanticide, high maternal mortality, and deliberate negligence of women (Moultrie, 2013). In societies like those from Asia, e.g. China, sex ratios often remain high partly because girls are often neglected to die, or they are hidden away from being captured in data collection processes. Such societies value boys at the expense of girls. On the other hand, selective migration may also act as a distorting factor with regard to sex ratios (Keane et al., 1985). Equally, more male immigrants than females may lead to an overall high sex ratio.

2.8 Matching of counts

One of the direct methods of testing for the accuracy of census counts is by comparing them to respective counts from other data sources that are independent from censuses (Keane et al..

1985). The approach is vital for this study because the findings obtained should provide answers to issues that have been contested in these censuses. For instance, if adjusted census counts match respective counts from other data sources, there would be suggestions that South Africa's last three censuses accurately estimated the actual population counts for the country. This would also imply that undercount estimates derived from the PES were generally accurate as the adjusted counts were mathematically derived using the undercount estimates. The accuracy of both adjusted counts and the undercount estimates would also suggest that the PES was a reliable method for estimating the undercount in the country's last three censuses.

2.9 The effect of undercount on demographic parameters and service delivery

Undercount can affect fertility, mortality and migration estimates in many ways. Fertility estimates are computed from live births recorded in a year in relation to the population from which the children are born. When there is equal undercount among children born and the population from which they are born, the two offset each other and the effect of undercount can be ignored (Rindfuss, 1974). In regular instances undercount has often varied between different subpopulation groups. For instance, South Africa's 1996 Census had an over count of women and an undercount of children aged 0 to 4 years (Dorrington, 1999). Such a case would therefore imply an inaccurate downward effect of fertility estimates.

In computing mortality rates the numerator is the number of deaths occurring in each census year, and the denominator is the total population estimate from a census (Johnson, 2012). Demographers and planners are more interested in mortality estimates, particularly mortality rate ratios for subpopulation groups (Johnson, 2012), e.g. white versus black people. The resulting mortality rate ratios suggest the difference in susceptibility to mortality between subpopulation groups. This information is important in strategizing for mortality reduction.

Mortality reduction is vital in achieving a population's better welfare as mortality levels are believed to be significant indicators of the health status of a population (Centre for Disease Control and Prevention, 1999). Low levels of undercount, e.g. among white people relative to higher levels of undercount among black people, may lead to a decrease in the mortality rate ratio of blacks to whites (Johnson, 2012). As for migration, similar to fertility and mortality, equal missing of moving migrants and out-moving migrants will lead to no effect on the net migration rate. However, when undercount varies between the two groups of migrants, net migration may either be inflated or deflated depending on which group was undercounted more than the other.

National funds received, e.g. by states and counties in the case of the US and by provinces in the case of South Africa, is largely based on the number of people living in these respective areas as estimated in the census. As in the case of California in its 2000 Census, if coverage measurement and adjustment were to omit between 1 and 1.2 million people, this would result in an estimated revenue loss between $1.1 and 1.5 billion in the ensuing decade (Steffey, 1997). Murray (1992) as quoted by Steffey (1997) noted that depending on a programme, 32% to 42% of areas would gain funds because of using corrected population counts in 1990. Therefore in the case of California, undercount would have negatively affected service delivery.

This would have implied poor and unfair service delivery for the state. In Wales, the council of Cardiff City fears that the 2001 Census may have produced an undercount of about 22 000 people, and this according to Cardiff council leader Rodney Berman, was most likely to have resulted in the city losing about £85 million by 2011 (BBC Politics Show on Census Day, 2011). Since undercount has the potential of undermining actual funds that should have been allocated to a particular subpopulation group, some local authorities have taken it upon themselves to motivate people within their councils to take census participation as an obligation (BBC Politics Show on Census Day, 2011). Berman, for instance, has been at the forefront in motivating residents of Cardiff City to participate in the 2011 Census and he even ensured that census forms were completed and returned (BBC Politics Show on Census Day, 2011). Besides the numbers obtained from a census being crucial in fund allocation, census also provides extra information about subpopulation groups in the country. That is important because the needs of different subpopulation groups vary.

Failure to efficiently provide service delivery can even result in violent protest from an expectant population. In South Africa, angry residents of Zandspruit informal settlement in Johannesburg blocked off busy roads and stoned buildings and vehicles as they protested against poor service delivery in their community (Think Africa Press, 2011). Even though the protesters believed the source of poor service delivery was due to corruption among municipal authorities, this incident serves as a reminder of what authorities can expect from the general population when service delivery is poor, for whatever reasons.

2.10 Deficiencies in existing literature

The existing gap in prior studies related to the discourse of census controversies in South Africa is the investigation of PES accuracy by comparing its undercount estimates with respective undercount estimates obtained from alternative methods. This is a direct procedure that investigates the accuracy of the PES in estimating and adjusting for the undercount in South Africa's last three censuses. Alternative methods that have been used to test for the undercount include the DA. This method has often been used in the United States of America (USA). The other method is the Bayesian model, which has been used in the United Kingdom (UK).

Other deficiencies in literature include investigations on which census counts between adjusted and unadjusted are better estimates of South Africa's population. Also lacking are studies that investigate the effect of undercount on both demographic parameters and service delivery. These gaps were addressed in this study.

CHAPTER 3: DATA SOURCES AND METHOS OF ANALYSIS

3.1 Introduction

This chapter describes the types of data that were used in this study: census and non-census data. The second section describes data processing which included descriptions of reconstruction of unadjusted counts at both national and at small areas level, construction of population estimates at national level using MORTPAK's PROJCT programme, and the GIS mapping of adjusted census counts at small areas level. The third section is on methods of analysis used in achieving the specific objectives. Last is a section that summarises software packages that were used in the study.

3.2 Data sources

3.2.1 Census

Census data was drawn from South Africa's 1996, 2001, and 2011censuses. The table below describes some of the features that were characteristic of the three censuses. [Table 3.1]

Table 3.1: Some of the features characteristic of the three censuses

Characteristic	Census 1996	2001	2011
Reference point	Night of 9th October	Night of 9th October	Night of 9th October
Information collected	Demographic, Social and Economic	Demographic, Social and Economic	Demographic, Social and Economic
Number of enumeration areas	86 000	Reduced to 80 782	Increased to 89 305
Harmonisation of questions to match other censuses in Southern Africa	No	Yes	Yes
Method used	De-facto	De-facto	De-facto
Translations	Questionnaires	Questionnaires	Questionnaires

	translated into all the eleven languages	translated into all the eleven languages	translated into all the eleven languages
Translations (continued)	All the 11 versions of translation were available in the field	Only the enumerator had the translated version	Only the enumerator had the translated version
Type of enumerators	Unemployed people	Unemployed people	Unemployed people
Map reading training	Enumerators not trained	Enumerators trained	Enumerators not trained
Manuals	Less detained	Very detailed	Very detailed
Pilot survey	Yes	Yes	Yes
Use of digital devices	Very little	Fairly digitalised	Highly digitalised
Processing centres	Nine processing centres	One processing centre	One processing centre
Default method of conducting interviews	Enumerators conducted face to face interviews Questionnaires were only left upon request and were returned via post	Enumerators conducted face to face interviews Questionnaires were only left upon request and were returned via post	Enumerators conducted face to face interviews Questionnaires were only left upon request and were returned via post
Conducted PES	Yes	Yes	Yes

Source Cronje and Budlender (2004); StatsSA (2011)

The data was collected from all individuals residing in the country on 9 October, in the year of each respective census. The information collected includes individuals' demographic and socioeconomic characteristics. The data is classified as either unadjusted or adjusted, and the former data carries the undercount error, and the latter corrects the error. The PES was used to determine the undercount rate, e.g. by, age, sex, race and province using about 600 enumeration areas that were randomly sampled countrywide immediately after the respective censuses. The undercount rate was used to determine the adjustment factor.

Adjusted counts are a product of the population enumerated multiplied by the respective adjustment factor.

The formula used to compute the adjustment factor is stated below.

$$\textit{Adjustment Factor} = \frac{1}{(1\text{-Undercount rate})}$$

Census data is available for public use from national to small areas level (SAs), and at times if granted permission by StatsSA the data can be obtained at enumeration areas level (EAs). The latter is the lowest level at which data is captured at group level, before the individual level. To use data at EAs level permission has to be granted by StatsSA. This is meant to ensure that confidentiality is maintained. Censuses 2001 and 2011 had data at both SAs and EAs level, yet Census 1996 did not have data at the latter level. Hence permission was sought and granted for the use of counts at EAs level in the case of Census 1996.

The method used in collecting census data in South Africa's last three censuses was the De-facto method. The method counts individuals as residents of the area from which they were captured. The default process of data collection in South African censuses is through enumerators visiting each dwelling unit on the census day, and handing over a questionnaire to the head of the household for completion. The form is completed in the presence of the enumerator and handed back upon completion. In exceptional cases the household head would be allowed upon request to keep the form and complete it later, before sending it back by post. This data is available for use from national to small areas level.

The country's adjusted counts are obtained as 10% samples from the StatsSA website, whereas the full census counts are obtained in SUPERCROSS format. This is software that was devised in Australia and is used for only cross-tabulation analyses. StatsSA adopted the use of this software since Census 1996. Census counts can be drawn from SUPERCROSS through GIS mapping. The main strength of census data is that it is free of sampling errors. However, its main limitations are that it is highly susceptible to non-sampling errors of both content and coverage. Adjustment for both errors is often recommended.

3.2.2 Agincourt HDSS data

Agincourt HDSS is located in the north-eastern part of South Africa, an area that shares the border with Mozambique. The surveillance site covers an approximate area of 475 km^2. The data collected is longitudinal, and the baseline census was carried out in 1992. The surveillance site is jointly run by the Medical Research Council and the University of the Witwatersrand. Agincourt HDSS is also a member of the International Network for Demographic Evaluation of Populations and Their Health (INDEPTH).

Besides regularly updating information on social and demographic characteristics of the population under surveillance since 1992, more than 20 censuses have also been conducted within the site. The updating of information is done on a yearly basis. The data can be accessed either at the entire surveillance site level or at village level. The latter is the lowest level at which this data can be accessed. In 1996, Agincourt HDSS covered a total of 20 villages, and these increased to 22 and 28 by 2001 and 2011 respectively. The estimated population at baseline was 57 000 from 8 900 households. The households increased to more than 12 000 by 2011. The population under surveillance is largely composed of Mozambican migrants and locals. A common characteristic among this population is the predominance of temporary migrants that maintain strong links with their rural homes covered by the site, yet they work and dwell outside the site. At analysis stage the temporary migrants were excluded from Agincourt to make it comparable to census data for the area covered by Agincourt HDSS that was drawn from SUPERCROSS, as will be explained in later sections.

The main strengths of this data, especially in relation to census data, is that it is regularly updated – a factor that improves quality in terms of both content and coverage. Secondly, the area covered by the surveillance site is small, which again reduces the creeping in of non-sampling errors noted above. Furthermore, like censuses it also free of sampling errors.

3.3 Data processing

3.3.1 Reconstruction of unadjusted census counts

StatsSA does not avail unadjusted counts for public use. The data can be obtained upon request. Even in case they agree to avail the data, they only give a small fraction of this data, often not above 5%. Permission to use this data was sought, but was not granted. Therefore the study applied the method proposed by Nyirenda et al. (2007) to reconstruct estimates of enumerated census counts from adjusted census counts. The method reconstructed estimates of enumerated census counts by using adjustment factors that StatsSA used to arrive at adjusted census counts from enumerated counts. Undercount rates have been used in South Africa's last three censuses for the adjustment of enumerated censuses (StatsSA, 2011). Firstly, broad adjustment classes of province, EA type, population group, and age categories of 0-19, 20-44 and 45+ were produced. Assumptions made over these adjustment classes were that the classes have the same coverage and undercount rates. As a result uniform adjustment factors were applied to each of these classes.

The procedure for computation of unadjusted data is based on the fact that adjusted counts were a result of enumerated counts multiplied by the adjustment factor. Therefore reconstruction of enumerated counts was achieved by firstly determining the proportion of population enumerated by reinventing the formula for determining the adjustment factor. The processes of computing the adjustment factor and proportion of population enumerated are described with examples below.

For a 10% undercount rate:

i) The adjustment factor would be: 1/1-0.1 = 1.1

ii) The proportion of population enumerated: (1-0.1)/1 = 0.9

For a 17% undercount rate:

i) Adjustment factor would be: 1/1-0.17 = 1.2

ii) The proportion of population enumerated: (1-0.17)/1 = 0.83

For a 14.5% undercount rate

i) Adjustment factor would be: 1/1-0.145 = 1.17

ii) The proportion of population enumerated: (1-0.145)/1 = 0.855

The reconstructed counts are consistently termed unadjusted counts in this study. The limitation of this method is that it may introduce further errors into unadjusted counts. However, the study has to improvise a credible way to manage this challenge of data restriction often impeding research in Africa. Census 1996 had both adjusted counts and their official unadjusted counts for provinces and this provided an opportunity to test how close the reconstructed unadjusted counts were to the official unadjusted counts. The outcome from this comparison was then assumed to reflect what would come from the other two censuses that did not have official unadjusted counts to match those that have been reconstructed.

Unadjusted counts were reconstructed at both national level, and at Sas level. At the former level, unadjusted counts were reconstructed from the 10% samples' adjusted counts. To match complete census counts, the adjusted counts drawn from respective censuses' 10% samples were weighted. At the latter level unadjusted counts for the area covered by Agincourt HDSS were reconstructed from complete adjusted counts from SUPERCROSS.

3.3.2 Constructed population estimates

Alternative population counts at national level were also needed in this study. The study therefore constructed population estimates using MORTPAK. In this study these counts will be consistently referred to as constructed population estimates. The software is a product of the United Nations Population Division and is used to compute various demographic estimates, e.g. for the entire world, for sub regions, and for specific countries. This includes Life Tables' construction, indirect estimations of either mortality, or fertility, and single-year population projections. The software uses the cohort component technique for projecting population counts.

MORTPAK has about 20 programmes that compute the various estimates as those listed above. The PROJCT programme was of interest to this study and was carried out in one-year population progressions according to age and sex up to 100 years. The construction of population estimates is built from a base population whose counts are grouped according to sex and five-year age groups. The information needed to construct the population estimate is listed in Table 3.2 below.

Table 3.1 Information for MORTPAK constructed population estimate

List of required information	Description of information
Year of base population	Indicates the year for the starting date of the projection
Month of base population	Indicates the month for the starting date of the projection
Day of base population	Indicates the day of the month for the starting date of the projection. Value must be between 1 and 31.
End year of projection	Indicates the ending year of the projection
Display/print projection results every (x) number of years	Indicates the print cycle for the projection results. For example, if a value of 5 is given, projections results are printed every fifth projection year
Final open-age group of base population	Indicates the final open-age group for the male and female population. Minimum 65+ and maximum at 85+
Sex ratio at birth	The sex ratio at birth must be between 0.75 and 1.5.
Model life-table pattern e.g. Coale-Demeny West	Indicates the model life-table pattern to be used, and the choice was Coale-Demeny North
Male base population	The male population by age for the base population
Female base population	The female population by age for the base population
Total fertility rate	Total fertility rates are required for the "initial projection period" and the "final projection period"
Net male migrants	Net migrants for males are required for the "initial projection period" and the "final projection period"
Net female migrants	Net migrants for females are required for the "initial

	projection period" and the "final projection period"
Life expectancy at birth for males	Life expectancy at birth for males are required for the "initial projection period" and the "final projection period"
Life expectancy at birth for females	Life expectancy at birth for females are required for the "initial projection period" and the "final projection period"
Current fertility pattern	The age-specific fertility rates corresponding to the first projection year.
Projected fertility pattern	The age-specific fertility rates corresponding to the last projection year.
Male Migration Pattern	The male migrants by age.
Female Migration Pattern	The female migrants by age
Label for user-defined model	This variable is used only if the model life-table pattern above is coded as zero
User defined model q(x,n) male values	This variable is used only if the model life- table pattern above is coded as zero. It consists of model male $_nq_x$ values supplied by the user.
User defined model q(x,n) female values	This variable is used only if the model life- table pattern above is coded as zero. It consists of model female $_nq_x$ values supplied by the user.

Source: UN MORTPAK for Windows Version 4.3 Manual 2013 pages 56-60

The main limitation of the procedure is that it is based on assumptions of constant mortality, fertility and migration for the projected period. However, the strength of this method is that it projects the counts for a yearly period.

3.3.3 Extraction of adjusted census counts for area covered by Agincourt HDSS

The adjusted counts for the area covered by Agincourt HDSS, for each of the three censuses, were drawn from SUPERCROSS. ArcGIS software was used to extract small areas layer (spatial data) completely falling within the Agincourt HDSS area. The data sets were then spatially overlaid in the GIS environment to extract SAs in the case of censuses 2001 and 2011 or EAs in the case of Census 1996, coinciding with Agincourt HDSS villages by running an overlay query for censuses 2001 and 2011. Three census counts by sex and five-year age groups were then extracted from small areas of each national census, in the municipality where Agincourt HDSS area belonged. The two data sets were joined by running a query that linked small areas extracted from ArcGIS to the national census data from SUPERCROSS. This produced counts from SUPERCROSS for each census, for the area covered by Agincourt HDSS. Unadjusted counts for the area covered by Agincourt HDSS were then reconstructed from the respective adjusted counts extracted from SUPERCROSS.

Agincourt HDSS counts for the years 1996, 2001, and 2011 were obtained in STATA format from the site's data processing unit. The counts were arranged according to both gender and five-year age groups. There were methodological differences for data collecting between censuses and Agincourt HDSS. The latter included temporary residents, whereas the former excluded them.

3.3.4 Exclusion of temporary migrants from Agincourt HDSS data

There were methodological differences between the census and Agincourt HDSS. A difference in methods of data collection between the two data sources that was of interest to this study was on how they recorded temporary migrants. The three censuses excluded temporary migrants among population counts for the area covered by Agincourt HDSS, yet population counts from Agincourt HDSS included temporary migrants. Therefore to make the counts from the two data sources comparable, the variable residence status in Agincourt HDSS was used to exclude temporary residents from Agincourt HDSS counts using STATA software.

3.3.5. Weighting

Adjusted data from the three censuses' 10% samples were weighted before use to ensure that the counts estimate actual adjusted counts. Each census had a weighting variable for the various data sets. STATA software was also used for weighting all the data used in this study.

3.3.6 Arranging counts according to relevant categories

For the convenience of the various analyses, counts were arranged according to desirable categories like five-year age groups, sex, race, and provinces.

3.4 Methods of analysis

This section outlines the methods that were applied to achieve each objective that was set in this study.

3.4.1 Estimation of levels of content error

This section describes the methods that were used to examine the level of content errors in the three censuses' adjusted counts. These counts should have no or less content errors if the PES was efficient in adjusting the undercount and high levels of content errors if the PES was inefficient. The analyses were done using indices. The main limitation of these indices is that they are based on the linearity assumption, which in certain instances is not met. However, their strength is that they test for content error mainly from a characteristic like age, which is known to be highly prone to reporting bias.

3.4.1.1 UN Age-Sex Accuracy Index

The Index measures the accuracy of age and sex distributions of a population by producing a joint score (UN Age-Sex Index) from age and sex ratios of the respective population. The first step is computation of age ratios and sex ratios.

$$\text{Age ratio } (5Arx) = 100 \quad \frac{5Px}{\frac{1}{2}(5Px\text{-}5 + 5Px\text{+}5)}$$

Where 5Arx = age ratio for ages x to x + 4; and 5Px = Population at ages x to x + 4. Large deviations of 5Arx from 100 suggest large errors in age reporting.

$$\text{Sex ratio } (5SRx) = 100 \quad \frac{5MPx}{5FPx}$$

Where 5SRx = sex ratio for ages x to x + 4; and 5MPx and 5FPx = male and female populations respectively, at ages x to x + 4. Also large deviations of this score from 100 suggest large errors in sex distribution.

The second step is combining the two scores to produce a joint score (UN Age-Sex Accuracy Index).

UN Age-Sex Accuracy Index = 3 * SRS + ARSM + ARSF

Where SRS = sex ratio score i.e. the mean difference between sex ratios for successive age groups (averaged regardless of whether positive or negative); ARSM = age ratio score for males i.e. mean deviation of the age ratios for males from 100%; and ARSF = age ratio score for females, i.e. mean deviations of age ratios for females from 100%, and in both cases it is regardless of signs. The South African censuses will be assumed to have accurate population distribution by age and sex if the UN Age-Sex Accuracy Index score is 20; inaccurate if is between 20 and 40; and excessively inaccurate if over 40.

An accurate score would suggest that there is less confounding effect of content error on findings from Demographic Analysis estimates for undercount.

3.4.1.2 Whipple's Index

The Index measures the extent of age heaping in respective censuses' data. The index is based on the assumption that certain digits are preferred in reporting for age than others. Precisely these are ages ending with digits 0 and 5.

The application of this technique in the South African censuses is guided by the formula specified below, which is often applied in investigating age preference when reporting for age, within an age range of 14 years to 73 years.

Whipple's Index = $\underline{\sum (P15 + P20 + P25 + P30 ... + P60 + P65 + P70}$

$$1/5\sum (P14 + P15+ P16 + \ldots P70 + P71 + P72 + P73)$$

Where ∑ = sum of, and P = population in the specified age number.

For each of the South African censuses analysed, when the Whipple Index below 105 suggests that there is no preference or heaping for or around ages ending with digits 0 or 5. As the index increases to a maximum of 500, this suggests errors in census data due to age preference or heaping on ages ending with the respective digits.

3.4.1.3 Meyer's Index

The Index measures digit preference or avoidance in reporting about age, for ages ending with not only 0 or 5 as is the case with the Whipple index, but ages ending with all digits from 0 to 9. The totals for each of the digits 0 to 9 are equally blended, and the assumption is that the totals for each of these ten digits should be 10% of the overall sum. Excess preference of a specific digit in age reporting will show a deviation from the 10% proportion of that specific digit to the overall grant percentage total, which is 100% for all the 10 digits. The same applies in the case of avoidance of a particular digit.

Steps followed when using this Index in analysing digit preference or avoidance in age reporting is as follows:

Firstly, it is the summation of populations ending in each digit over the whole range starting with the lower limit of the range (e.g., 10, 20, 30, 40… 11, 21, 31….)

Secondly, ascertain sum excluding the first population combined in step 1 (e.g., 20, 30, 40….; 21, 31, 41…)

Thirdly, weight the sums in steps 1 and 2 and add the results to obtain a blended population (e.g., weights 1 and 9 for 0 digit, weights 2 and 8 for 1, etc.)

Fourthly, convert distribution in step 3 into percentages.

Fifthly, take the deviation of each percentage in step 4% from 10% which is the expected value of each percentage.

Finally, a summary index of preference for all terminal digits is derived as one half of the sum of the deviations from 10.0%, without regard to signs.

A Meyers Index of 0 implies no digit preference when reporting for age, whereas 90 imply only one digit was preferred.

3.4.1.4 Bachi Index

This index is follows almost the same procedures as the Meyer's index. The only difference is that this index is drawn from deviations that were only positive. This is unlike the latter index which is based on a summation of either of the deviations regardless of either being negative or positive.

3.4.2 Estimation of coverage error

This section investigates coverage errors in adjusted counts. Less coverage errors indicates the efficiency of PES in adjusting counts for the undercount error. Coverage error indicates PES inefficiency. Methods used in this section are: graphical cohort analyses, age ratio analyses, sex ratio analyses, and growth rate analyses to test for coverage error by examining if population distributions from tested counts follow expected trends or patterns. However, there are certain distortive factors like migrations or sudden changes in mortality and fertility that can produce distorted tends or patterns. Hence such confounding factors lead to distorted distribution and erroneous conclusions that the data has coverage errors. The strength of these tests is the manner in which the data is presented, i.e. graphical. Such presentations of findings are vivid and articulate to readers. The strength of this approach is that the test for coverage errors is done by comparing exact counts rather than using weighted counts from

10% samples, which is likely to incur further bias during the weighting process. Furthermore, there are no assumptions associated with the analyses.

3.4.2.1 Comparison of adjusted census and Agincourt HDSS counts

Adjusted census counts by age and sex for the area covered by Agincourt HDSS were compared against respective counts from Agincourt HDSS data, for the three censuses. Respective line graphs were plotted. The extent of accuracy of the census counts was measured by the closeness of comparative lines and similarity in the distribution of counts across age groups.

3.4.2.2 Growth rate analysis

Developing countries are often associated with very high growth rates. Hence in examining the growth rates in South African censuses, the exponential growth rate was applied. Growth rates associated with less coverage errors range between 0 and 3.5, even in cases of excessive migrations. Any growth rate outside this range suggests undercount or over count, respectively.

The growth rates were computed per age group, for the two inter-census periods of Census 1996 to Census 2001, and Census 2001 to Census 2011.

The exponential growth rate is derived from the following formula:

$P(t) = P_0 e^{rt}$

Where:

$P(t)$ = final population at time t

P_0 = initial population at time $t = 0$

r = growth rate

t = time

3.4.2.3 Graphical cohort analysis

The technique is based on the fact that a birth cohort in any given population is expected to reduce in number over time due to mortality effect. This implies that a trace of each respective age cohort from the earliest census should show a decline trend in numbers over each successive census. Therefore, if population distributions by age are plotted from two or more successive censuses, the population distribution plot for the earliest census should be at the top, followed by second earliest, then third earliest, in that descending order, with the last census being at the bottom. Any inconsistence to such order, and any overlaps of the respective censuses' population distributions, is assumed to suggest coverage errors.

Two Graphical Cohort Analysis graphs, one for males and the other one for females were plotted.

3.4.2.4 Age ratios

When other possible confounding factors like migrations have no effect on a population, age ratios deviating from 100 suggest coverage errors.

Age ratios are computed as:

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

Where:

${}_nAR_x$ = age ratio for a given age group

nNx = population in that given age group (x)

nNx-5 = population in the age group that is n years before x

nNx+5 = population in the age group immediately after that given age group

3.4.2.5 Sex ratios

Sex ratios can also be an indicator of coverage errors, e.g. low sex ratio at birth can be an indicator of undercount of males' births, i.e. if other factors like migration have no effect.

The formula for computing sex ratios is:

$$ {}_{n}SR_{x} = \frac{{}_{n}N_{x\ males}}{{}_{n}N_{x\ Females}} * 100 $$

Where:

${}_{n}SR_{x}$ = sex ratio for that particular age

${}_{n}N_{x\ males}$ = male's population in that particular age

${}_{n}N_{x\ females}$ = female's population for that particular age

3.4.3 Which census counts are better estimates of South Africa's population?

The tests to determine which census counts between adjusted and unadjusted were better estimates of South Africa's actual population involved comparison of the two against respective counts from various non-census data.

3.4.3.1 Adjust for census undercount or not

Data for this section's analyses came from the three censuses and Agincourt HDSS. The analyses were at both national and SAs level. For analyses at national level, counts from three censuses' 10% samples and MORTPAK PROJCT programme's constructed population estimates were used. For analyses at SAs level census counts for the area covered by Agincourt HDSS that were extracted from SUPERCROSS and counts from Agincourt HDSS data were used.

The purpose of the investigation was to determine whether it was better to make adjustments for census undercount or not. HDSS counts were set as a gold standard based on the strengths of this data, as explained earlier. In the analyses, adjusted and unadjusted counts from the three censuses were compared to respective counts, both at national and SAs levels. The limitation of the analyses in this section is that the unadjusted counts used were reconstructed, and the process may have ushered further bias in the data. However, the strength of these analyses is that comparison of both adjusted and unadjusted counts are made against a credible gold standard. Furthermore, both adjusted counts for area covered by Agincourt HDSS and Agincourt HDSS counts are official counts that are not built on any assumption or weighting. Analyses were also robust since they were done at both national and SAs level. At the former level the analyses were for censuses 2001 and 2011. At the latter level the analyses were for the three censuses.

The adjusted counts at national level were drawn from respective censuses' 10% samples; adjusted counts for the area covered by the surveillance site were drawn from SUPERCROSS, as explained earlier. Unadjusted counts were then reconstructed from respective adjusted counts. Adjusted and unadjusted counts at national and SAs level for each census and respective counts from constructed population and Agincourt HDSS data were then arranged according to age and sex. Various descriptive tools were used to summarise and describe the findings from the comparisons of adjusted, unadjusted, and Agincourt HDSS counts.

3.4.3.2 Tables for adjusted, unadjusted and constructed population estimates' counts

These analyses compared adjusted and unadjusted counts against respective counts from a constructed population estimate, at national level. Comparisons were for five-year age groups and sex.

3.4.3.3 Tables for adjusted, unadjusted and Agincourt HDSS counts

Respective adjusted and unadjusted counts for area covered by Agincourt HDSS were then compared to respective counts from Agincourt HDSS. Comparisons were also for five-year age groups and sex.

3.4.3.4 Line graphs comparing adjusted, unadjusted, HDSS counts

Secondly, population distributions by age groups for adjusted and unadjusted counts for area covered by Agincourt HDSS relative to counts from Agincourt HDSS were plotted on line graphs.

3.4.3.5 Bar graphs comparing adjusted, unadjusted, HDSS counts

Thirdly, bar graphs were also plotted to compare total counts by sex for adjusted and unadjusted counts for area covered by Agincourt HDSS against respective counts from Agincourt HDSS.

3.4.3.6 Population pyramids comparing adjusted, unadjusted, HDSS counts

Finally, population pyramids from adjusted and unadjusted counts also for the area covered by Agincourt HDSS were plotted relative to the respective population pyramids from Agincourt HDSS's counts.

3.4.4 Compassion of PES and DA undercount estimates

Data came from the three censuses' 10% samples and constructed population estimates from MORTPAK's PROJCT programme. The analyses were at national level. The analyses in this section addressed the main research objective of this study. Overall undercount estimates for males, females, for both sexes combined, and for five-year age groups obtained from the two methods were compared. The compared undercount estimates were for censuses 2001 and 2011. The computed DA undercount estimates were for similar broad categories used in the PES adjustment process. This was achieved by averaging the undercount estimates for all five-year age groups from DA-constructed population estimates that coincided with respective broad categories from the PES. There were no comparisons for Census 1996 because the Census 1991 which was supposed to be the base population for constructed population estimates for 1996 was beyond the scope of this study.

The PES undercount estimates for censuses 2001 and 2011 for the various categories described above were obtained from StatsSA. They are also accessible from the StatsSA website. Alternative undercount estimates from DA were computed using constructed population estimates and respective unadjusted counts for respective censuses.

The formula for computing DA undercount estimates is:

$$\textbf{DA undercount estimate} = \left(\frac{\textbf{Constructed Pop estimate} - \textbf{Unadjusted counts}}{\textbf{Constructed Pop estimate}}\right) * 100$$

Source: US Census monitoring board, page 86

Where the DA undercount estimate refers to the undercount rate obtained from the DA method, the Constructed Pop estimate is the population count that was constructed using the DA method. Unadjusted counts are respective census counts that were reconstructed from adjusted counts.

Using this formula, DA's undercount estimates for males, females, males and females combined, and for five-year age groups were computed for censuses 2001 and 2011.

The limitations of analyses in this section are that firstly, only two censuses' undercount estimates were investigated. Secondly, the data management processes described above were also likely to further introduce bias into the counts. Also undercount estimates for Census 1996 were not investigated because the population estimate for 1996 could not be constructed since the base population of Census 1991 was beyond the scope of the study. However, the strength lies in the fact that at least more than one census was used in these analyses. Undercount estimates for the other two censuses still gave robust analyses since undercount estimates for various subpopulation groups were conducted. A further strength of these analyses is that by comparing undercount estimates from the two methods this was a direct test for the accuracy of the PES method as the undercount estimates were directly computed from their respective methods.

3.4.4.1 Comparisons for overalls

Bar graphs were used to compare respective PES and DA undercount estimates for males, females, and combine males' and females' overalls.

3.4.4.2 Comparisons for age groups

Line graphs were used to compare PES and DA undercount estimates by age groups.

3.4.5 Undercount effect on demographic parameters

Demographic parameters from adjusted census counts were computed and compared to those from "unadjusted" counts. Data from the 10% samples were used and the comparisons were at national level. The comparisons examined whether undercount in the country's three censuses underestimated, overestimated, or had no effect on demographic parameters. A limitation of the analyses in this section is that only fertility demographic parameters were used in the comparisons. However, these demographic parameters still achieved the purpose of the analyses which was to test if there are differences between estimates from adjusted counts and those from unadjusted counts.

3.4.5.1 Crude Birth Rate

Formula:

$$CBR = \frac{\text{Total number of births recorded in the census}}{\text{Total population estimated in the census}} * 1000$$

3.4.5.2 Parity Progression Ratio

Procedure for computation

For example, the probabilities of parity progression from birth order 0 (i.e. parity 0) to the next birth order (i.e. parity 1) was computed as:

$$PPR\ (0, 1) = \frac{\text{Number of women at parity 1}}{\text{Number of women at parity 0}}$$

From parity 1 to parity 2 as:

$$PPR\ (1, 2) = \frac{\text{Number of women at parity 2}}{\text{Number of women at parity 1}}$$

3.4.5.3 Age-specific Fertility Rates

Fertility rates for a specific age group were computed as:

$$ASFR = \frac{\text{Number of live births to women in specified age group}}{\text{Number of women in same age group}}$$

3.4.5.4 General Fertility Rate

This rate is computed as:

$$GFR = \frac{\text{Total number of births recorded in the census}}{\text{Total number of women aged 15 to 49 recorded in the census}}$$

3.4.5.5 Total Fertility Rate

Total fertility rate, e.g. for five-year age groups, used in this thesis was computed as:

$TFR = 5 \sum ASFR_{x}$ (for 5-year age groups)

Where:

$ASFR_x$ = age-specific fertility rate for women in age group x

3.4.6 Undercount effect on service delivery

3.4.6.1 The PRICEWATERHOUSECOOPERS method

The study applied the above method to estimate the undercount impact on service delivery using the BSGF allocations. This method has been applied in the USA to estimate the undercount effect on service delivery (Steffey, 1997; PRICEWATERHOUSECOOPERS Final Report to Congress, 2000). The method firstly determines the amount of funds from central government that are to be distributed. Subpopulation groups to benefit from the allocations are also identified and a formula that uses census counts to allocate the funds is determined. The subpopulations are then allocated funds based on the proportion of their population relative to total population of all benefiting subpopulation groups.

About R350 billion is budgeted for provincial disbursement every year in South Africa for various programmes. Such state funding will be used as proxy measurement for service delivery. This method was first used to estimate the impact of undercount on state funding among the USA's 31 states and the District of Columbia in 2000 for the projected inter-census period of 2002 to 2012. The method was built on the premise that funding received by a province in the case of South Africa, is determined by the proportion of the respective province's total population relative to the total population of the country. Each total population is as determined in the respective censuses. The proportion of each province's population in relation to the country's total population is then multiplied by the total amount of the national grant being distributed. In the case of the USA, grant allocations were distributed to states and counties, and these are replaced by provinces in the case of South Africa. Both adjusted and unadjusted counts for each province and at national level for all the three censuses were obtained.

The formulas for allocating funds for different grants were obtained from the Treasury, a subsection of the Department of Finance, mandated with the responsible of disbursing national grants. The Basic Services Grant of South Africa was the only grant whose formula was entirely based on census counts. The total amount allocated in a year was distributed to the country's nine provinces based on the size of each province relative to the overall national population size. The formula used to allocate Basic Services Grant to the nine provinces was: *(Provincial Pop/National Pop)*Total Funds Budgeted for particular year.*

1) The difference between allocations from adjusted counts and allocations from unadjusted counts for each province was its estimate of the effect of undercount on service delivery for a particular year.
2) The second step involved computing projections of undercount effect on service delivery for the entire inter-census period by province. For example inter-census

period for the first two censuses was five years, and so projections were done for five years. The inter-census period between the last two censuses was 10 years, and hence projections were for 10 years' period. The effect of undercount on service delivery for a projected period was obtained by multiplying the number of years from the respective inter-census period by the amount of funds a province was expected to have either been prejudiced or gained in a year.

The main limitation of this method was that adjusted counts were weighted and unadjusted counts were reconstructed. Both processes have the potential of introducing bias in the counts obtained. Also there was no direct measurement for service delivery; hence proxy measurements like BSGF and parliamentary seat allocations were used. However, the two proxies have been used in other studies conducted in the USA, and they yielded the expected results. The strength of the method is that analyses were done repeatedly as they involved three censuses. This made the analyses robust enough to compensate for limitations.

3.4.6.2 Comparison of parliamentary seat allocation

Parliamentary seat allocations based on adjusted data were compared to parliamentary seat allocation based on 'unadjusted' data. This was meant to give an estimate on the likely number of seats to be allocated per each province based on either two data sets.

According to the Independent Electoral Commission of South Africa (ICC), there are 200 parliamentary seats allocated to the nine provinces. The formula for their distribution during each national election is:

One parliamentary seat = 100 000 eligible voters, and these are individuals aged 18 years and above who are registered to vote in a particular election. This analysis was based on the assumption that all people 18 years and above, based on either adjusted and 'unadjusted' counts, were registered as voters in their respective provinces. Hereafter parliamentary seats to be expected, based on each of the two data sets for each province, were computed.

3.5 Software packages

Software packages that were used included STATA version 12. Firstly, the software was used for processing Agincourt HDSS counts into the respective statuses that they were in as at exact dates that coincided with enumeration dates for the respective censuses investigated in this study. The software was further used for all other data management that was deemed necessary for various analyses to be conducted, like data weighting and arranging counts according to desirable categories. For 10% samples data, the software was used for weighting and processing counts according to age and sex. For data from SUPERCROSS, the software was used for converting the data from Excel to STATA format. The data was already arranged according to five-year age groups and sex. The software was therefore only used to restrict the age groups to desirable cut-offs. For data from Agincourt HDSS, the software was used to exclude temporary migrants from the surveillance site's counts, and also for arranging counts according to age and sex.

ArcGIS was used for extracting the fully adjusted census counts for area covered by Agincourt HDSS. The mapping of these counts using this software was done for the three-census studies. Excel was used for plotting all graphs that were used to describe the various findings from the study. The software was also used for all computations that were needed for certain analyses to be conducted.

Population Analysis Spreadsheets (PAS) that were used included AGESEX Excel spread sheets for computing the UN Age-Sex Accuracy Index, together with age ratios, sex ratios, age-sex ratio score, and sex ratio score; SINGAGE Excel spread sheets for computing

Whipple and Meyers indices. TFR-GFR spreadsheets were also used for computing estimates of age-specific fertility rates.

3.6 Ethical issues

The use of secondary data did not present any serious ethical constraints for this research. Users of such secondary data only have to abide by the terms and conditions prescribed by custodians of the data. Above all, the study only investigated the PES and its undercount estimates and counts from relevant data sets, and did not use any information that related to details of individuals' characteristics or conditions.

CHAPTER 4: ESTIMATES OF LEVELS OF CONTENT ERROR

4.1 Introduction

The focus of this chapter was to test the levels of content error in the three censuses that were investigated. This type of error occurs when incorrect information on captured individuals is recorded. Examples of the error include misreporting of individuals' characteristics, like their age, level of education, and place of residence. The methods that were used to estimate levels of content error were: the United Nations Age Sex Joint Score, Whipple, Meyers, and Bachi's indices.

The test for the magnitude of content errors in adjusted counts directly investigated the accuracy of the PES. This is because adjusted counts were corrected for both content and coverage errors using this method. Therefore, the presence of content error in adjusted counts suggests inaccuracy of the PES, whereas absence of the error confirms accuracy of the PES. The importance of such findings to this study is that they contribute to reducing controversies around the PES as they determine whether the PES is accurate in adjusting errors or not. Furthermore, testing for content error is an initial step in evaluating census coverage, which was the focus of the next chapter. The presence of content error can confound outcomes from certain techniques used for estimating coverage error. Determining the magnitude of coverage error equally contributes towards reducing controversies around the PES, as is explained in the next chapter.

4.2 Test for content errors

The following subsections reported findings on levels of content errors that were confirmed from the different methods used.

4.2.4 UN Joint Age-Sex Score

When estimating level of content error in a census, a UN Joint age-sex score of <20 suggests accurate data, whereas >=20 to<=40 suggests inaccurate data, and >40 indicates highly inaccurate data. At national level this index suggested overall accurate census data with regards to reporting different characteristics among individuals captured in all three censuses. Tests for the index were done on census data that firstly included both South African citizens and immigrants, and secondly on census data that excluded immigrants. From the data that included both, e.g. Census 1996, the sex ratio score, age ratio score for males and females, and UN joint age-sex score were 3.3; 4.2; 3.9; and 18.1 respectively, all suggesting high levels of internal consistence within the respective census data. The indices slightly changed upwards when excluding immigrant counts from the tests, but overall remained within the range, which suggests internal consistence within the data. The other two censuses also had indices that suggested high levels of internal accuracy within their data. [Table 4.1]

Indications are that internal consistence was best in Census 2011, where sex ratio score, age ratio score for males and females, and UN joint age-sex score reduced to 3.3; 3.8; 4.1; and 17.8 respectively, when including both the citizens and immigrants. Even when excluding the immigrants, the scores remained lower compared to corresponding scores from the preceding censuses' data. However, the UN joint age-sex score for 2001 when excluding immigrants was an outlier, as it was the only one from the three censuses that was not less than 20. [Table 4.1]

Table 4.1 UN Joint Age-Sex Scores for the three censuses

Country and Province	1996 census				2001 census				2011 census			
	SRS	ARS		UNS	SRS	ARS		UNS	SRS	ARS		UNS
		M	F			M	F			M	F	
National Incl immigrants	3.3	4.2	3.9	18.1	3.6	4.3	4.6	18.1	3.3	3.8	4.1	17.8
National Excl immigrants	3.6	4.3	4.0	19.2	3.7	4.3	4.6	20.0	3.2	3.5	3.9	17.0

SRS=sex ratio score; ARS=age ratio score; UNS=United Nations Joint age-sex Score, M=male; F=female; Incl= including; Excl=excluding.

4.2.5 Whipple's indices for the three censuses

Table 4.2 below describes evaluation criteria for age heaping using the Whipple's index as defined by the United Nations. Five categories are demarcated for the index. When the index predicts highly accurate data with regards to age heaping, the score has to be less than 105, and the departure from perfect is estimated to be less than 5%. When data is said to be bad, i.e. when age preference/heaping is around digits ending with 0 or 5, the index would be above 124. The percentage departure from perfect will be at 25% or above. [Table 4.2 here]

Table 4.2 United Nations evaluation criteria for age heaping using Whipple index

Whipple Index	Data quality	Variance from perfect
<105	Highly accurate	<5%
105-109.9	Fairly accurate	5-9.99%
110-124.9	Approximate data	10-24.99%
125-174.9	Bad data	25-74.99%
>175	Very bad data	>=75%

The Whipple's indices described here were for census data at national level. Indications from Census 1996 are that Whipple's indices for males, females and for both sexes combined remained the same at 101. The three indices suggest highly accurate data with regards to age heaping. Such a suggestion is however inconsistent with results in Table 4.2, which clearly suggested consistent age heaping around digits ending with 0 in the 1996 Census. Census 2001 has the lowest scores of Whipple's indices for males, females, and for the combined sexes, i.e. 97 for each of the three categories. The indices were almost the same for Census 2011, as the indices for females and for both sexes were at 97, with that for males just one up, at 98. [Table 4.3]

Table 4.3 Whipple Indices for the three censuses

Year of census	Male	Female	Total
1996	1.01*101=101	1.01*100=101	1.01*100=101
2001	0.97*100=97	0.97*100=97	0.97*100=97
2011	0.98*100=98	0.97*100=97	0.97*100=97

4.2.6 Meyer's Index

The Meyer's Index ranges between 0 and 90, the former denoting no age heaping at any of the terminal digits, and the latter denoting extreme age heaping at one of the digits. As for the males, ages ending with digits 0, 4, 6 and 8 suggested slight over-reporting. Their deviations from away from 10% were 0.4, 1.2, and 0.3 respectively. The other age ending with digits 1, 3, 5, 7, and 9 were reportedly under-reported, though also not high to suggest bad data. The deviations from 10% suggested for all digits among females also did not suggest bad data. However, there was slight over-reporting as well as under-reporting of certain digits when reporting on age, as was the case in males. Over-reported digits in females were the same as for males. For combined sexes, over-reporting or under-reporting of digits in reporting on age was very minimal across the entire range of digits. For combined sexes preferred digits, i.e. those over-reported that had the highest deviations from 10%, were 6 with 1.2, followed by 0 with 0.5, and avoided digits, i.e. those under-reported that had the highest deviations were 9 which had -0.7, followed by 7 which had -0.4. The overall indices were 4.4 for males, 4.9 for females and 4.6 for combined sexes. All the indices suggested no significant preference or avoidance for any of the ages ending with digits between 0 and 9 [Table 4.4].

The Meyer's indices for males, females, and for combined sexes improved to 2.7 for each of the three categories noted for Census 2001. This was naturally expected to follow since deviation from 10% for all ages ending with digits 0 to 9 had also reduced for males, females, and for combined sexes in Census 2001. The highest deviation among preferred ages was for digits ending with 1 for males, females and for combined sexes, all recording 0.8. The highest deviation among avoided digits was: -0.4, -0.6, and -0.5 for males, females and for combined sexes, all being for digit 1 [Table 4.4].

Table 4.4 Meyer's Indices for the three censuses

Terminal Digit	Census year								
	1996			2001			2011		
	Ma le	female	total	male	female	total	male	female	total
0	0.4	0.5	0.5	-0.2	-0.3	-0.3	-0.1	-0.3	-0.2
1	-0.4	-0.5	-0.5	0.8	0.8	0.8	0.6	0.6	0.6
2	-0.1	0.1	0.0	-0.2	-0.1	-0.2	-0.1	-0.1	-0.1
3	-0.4	-0.3	-0.3	-0.2	-0.1	-0.1	-0.2	-0.1	-0.1
4	0.4	0.2	0.3	-0.4	-0.6	-0.5	-0.4	-0.4	-0.4
5	-0.2	-0.4	-0.3	-0.2	-0.1	-0.2	-0.1	-0.1	-0.1
6	1.2	1.2	1.2	0.0	0.0	0.0	0	0	0
7	-0.4	-0.4	-0.4	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
8	0.3	0.4	0.3	0.1	0.2	0.1	0	0.1	0.1
9	-0.6	-0.8	-0.7	0.5	0.5	0.5	0.3	0.3	0.3
Meyer's Index	**4.4**	**4.9**	**4.6**	**2.7**	**2.7**	**2.7**	**1.9**	**2.2**	**2.0**

The Meyer's indices obtained for the three censuses suggest that the internal quality of data has been improving. The Meyer's indices for Census 2011 improved to 1.9, 2.2, and 2.0 for males, females, and for combined sexes respectively. In the previous two censuses, there seems to have been a general pattern of preference for positive digits and avoidance of those which were negative. However, Census 2011 was highly inconsistent with such patterns. For example, positive digits like 2 and 4 were mostly avoided than being preferred. The only positive digit that was mostly preferred rather than being avoided was 8. Negative digits that were mostly preferred rather than avoided were 1 and 9. [Table 4.4]

4.2.7 Bachi Index

With regard to the Bachi Index, the range within which the index falls is from 0 to 180. The former implies no age heaping at all, whereas the latter implies that all the ages were reported at one digit. The Bachi indices for males, females, and combined sexes for all the censuses were below 4, the highest being the index for females from Census 1996 which was 3. This suggests that the censuses' data had less internal inconsistence with regards to age preference or avoidance. In Census 1996, age preference was noted for digits ending with 0, 4, 6, and 8 among males. For females this was for digits 0, 2, 4, 6, and 8. The ages which ended with negative digits were largely avoided. As for both censuses 2001 and 2011, percentages of age preference or avoidance across all ages ending with digits 0 to 9 were overall very small. [Table 4.5]

Table 4.5 Bachi Indices for the three censuses

Terminal Digit	Census year								
	1996			2001			2011		
	male	female	total	male	female	total	male	female	total
0	0.6	0.8	0.7	-0.4	-0.4	-0.4	-0.4	-0.6	-0.5
1	-0.9	-0.8	-0.8	1.0	1.0	1.0	0.6	0.5	0.6
2	-0.2	0.0	-0.1	-0.1	0.0	0.0	-0.1	-0.1	-0.1
3	-0.5	-0.3	-0.4	0.0	0.0	0.0	-0.2	-0.1	-0.1
4	0.4	0.2	0.3	-0.4	-0.7	-0.6	-0.4	-0.4	-0.4
5	-0.1	-0.4	-0.3	-0.3	-0.2	-0.2	-0.2	-0.1	-0.1
6	1.5	1.6	1.5	-0.2	-0.2	-0.2	0.1	0.1	0.1
7	-0.4	-0.4	-0.4	-0.2	-0.2	-0.2	-0.1	0	0
8	0.4	0.5	0.5	0.0	0.1	0.1	0.1	0.2	0.2
9	-0.7	-1.0	-0.8	0.6	0.5	0.6	0.4	0.4	0.4

Bachi's Index	**2.9**	**3.0**	**2.9**	**1.6**	**1.7**	**1.6**	**1.2**	**1.2**	**1.2**

4.2.9 Concluding remarks

The findings from various tests conducted in this chapter largely indicated that adjusted counts had insignificant levels of content error. The data tested for content error is already adjusted for the error using the PES. For this reason the data should have minimal or no content error. As expected the findings confirmed that PES adjusted counts had minimal levels of content error. Such findings therefore supported the view that the PES was credible; as its adjusted counts for the three censuses indicated insignificant levels of content errors. In a way such findings contribute in reducing controversies associated with South African censuses which questioned accuracy of PES in estimating and adjusting for undercount.

CHAPTER 5: ESTIMATION OF COVERAGE ERROR

5.1 Introduction

This chapter evaluated coverage error in the country's three censuses. The level of coverage error in a census indicates both accuracy and completeness of the respective census. The methods used to estimate coverage error in this study included: comparison of adjusted counts against respective Agincourt HDSS counts, growth rate analyses, graphical cohort analyses, and age and sex ratio analyses. The selection of these methods was based on their applicability to the data available.

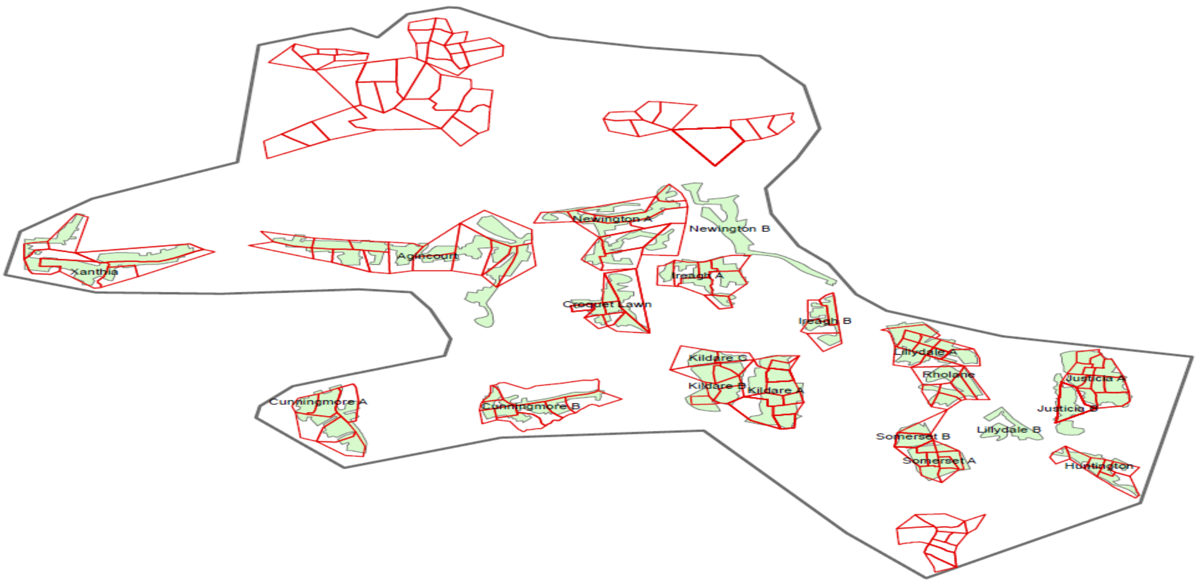
Findings suggesting a high level of coverage error in the respective censuses would suggest that the PES is not an efficient method for estimating and adjusting the undercount in these censuses. Such findings would be confirming that the criticism of the PES as an inaccurate method in correcting the undercount is credible. Whereas, suggestions of completeness of coverage in these censuses would suggest that the PES adjustment process was credible, thereby confirming the accuracy of the PES. Either way, the findings would reduce controversies as there would be empirical evidence that back up the assertion.

5.2 Matching of coinciding boundaries

The area covered by Agincourt HDSS had 20 villages in 1996. After an overlay of Census 1996's EAs over Agincourt HDSS villages' boundaries, three conglomerations of EAs had no coinciding villages. Two are on the upper boundary of Agincourt HDSS area and one at the lower boundary. There were also villages which never coincided with any EAs boundary, i.e. Newington B and Lillydale B. The third scenario was when EAs boundaries overlapped beyond the boundaries of coinciding villages' boundaries. Finally, there were instances when villages' boundaries overlapped coinciding boundaries of EAs, e.g. Agincourt, Justicia A and B, and Kidare B.

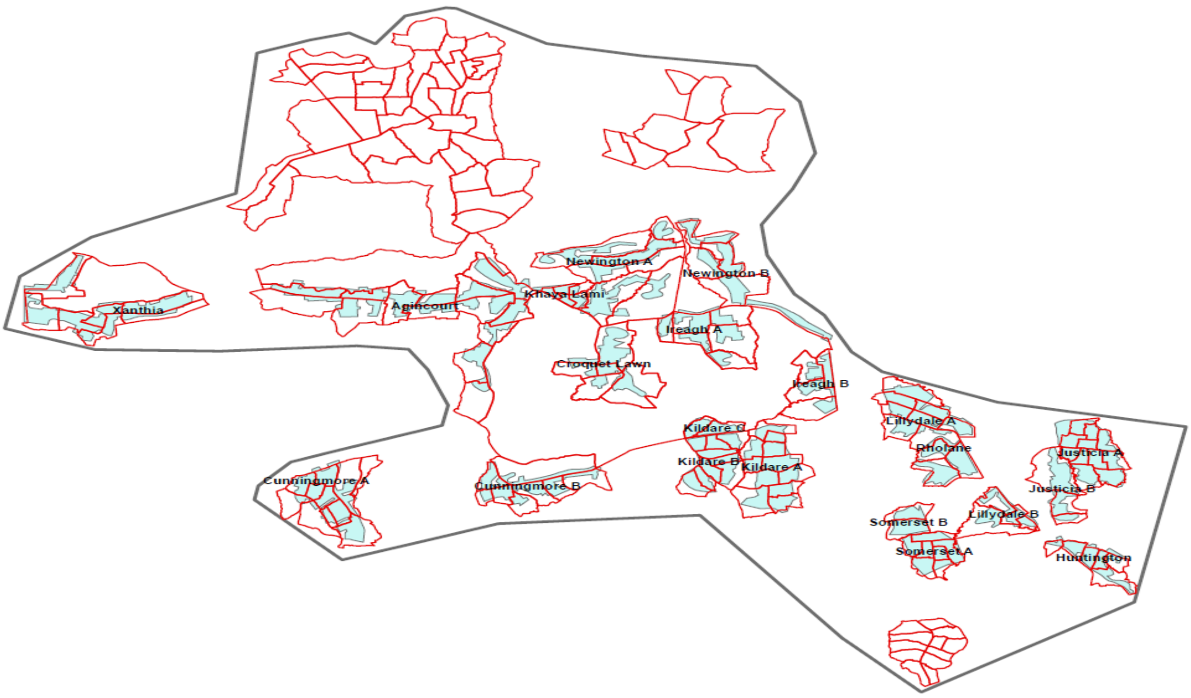
The first and third scenarios should lead to more census counts for the area covered by Agincourt HDSS relative to counts from the surveillance site. The second and fourth scenarios should lead to access of counts from the surveillance site relative to census counts. The effect from the contrasting scenarios should cancel each other out, hence expectations would be that census and Agincourt HDSS counts should be the same. [Fig 5.1]

Fig 5.1 Overlay for Agincourt HDSS area, 1996



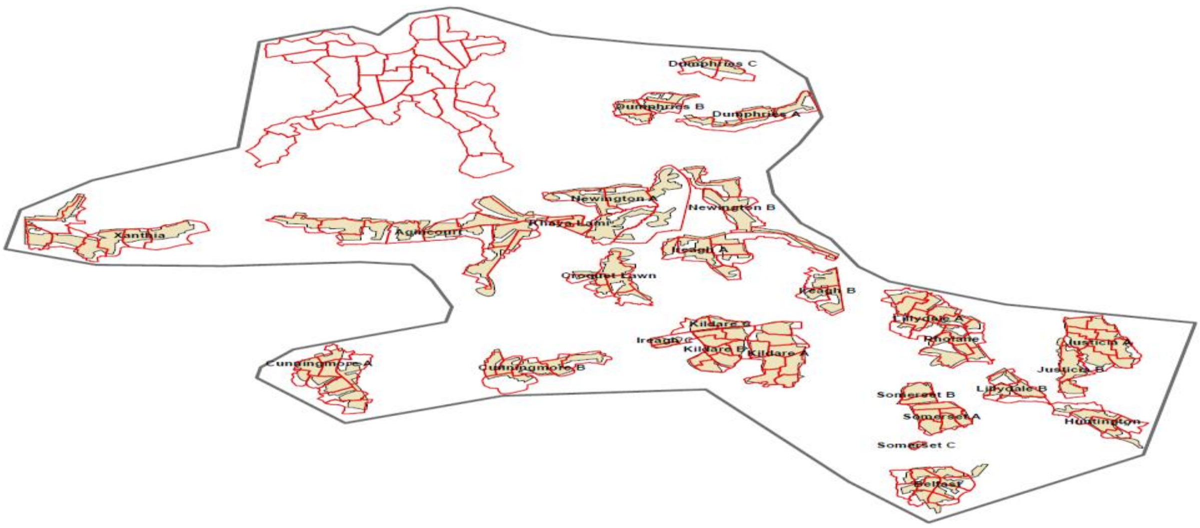
Villages from Agincourt HDSS area had increased to 22 by 2001. GIS mapping of Agincourt HDSS villages with coinciding small areas from Census 2001 resulted in almost all villages being completely submerged in the latter's boundaries. The only exceptions were in small strips of Newington B and Justicia A villages. This meant that, unlike in the 1996 scenario, 2001 mapping was mainly dominated by boundaries of small areas overlapping boundaries of villages. Therefore it is expected that counts for Agincourt HDSS area obtained from Census 2001 should be slightly more than those from Agincourt HDSS data. [Fig 5.2 here]

Fig 5.2 Overlays for Agincourt HDSS area, 2001



The 2011 overlays for small areas and village boundaries for the area covered by Agincourt HDSS produced an almost perfect match. There were only a few slight overlaps of boundaries of small areas over boundaries of villages. It was indicated that the majority of the boundary overlays for small areas and villages approximately matched with 95% or better. For instance, the matching of boundaries of small areas and boundaries of villages in Somerset A, B, and C, and Justicia A and B were almost 100%. It is therefore expected that firstly due to a few cases of boundaries of small areas overlapping boundaries of villages, adjusted census counts should still be higher than Agincourt HDSS counts. Secondly, the difference between adjusted census and Agincourt HDSS counts should be very slight. [Fig 5.3 here]

Fig 5.3 Overlays for entire Agincourt HDSS area, 2011

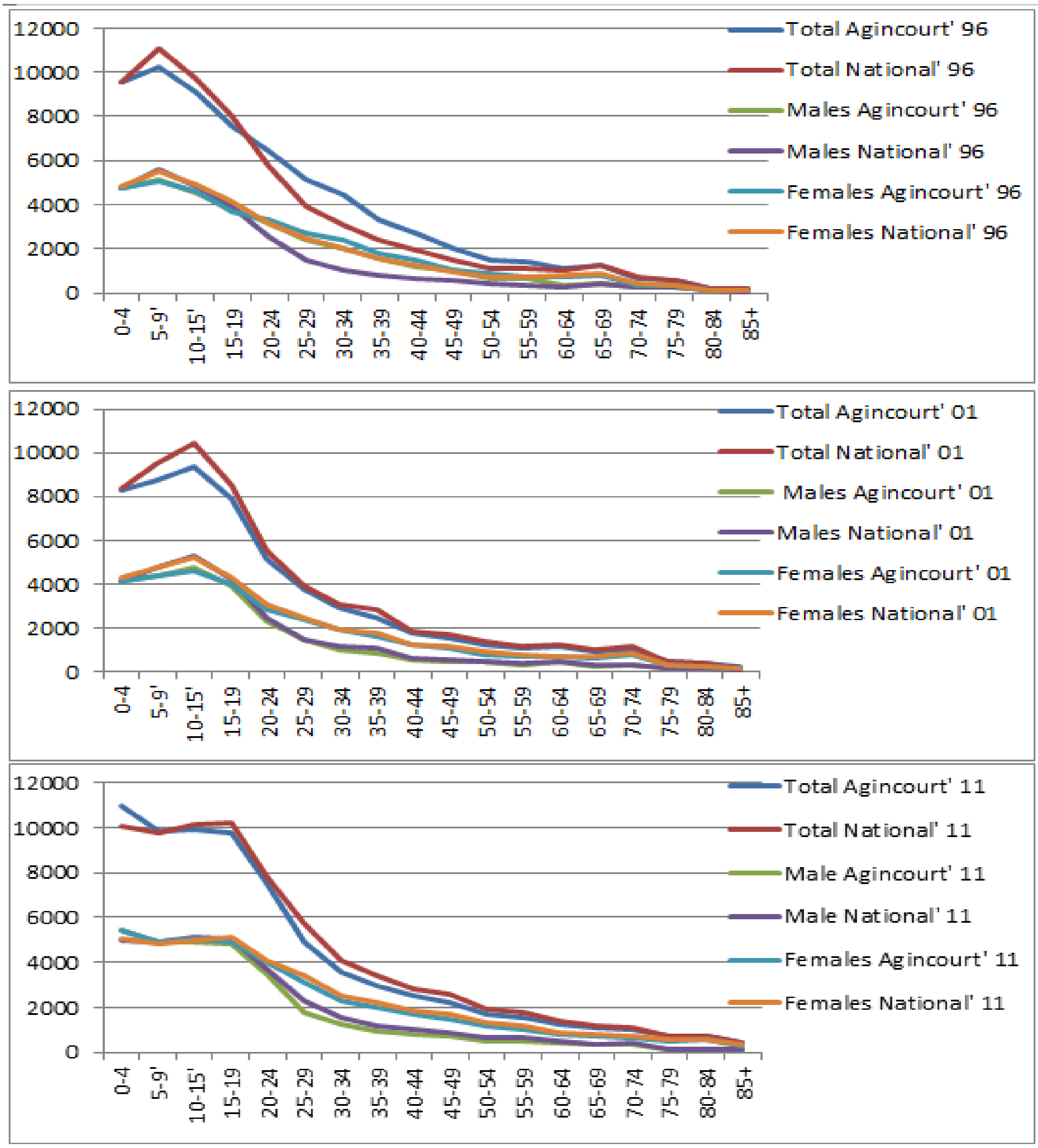


5.3 Matching of adjusted count and Agincourt HDSS counts

The analyses were carried out by five-year age groups for males, females, and for totals from the combined counts of both sexes. The purpose was to establish that the adjusted census counts were the same as the respective counts of the surveillance site. It was indicated that the adjusted counts of the three censuses closely approximated respective counts of Agincourt HDSS. This largely suggested the accuracy of the adjusted census counts. For example in the case of 2001 and 2011 comparisons, this confirmation was evidently clear across all age groups, in the three categories whose counts were analysed. Besides the close matching of the compared counts, further confirmation of the accuracy of the adjusted counts was that the trends across all age groups of the compared counts matched each other very well. Most important was the fact that the adjusted counts were slightly higher than the respective counts of Agincourt across age groups. Such a finding was consistent with the outcome of the mapping of census counts by ArcGIS, where boundaries of small areas for censuses 2001 and

2011 slightly exceeded the coinciding boundaries of villages covered by Agincourt HDSS. This meant extracted census counts for areas covered by Agincourt HDSS would be slightly higher than respective counts of Agincourt HDSS. Regarding the 1996 comparisons, it was also indicated overall that the adjusted counts closely matched respective counts of Agincourt HDSS, and that counts for middle-aged groups deviated clearly. [Figs 5.4 to 5.6]

Fig 5.4-5.6 Matching of Agincourt and adjusted census count



5.4 Assessment of growth rates

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 for population change (Moultrie et al, 2013). If growth rates approach or overlap these limits and there have been no heavy migrations, this suggests that the respective counts may have been affected by coverage errors.

5.4.1 Growth rates for total

Growth rates were computed for five-year age groups starting with 0-4 up to open-age group of 90 years and above. Total growth rates for males and females combined between 1996 and 2001, were within the expected range, except for age groups 0-4, and 75 to 79 years. Both age groups had growth rates that were below 0. Growth rates for the period 2001 to 2011 also fell within the expected range, except for age groups 5-9, and 10-14 years. Overall growth rates for the two periods were consistent with those expected. Even after disaggregating by sex, the computed growth rates are largely within the expected range, and only a few would be outside the range. [Tables 5.1]

Table 5.1 Growth rates by age group, and for total counts

Age group (Years)	1996 Census	2001 Census	Growth Rate (r) (%)	2001 Census	2011 census	Growth rate (r) (%)
0-4	4439089	4429376	-0.04	4429376	5685452	4.99
5-9	4672396	4849900	0.75	4849900	4819751	-0.12
10-14	4668724.6	5052172	1.58	5052172	4594886	-1.90
15-19	4183880	4982066	3.49	4982066	5003477	0.09
20-24	3982645	4289409	1.48	4289409	5374542	4.51
25-29	3452278	3927371	2.58	3927371	5059317	5.07
30-34	3069629	3337991	1.68	3337991	4029010	3.76
35-39	2654043	3073972	2.94	3073972	3467767	2.41
40-44	2131860	2607302	4.03	2607302	2948618	2.46
45-49	1672336	2088518	4.45	2088518	2620283	4.54
50-54	1264881	1641695	5.22	1641695	2218289	6.02

55-59	1069173	1201962	2.34	1201962	1797408	8.05
60-64	889192.932	1070172	3.71	1070172	1385768	5.17
65-69	762688.427	788238.21	0.66	788238.21	957804.66	3.90
70-74	481885.622	626843.65	5.26	626843.65	750487.93	3.60
75-79	379902.926	368945.57	-0.59	368945.57	481365.76	5.32
80-84	178267.524	271922.76	8.46	271922.76	322200.62	3.39
85-89	91116.65	95840.1802	1.01	95840.1802	146403.64	8.47
90+	46089.518	64986.494	6.87	64986.494	107728.37	10.11
Total	40090077.4	44768684.4	2.21	44768684.4	51770560.2	2.91

5.4.2 Provincial growth rates

All provincial growth rates were within the expected range except for the Northern Cape, i.e. for the inter-census period of 1996 to 2001. Gauteng had the highest growth rates, followed by Mpumalanga and KZN respectively. During the 2001 to 2011 inter-census period, all provinces except North-West had growth rates that were between 0 and 3.5. The growth rates in the 10-year inter-census period for respective provinces were generally lower compared to the growth rates in the five-year inter-census period. [Fig 5.2]

Table 5.2 Provincial growth rates

Province	1996 Census	2001 Census	Growth Rate (r) (%)	2001 Census	2011 census	Growth rate (r) (%)
1	3914326	4513206	2.85	4513206	5822734	2.548
2	6252014	6415451.7	0.52	6415451.7	6562053	0.226
3	831156.834	823429.16	-0.19	823429.16	1145861	3.304
4	2603678	2715587	0.84	2715587	2745590	0.11
5	8300918	9420961	2.53	9420961	10267300.4	0.86
6	3327187	3662194	1.92	3662194	3509953	-0.425
7	7247417.9	8830155	3.95	8830155	12272262.9	3.292
8	2749157.5	3125664	2.57	3125664	4039939	2.566
9	4864221.4	5262037	1.57	5262037	5404868	0.268

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

5.4.3 Growth rates by race

Growth rates for South Africa's four major races, i.e. black, white, coloured, and Asian people were also examined for both inter-census periods. All races except for the white people in the five-year inter-census period recorded growth rates that fell within the recommended range. Growth rates for black people in both inter-census periods were higher than in any other race. The coloured race's growth rates were second, followed by the Indian and lastly the white race, in both inter-census periods. [Fig 5.3]

Table 5.3 Growth rates for the four main races

Race	1996 Census	2001 Census	Growth Rate (r) (%)	2001 Census	2011 census	Growth rate (r) (%)
Black	31141366.7	35358487.9	2.54	35358487.9	41000937.5	1.481
coloured	3593966	3994571	2.114	3994571	4615401	1.445
Asian/Indian	1045165	1118985.5	1.365	1118985.5	1286930	1.398
Whites	4426121	4296640	-0.59	4296640	4586838	0.654

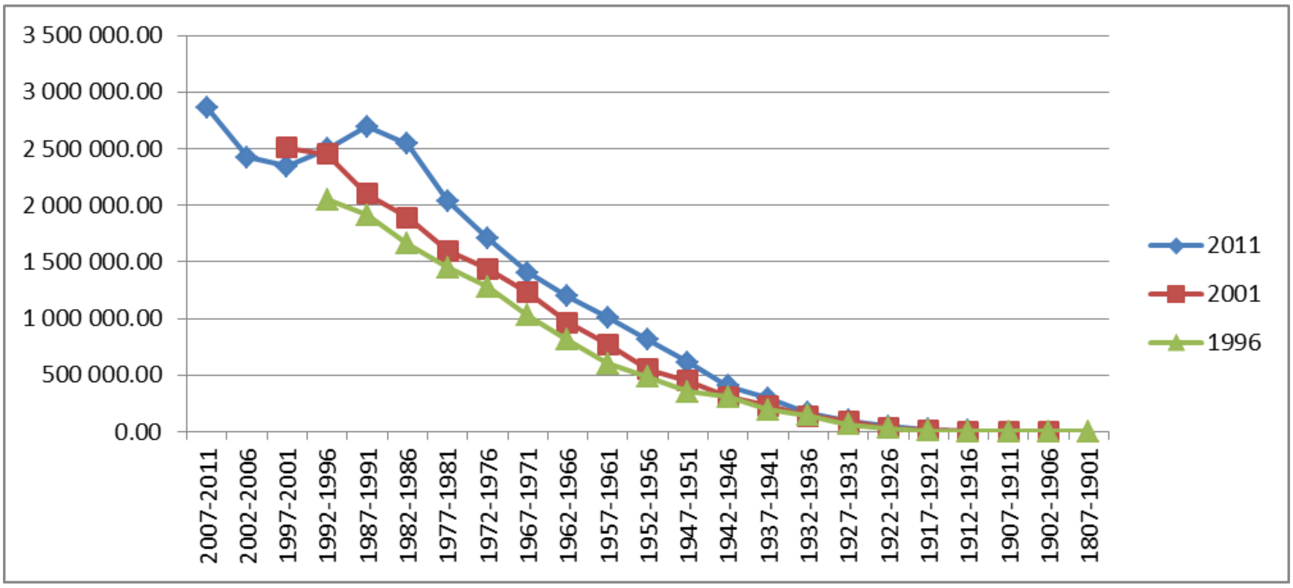
5.5 Graphical Cohort Analysis of the three census enumerations

The three lines representing population distribution by age for the counts of each census (adjusted) should follow the same trend, and should never cross (Keane et al., 1985). Violations of these conditions suggest coverage errors.

5.5.1 Males

The graphical cohort analysis for males in the counts of the three censuses generally does not conform to distributions expected from accurate census counts. Firstly, Census 2001 and 2011 cross each other at age group 20-24 years. Secondly, there are deviations in terms of the order of arrangement expected from the population distribution of the three censuses. The population distribution for Census 1996 is expected to be at the top, followed below by that of Census 2000, and lastly Census 2011. However, the order of distribution is opposite: the population distribution of Census 2011 is at the top, and that of 1996 at the bottom. Another notable characteristic on the relationship among the three lines is that they are wider apart at younger ages, especially censuses 2001 and 2011. With increase in age, the three lines seem to eventually converge at around 75 years onwards. [Fig 5.7]

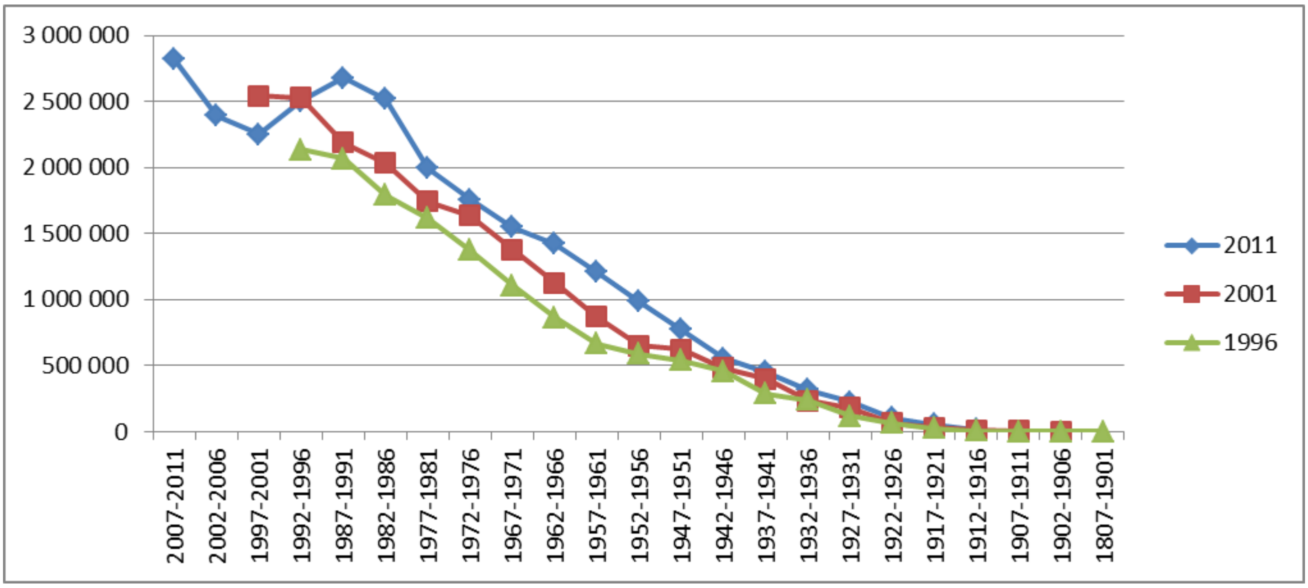
Figure 5.7 Graphical Cohort analyses, males



5.5.2 Females

The graphical cohort analysis of females was similar to that of males in that there was an overlap between the population distribution lines for censuses 2001 and 2011 at age group 20-24 years. A further similarity is that generally the population distribution line of Census 1996 remained largely at the bottom, with 2011 at the top, followed by Census 2001. However, a notable difference is that the female graphical cohort analysis indicated population distribution lines with constant ups and downs which were not consistent with each other as they occurred at different points among the different lines. Moreover, though the population distribution lines for the three censuses eventually merge, this occurs at later ages (i.e. 90 years onwards) relative to that of males. [Fig 5.8]

Figure 5.8 Graphical Cohort analyses, females



5.6 Age ratios

The PES adjusted age ratio counts of males in all three censuses were inconsistent overall compared to those expected from censuses that suggest no errors. Census 1996 in particular had some age ratios where the deviations from expected age ratios were too high. For instance, the ratios age group 65-69 were 112, and 89 for those aged 60-64 years. Since age ratios were computed per 100, this would be the expected age ratio if census counts are to be presumed free of error. The age group with age ratios that were closer to those expected was 40-44 years. Census 2001 appears to have the worst age ratios in terms of deviating from those expected. Except for the age group 35-39 years, the age ratios for all other age groups were clearly away from 100, and either below or above 100. In fact, among all age ratios below 100 all were below 97. Census 2011 probably had census counts that produced age ratios that were better in terms of being close to those expected. Age groups 50-54 and 60-64 had age ratios that were just slightly above 100, whereas ratios of age groups 15-19 and 55-were exactly 100. On the other hand, age ratios for age groups 35-39, 40-44, and 45-49 years had odds ratios which were slightly below 100. Among computed age ratios, deviations in age group 65-69 years were the worst in Census 2011. [Table 5.4]

Table 5.4 Males' age ratios from PES adjusted counts relative those expected

Age group	1996 census	Age ratio		2001 census	Age ratio		2011 census	Age ratio	
	Count	PES	Ex	Count	PES	Expec	Count	PES	Expec
0-4	2215745			2214369			2867585		
5-9	2340363	103.4	100	2423906	102.6	100	2425181	93.1	100
10-14	2309587	105.3	100	2510361	102.9	100	2344275	95.2	100
15-19	2048208	97.0	100	2454284	106.5	100	2498572	99.2	100
20-24	1914992	103.2	100	2100064	96.6	100	2694646	106.9	100
25-29	1661866	98.7	100	1893200	102.4	100	2542682	107.5	100
30-34	1452053	98.8	100	1596760	95.9	100	2036206	95.8	100
35-39	1278644	103.2	100	1438418	101.8	100	1709347	99.4	100
40-44	1026535	98.3	100	1230423	102.5	100	1402328	96.5	100
45-49	809068	99.5	100	962658	96.2	100	1195740	99.1	100
50-54	599709	92.9	100	770704	101.8	100	1011349	100.7	100
55-59	482332	101.4	100	551102	90.5	100	811950	100.0	100
60-64	351752	89.1	100	447549	104.5	100	612364	100.9	100
65-69	307073	112.3	100	305169	90.1	100	401548	88.3	100
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

Female age ratios almost followed the same trends as those exhibited by males. None of the age groups had counts that produced age ratios of 100 in Census 1996. Just as in the case of males, age group 65-69 had the worst age ratios of 110.6, which was just slightly lower than that for males in the same age category. Among age ratios higher than 100, age group 35-39 had an age ratio of 101, which was the closest to the expected 100. On the other hand, among the age ratios below 100, age group 40-44 whose age ratio was 98.7 was the closest to the expected ratio. As for Census 2001, no age ratio was at 100. In those age groups with age ratios below 100, age group 50-54 with 98 had the closest age ratios, as was expected. In the age groups with age ratios above 100, age group 5-9 whose age ratio was 102 was the closest to expected. The age group with age ratios that had the worst deviation from expected was

60-64 years, with 109.8. Census 2011, just as in the case with males had some age groups that had age ratios that were almost at 100. These were 50-54, ad 55-59 and 60-64, where the age ratios were 101.2, 99.5 and 100.3 respectively. [Table 5.5]

Table 5.5 Females PES adjusted census counts' age ratios relative to those expected

Age group	1996 census	Age ratio		2001 census	Age ratio		2011 census	Age ratio	
	Count	PES	Expd	Count	PES	Expd	Count	PES	Expd
0-4	2223343			2215008			2817867		
5-9	2332033	101.8	100	2425994	102.0	100	2394570	94.5	100
10-14	2359138	105.6	100	2541811	102.6	100	2250611	91.9	100
15-19	2135672	96.5	100	2527782	106.9	100	2504905	101.6	100
20-24	2067653	105.3	100	2189344	96.0	100	2679896	106.7	100
25-29	1790412	97.2	100	2034172	103.5	100	2516635	107.7	100
30-34	1617576	102.2	100	1741231	94.9	100	1992804	93.2	100
35-39	1375399	101.0	100	1635554	104.9	100	1758420	99.4	100
40-44	1105325	98.7	100	1376879	99.7	100	1546291	97.2	100
45-49	863268	97.5	100	1125861	100.2	100	1424543	103.5	100
50-54	665172	91.7	100	870991	98.0	100	1206940	100.2	100
55-59	586841	97.6	100	650860	87.2	100	985458	99.5	100
60-64	537440	103.1	100	622623	109.8	100	773404	100.3	100
65-69	455615	110.6	100	483069	94.8	100	556257	90.7	100
Missing	236530			0			0		

5.7 Sex ratios

The sex ratio at birth for Census 1996, estimated from 0-4 age, was below expected as it was even below 100. At age 5-9 years, the sex ratio had increased to slightly above 100. From ages 10 years until 34 years the sex ratios follow a decreasing trend as should be expected. However, the magnitudes of declines in these sex ratios appear to be excessively high. For instance, at age 15-19 sex ratios declined to as much as 95.9, then to 92.6 for 20-24 years, to 89.9 and 82.2 for 34-35 years and 55-59 years respectively. At age group 75-79, the sex ratio had fallen to as low as 59.2, suggesting that the number of males in that age group had reduced to almost half compared to females in the same age group. The overall sex ratio for Census 1996 was 92.7. Counts from Census 2001 produced sex ratios that suggested consistent declines with increasing age, a trend that should be expected. However, like in the previous census, the decline of sex ratios with increase in age appears to have been too rapid. For example, at early ages like 30-34 years the sex ratio had fallen to 87.9, and to 85.5 at age group 40-49 years and to 71.9 at age group 60-64. The overall sex ratio changed slightly in Census 1996 by falling to 92.2. The sex ratio at birth was slightly higher than that for Census 1996, as it was pegged at exactly 100. However, Census 2011 seems to have age group counts that produced sex ratios that were better than expected patterns and trends. For instance, the sex ratio at birth was high, at almost 102. The sex ratio for all subsequent age groups between 5 and 34 years remained high, except for age group 15-19. From age group 35-39 age ratios then consistently fell rapidly with increasing age as in the other two censuses. For example, the sex ratio for age group 75-79 was almost at 50. The overall sex ratio for Census 2011 was 95.5. [Table 5.6]

Table 5.6 PES adjusted census counts' sex ratios

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

5.8 Concluding remarks

The chapter's main findings from comparison of adjusted counts relative to respective counts from Agincourt, and also those from growth rate analyses, confirmed less coverage error in the respective censuses. Whereas, findings from age ratio analyses, sex ratio analyses, and those from graphical cohort analyses suggested the presence of a substantial amount of coverage error in the censuses.

However, it is possible that the latter findings were affected by certain distortive factors other than inaccurate counts. Distortive factors like migration may have disrupted the expected trends and patterns that are associated with census counts that have less or no coverage errors. Therefore, evidence gathered from the chapter's finding largely suggests less coverage errors in the adjusted counts from the three censuses. For adjusted census counts for area covered by Agincourt HDSS reasonably matched respective counts from Agincourt HDSS data. Secondly, because the growth rates for both intercensal periods 1996-2001 and 2001-2011 largely fell within the expected range. Thirdly, even in the case of those findings suggesting inaccuracy of investigated census counts, distortive factors of migration and inconsistent mortality and fertility seem to have accounted for suggested inaccuracy of counts.

CHAPTER 6: TO ADJUST OR NOT TO ADJUST?

6.1 Introduction

This chapter investigated which census counts between adjusted and unadjusted were better estimates of South Africa's actual population. This was achieved by comparing the two census counts against respective counts from respective counts from non-census data. The comparisons were at both national and SAs level. At the former level, adjusted and unadjusted census counts from the 10% samples were compared against respective counts from constructed population estimates obtained using MORTPAK's PROJCT programme. At the SAs level, adjusted and unadjusted census counts for the area covered by Agincourt HDSS that were extracted from SUPERCROSS were compared against respective counts from Agincourt HDSS data.

The latter counts were used as a gold standard. This was because of the following reasons. Firstly, Agincourt HDSS data is longitudinally collected, is regularly updated and there is constant correction of errors. Secondly, the data is collected at small areas level, which makes it less susceptible to coverage errors relative to data collected at a large scale like national censuses.

The findings from this chapter's investigations are vital in addressing the study's main goal of reducing controversies around the accuracy of the PES in determining and adjusting the undercount in South Africa's last three censuses. The accuracy of the PES was tested using the following rationale and strategies. Firstly, unadjusted counts which carry the undercount error are expected to have wider deviations from the gold standard's respective counts. On the other hand, respective adjusted counts are expected to closely match compared counts from Agincourt HDSS since the undercount has been corrected. If the PES was accurate, this should hold.

Findings from these analyses reduce controversies around the PES as they provide empirical evidence of how accurate the PES has been in determining and adjusting the undercount.

6.2. Comparison of counts at national level

6.2.1 Comparisons for 2001

Adjusted counts for males in Census 2001 were consistently very close to the projected counts. The only wider differences between adjusted and projected counts were noted among the earliest age groups. However, from age group 20-24 years, the differences between the compared counts became small, often below 4%. On the other hand, unadjusted counts were wide apart from projected counts. For instance, while the overall percentage difference between adjusted and projected counts was only 1.75%, it was 17.4% in unadjusted and projected counts. However, it was indicated in age groups 5-9, 10-14 and 80 years and above, that unadjusted counts were closer to projected counts than adjusted counts. [Table 6.1]

Table 6.1 Projected and census counts, males 2001

Age group	Projected (Proj) 2001	Adjusted (Adj) 2001	Unadjusted (Unadj) 2001	% Diff (Proj-Adj)	% Diff (Proj-Unadj)
0-4	2576368	2214369	1837926.27	14.1	28.7
5-9	2121494	2423906	2026385.416	-14.3	4.48
10-14	2300693	2510361	2103682.518	-9.11	8.57
15-19	2276139	2454284	2044418.572	-7.89	10.2
20-24	2011466	2100064	1631749.728	-4.4	18.9
25-29	1880342	1893200	1471016.4	-0.68	21.8
30-34	1642258	1596760	1253456.6	2.77	23.7
35-39	1435825	1438418	1129158.13	-0.18	21.3
40-44	1252592	1230423.1	965882.1335	1.77	22.9
45-49	993390	962657.87	787454.1377	3.09	20.7
50-54	770009	770704.03	630435.8965	-0.09	18.1
55-59	557036	551102.11	450801.526	1.07	19.1
60-64	431450	447549.3	366095.3274	-3.73	15.1
65-69	297386	305168.98	256647.1122	-2.61	13.7
70-74	238491	230192.45	193591.8505	3.48	18.8
75-79	134067	136967.29	115189.4909	-2.16	14.1
80+	118299	139578.25	117385.3083	-18	0.77

Total	21037305	21405705	17381276.42	-1.75	17.4

Overall, the same trends were noticed in the 2001 male's comparisons remained evident in the females' comparisons. For instance, the adjusted counts were very close to projected counts in comparison to unadjusted counts in all age groups, except 5-9, 10-14, and 80 years and above. The closeness of adjusted counts to projected counts is also evidenced by the overall difference of only about 3%. On the other hand, the overall difference between unadjusted counts and projected counts was more than 14%. [Table 6.2]

Table 6.2 Projected and census counts, females 2001

Age group	Projected (Proj) 2001	Adjusted (Adj) 2001	Unadjusted (Unadj) 2001	% Diff (Proj-Adj)	% Diff (Proj-Unadj)
0-4	2661248	2215008	1840671.648	16.77	30.8
5-9	2152354	2425994	2032982.972	-12.7	5.55
10-14	2301008	2541811	2135121.24	-10.5	7.21
15-19	2333691	2527782	2108170.188	-8.32	9.66
20-24	2109469	2189344	1766800.608	-3.79	16.2
25-29	2039131	2034172	1641576.804	0.243	19.5
30-34	1765191	1741231	1426068.189	1.357	19.2
35-39	1591068	1635554	1339518.726	-2.8	15.8
40-44	1346155	1376879	1127663.901	-2.28	16.2
45-49	1074626	1125861	949100.823	-4.77	11.7
50-54	831704	870990.936	734245.359	-4.72	11.7
55-59	630896	650859.782	548674.7962	-3.16	13
60-64	541570	622622.77	524870.9951	-15	3.08
65-69	473983	483069.23	415439.5378	-1.92	12.4
70-74	372964	396651.2	341120.032	-6.35	8.54
75-79	207343	231978.27	199501.3122	-11.9	3.78
80+	213367	293171.17	252127.2062	-37.4	-18
Total	22645766	23362979.1	19383654.34	-3.17	14.4

6.2.2 Comparisons for 2011

The difference between 2011's adjusted and projected counts for males remained consistently very slight across comparative age groups, whereas that between unadjusted and projected counts was wide. As in the 2001 comparisons, unadjusted counts of age group 80 years and above were closer to projected counts than adjusted counts. Other age groups with similar patterns were 10-14, and 20-24 years. The overall difference between adjusted counts and projected counts was less than 1%, yet unadjusted counts differed by about 14%. [Table 6.3 here]

Table 6.3 Projected and census counts, males 2011

Age group	Projected (Proj) 2011	Adjusted (Adj) 2011	Unadjusted (Unadj) 2011	% Diff (Proj-Adj)	% Diff (Proj-Unadj)
0-4	2703556	2867584.9	2431711.995	-6.07	10.1
5-9	2423016	2425181	2151135.547	-0.09	11.2
10-14	2074582	2344275	2091093.3	-13	-0.8
15-19	2355025	2498572	2178754.784	-6.1	7.48
20-24	2473008	2694646	2349731.312	-8.96	4.98
25-29	2462557	2542681.7	2217218.442	-3.25	9.96
30-34	2124380	2036206	1643218.242	4.15	22.6
35-39	1877162	1709346.5	1379442.626	8.94	26.5
40-44	1544284	1402328	1131678.696	9.19	26.7
45-49	1351880	1195740	1017574.74	11.6	24.7
50-54	1121600	1011349	860657.999	9.83	23.3
55-59	842490	811949.96	690969.416	3.63	18
60-64	633820	612363.96	521121.73	3.38	17.8
65-69	412380	401548.2	356574.8016	2.62	13.5
70-74	289910	297144.509	263864.324	-2.5	8.98
75-79	159747	163690.73	145357.3682	-2.47	9.01
80+	133704	174182.94	154674.4507	-30.3	-16
Total	24983200	25188790.9	21584779.77	-0.82	13.6

There was also not much difference in terms of both trends and patterns observed in the 2011 comparison of males' counts and those observed from comparison of females' counts. Adjusted counts for most age groups were closer to projected counts than the close counts observed in the female count comparisons for year 2001. Hence the overall difference between adjusted and projected counts was almost 0%. On the other hand, though unadjusted counts were further away from projected compared to respective adjusted counts, as from previous analysis, the discrepancy had reduced. The difference between adjusted and projected counts, and also between unadjusted and projected counts, was excessively wide, as both were above 150%. From all the comparison above, the total projected counts were lower than both adjusted and unadjusted counts. The total adjusted counts would be very slightly above projected counts, yet unadjusted total counts would be above projected counts by wider margins. [Table 6.4]

Table 6.4 Projected and census counts, females 2011

Age group	Projected (Proj) 2011	Adjusted (Adj) 2011	Unadjusted (Unadj) 2011	% Diff (Proj-Adj)	% Diff (Proj-Unadj)
0-4	2786569	2817867	2398004.817	-1.12	13.9
5-9	2523789	2394570	2119194.45	5.12	16
10-14	2096320	2250611	1996291.957	-7.36	4.77
15-19	2365610	2504905	2189286.97	-5.89	7.45
20-24	2506685	2679896	2342229.104	-6.91	6.56
25-29	2517076	2516635	2199538.99	0.02	12.6
30-34	2176850	1992804	1703847.42	8.46	21.7
35-39	1984577	1758420	1503449.1	11.4	24.2
40-44	1669206	1546291	1322078.805	7.36	20.8
45-49	1541761	1424543	1274965.985	7.60	17.3
50-54	1274931	1206940	1080211.3	5.33	15.3
55-59	1014969	985458.39	881985.2591	2.91	13.1
60-64	751621	773404.01	692196.589	-2.9	7.91
65-69	520820	556256.456	505637.1185	-6.8	2.92
70-74	438134	453343.42	412089.1688	-3.47	5.94
75-79	275563	317675.03	288766.6023	-15.3	-4.8
80+	137093	402149.69	365554.0682	-193	-167
Total	26681574	26581769.3	23275327.7	0.374	12.8

6.3 Comparison of counts at small areas level

6.2.1 Comparison by age groups for males 1996

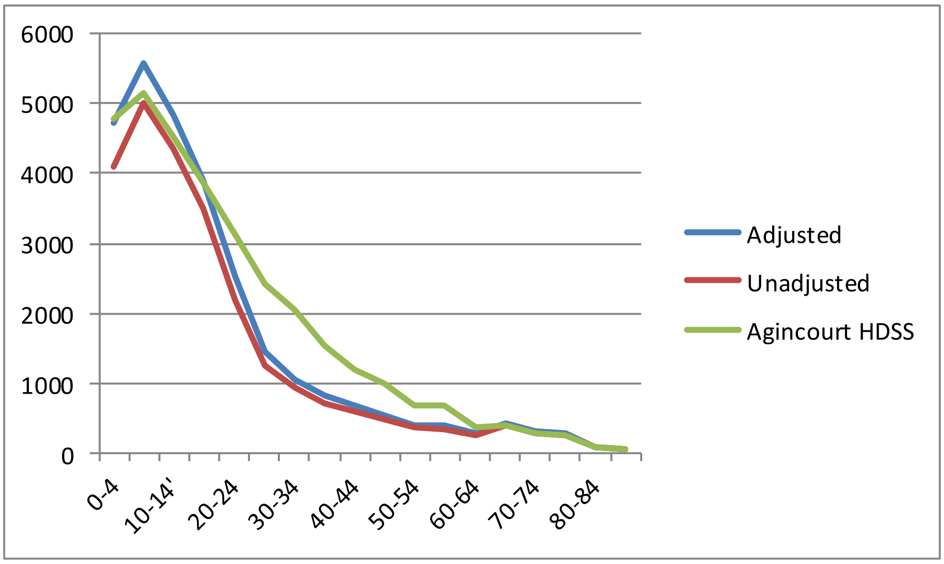
Adjusted counts were generally closer to Agincourt HDSS counts than the unadjusted counts. It was only among five age groups from a total of eighteen five-year age groups where unadjusted counts were closer to Agincourt HDSS counts than adjusted counts. Though adjusted counts were often closer to Agincourt HDSS counts than unadjusted counts, some discrepancies between adjusted and Agincourt HDSS counts were quite big. Such notable discrepancies existed in middle-aged groups. In the old-aged groups it was indicated that both adjusted and unadjusted counts were closer to Agincourt HDSS counts. [Table 6.5]

Table 6.5 Males' counts for 1996

Age grp	Agin HDSS 1996	Census Adjusted 1996	Census Unadjusted 1996	Differences	
				HDSS-Adj	HDSS-Unadj
0-4	4785	4724	4100	61	685
5-9	5148	5592	4999	-444	149
10-14	4513	4827	4359	-314	154
15-19	3880	3889	3492	-9	388
20-24	3136	2536	2201	600	935
25-29	2417	1453	1261	964	1156
30-34	2050	1053	930	997	1120
35-39	1526	809	714	717	812
40-44	1203	677	598	526	605
45-49	992	542	492	450	500
50-54	669	395	358	274	310
55-59	669	379	344	290	325
60-64	372	291	264	81	108
65-69	400	414	384	-14	16
70-74	272	295	274	-23	-2
75-79	239	269	250	-30	-11
80-84	82	86	80	-4	2
85+	62	59	55	3	7
Unclassified	383	414			
Missing	7				
Total	32420	28424	25155	3996	7265

Generally, both adjusted and unadjusted counts did not produce trends that were consistent with Agincourt HDSS counts across the majority of age groups. In fact, adjusted and unadjusted counts produced trends that were more identical to each other than to trends from Agincourt HDSS counts. They were both particularly far off from Agincourt HDSS counts in age group 20-24 to age group 60-64 years. It was only among the old-age groups, i.e. from 65-years and above, where both counts closely estimated Agincourt HDSS counts. [Fig 6.1]

Fig 6.1 Distributions of males counts by age group, 1996



6.2.2 Comparison of counts by age: females 1996

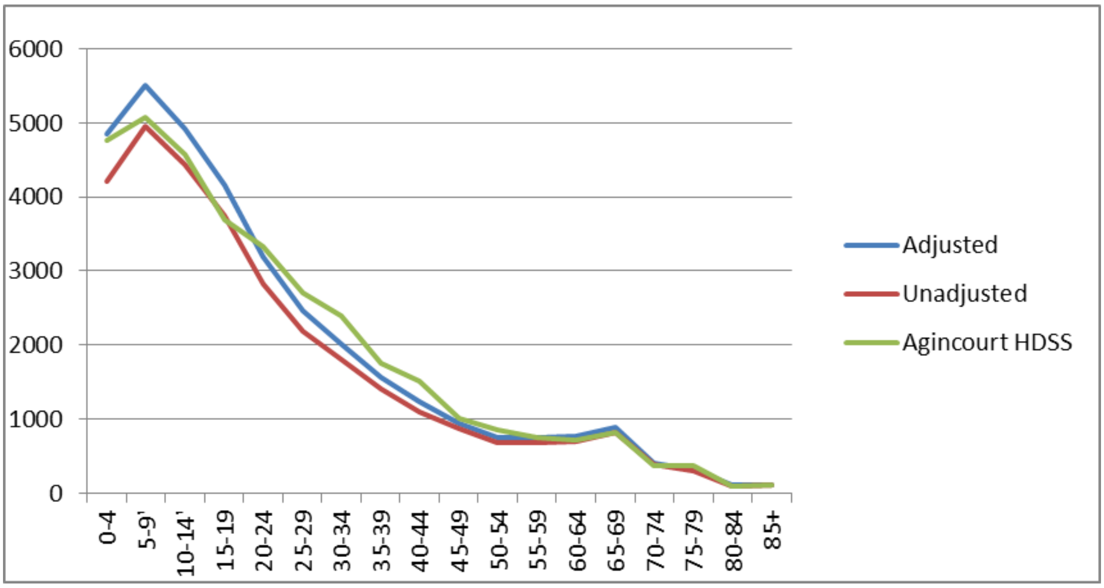
Comparisons of female counts largely followed the same trends and patterns as those picked up in males. However, the discrepancies between adjusted counts and Agincourt HDSS counts for females were much narrower than those observed in males. This meant adjusted counts for females were closer to those by Agincourt HDSS than the respective males. [Table 6.6]

Table 6.6 Comparison of female counts for 1996

Age grp	Agin HDSS 1996	Census Adjusted 1996	Census Unadjusted 1996	Differences	
				HDSS-Adj	HDSS-Unadj
0-4	4757	4843	4208	-86	549
5-9	5074	5513	4951	-439	123
10-14	4573	4920	4438	-347	135
15-19	3686	4164	3739	-478	-53
20-24	3333	3193	2819	140	514
25-29	2709	2469	2180	240	529
30-34	2388	2011	1808	377	580
35-39	1761	1560	1402	201	359
40-44	1518	1229	1105	289	413
45-49	1017	941	869	76	148
50-54	852	744	687	108	165
55-59	751	747	689	4	62
60-64	716	765	706	-49	10
65-69	828	882	825	-54	3
70-74	379	410	383	-31	-4
75-79	368	319	298	49	70
80-84	98	107	100	-9	-2
85+	104	115	107	-11	-3
Unclassified	570	541			
Missing	4				
Total	34916	34932	31314	-16	3595

The trends depicted in the graph below indicated that from age 40, adjusted census counts, closely matched respective counts from Agincourt HDSS. This was evidenced by consistent and closely identical trends from the respective counts drawn from the two data sets. However, there were also certain instances, e.g. among young age groups, where trends from unadjusted counts closely matched those by Agincourt HDSS better than the trends in the adjusted counts. [Fig 6.2]

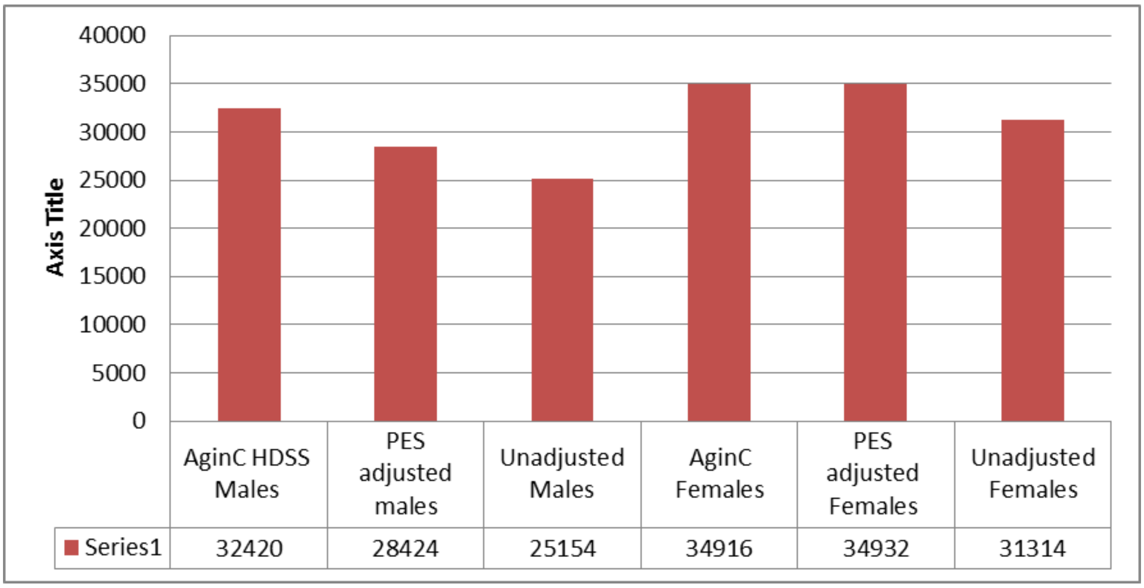
Fig 6.2 Distributions of females counts by age group, 1996



6.2.3 Comparisons of total counts

The totals of both adjusted and unadjusted counts in males were clearly different from respective totals obtained from Agincourt HDSS data. It was expected that after GIS mapping of counts, the adjusted counts would be the same as Agincourt counts. However, between adjusted and unadjusted total counts, the former were closer to respective counts from Agincourt HDSS data compared to unadjusted counts. As for females, the total in adjusted counts was virtually the same as that in Agincourt HDSS counts. This is 34 932 and 34 916 respectively, yielding a difference of only 16 counts in such a large population. The difference between the totals of the unadjusted counts and Agincourt HDSS counts was more than 3 000. [Fig 6.3]

Fig 6.3 Comparison of total counts by gender, 1996



6.2.4 Comparison of counts by age: males 2001

Adjusted counts for males across all age groups in Agincourt HDSS data were consistently lower than respective counts in census data, as was expected. However, the differences were not consistent across age groups, as they were generally wide in early age groups and narrowing with increase in age. Adjusted counts were also consistently closer to Agincourt HDSS counts than unadjusted counts, except for four age groups: 35-39, 55-59, 70-74, and 80-84 years. [Table 6.7]

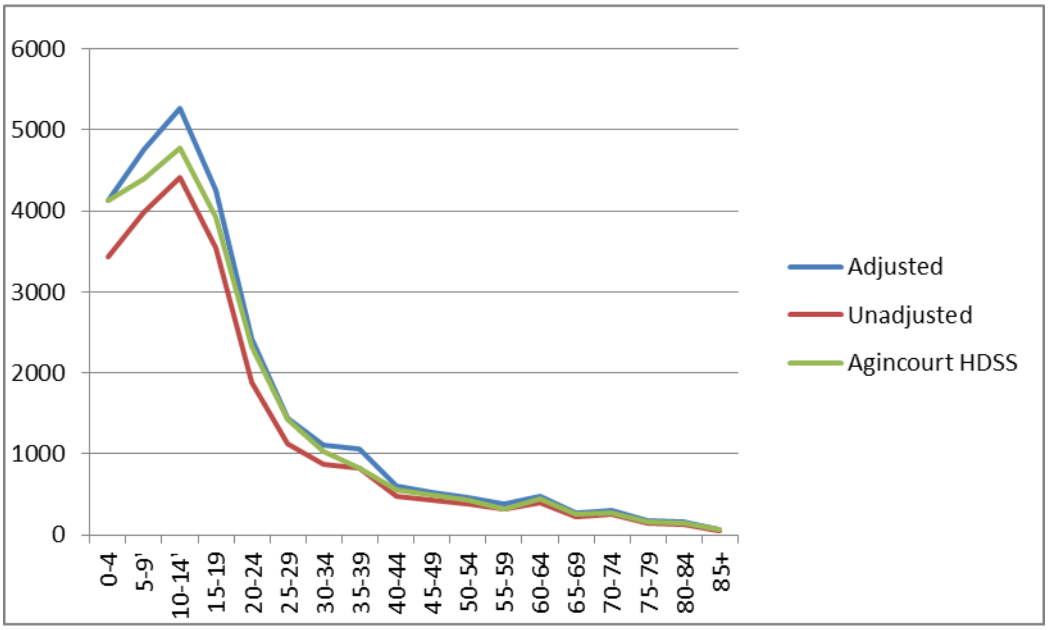
Table 6.7 Male counts for 2001

Age grp	Agin HDSS	Agin Adj	Agin Unadj	Differences	
				HDSS-Adj	HDSS-Unadj
0-4	4125	4135	3432.05	10	-692.95
5-9	4404	4759	3978.524	355	-425.476
10-14	4777	5267	4413.746	490	-363.254
15-19	3926	4248	3538.584	322	-387.416
20-24	2320	2417	1878.009	97	-441.991
25-29	1432	1444	1121.988	12	-310.012
30-34	1026	1111	872.135	85	-153.865
35-39	823	1057	829.745	234	6.745
40-44	557	608	477.28	51	-79.72
45-49	487	531	434.358	44	-52.642

50-54	432	470	384.46	38	-47.54
55-59	314	384	314.112	70	0
60-64	448	485	396.73	37	-51.27
65-69	251	272	228.752	21	-22.248
70-74	275	307	258.187	32	-16.813
75-79	166	178	149.698	12	-16.302
80-84	145	162	136.242	17	-8.758
85+	73	62	52.142	-11	-20.858
				0	0
Unspecified	384	402		18	-384
Missing	3			-3	-3
Total	25984	27897	22896.74	1913	-3087.258

Trends in census counts across age groups, i.e. both adjusted and unadjusted, were highly consistent with trends in Agincourt HDSS counts. Furthermore, trends in adjusted counts largely duplicated those in Agincourt HDSS, an indicator of how close the counts of the two sources were. The trend of counts in unadjusted data, though consistent with those in adjusted and Agincourt HDSS, was however clearly further away, suggesting a wider discrepancy between unadjusted counts and those in adjusted and Agincourt HDSS data. However, in old-age groups trends in three sources of compared counts were duplicated, suggesting that the old-aged group counts of adjusted, unadjusted, and Agincourt HDSS were largely the same. [Fig 6.4]

Fig 6.4 Distribution of male counts by age group, 2001



6.3.5 Comparison of counts by age group for females 2001

As in the comparisons of the 2001 males' counts, the females' adjusted counts were very close to Agincourt HDSS counts compared to unadjusted counts. As expected from the outcome of GIS mapping of counts, adjusted census counts were slightly higher than respective counts of Agincourt HDSS. [Table 6.8]

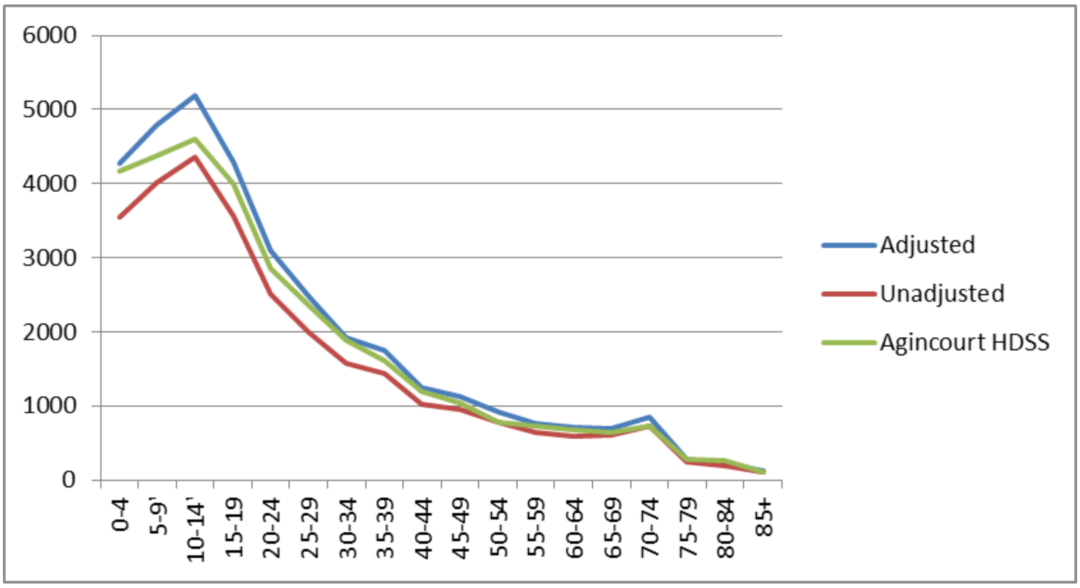
Table 6.8 Comparison of females counts, 2001

Age grp	Agin HDSS	Agin Adj	Agin Unadj	Differences	
				HDSS-Adj	HDSS-Unadj
0-4	4171	4265	3544.215	94	-626.785
5-9	4383	4790	4014.02	407	-368.98
10-14	4601	5191	4360.44	590	-240.56
15-19	3996	4282	3571.188	286	-424.812
20-24	2858	3104	2504.928	246	-353.072
25-29	2351	2466	1990.062	115	-360.938
30-34	1888	1928	1579.032	40	-308.968
35-39	1614	1744	1428.336	130	-185.664
40-44	1195	1251	1024.569	56	-170.431
45-49	1037	1128	950.904	91	-86.096
50-54	787	914	770.502	127	-16.498
55-59	720	763	643.209	43	-76.791

60-64	680	706	595.158	26	-84.842
65-69	638	698	600.28	60	-37.72
70-74	724	852	732.72	128	8.72
75-79	277	283	243.38	6	-33.62
80-84	252	218	187.48	-34	-64.52
85+	112	118	101.48	6	-10.52
				0	0
Unspecified	641	619		-22	-641
Missing	1			-1	-1
Total	32284	34701	28841.9	2417	-3442.097

There was a highly consistent and close match between adjusted and Agincourt HDSS counts across almost all age groups. The only clear differences noted are for age groups 0-4 and 5-9 years. The trend in unadjusted counts also remained consistent with those in adjusted and Agincourt HDSS counts, though remaining clearly far from the two, particularly at young and middle age groups. [Fig 6.5]

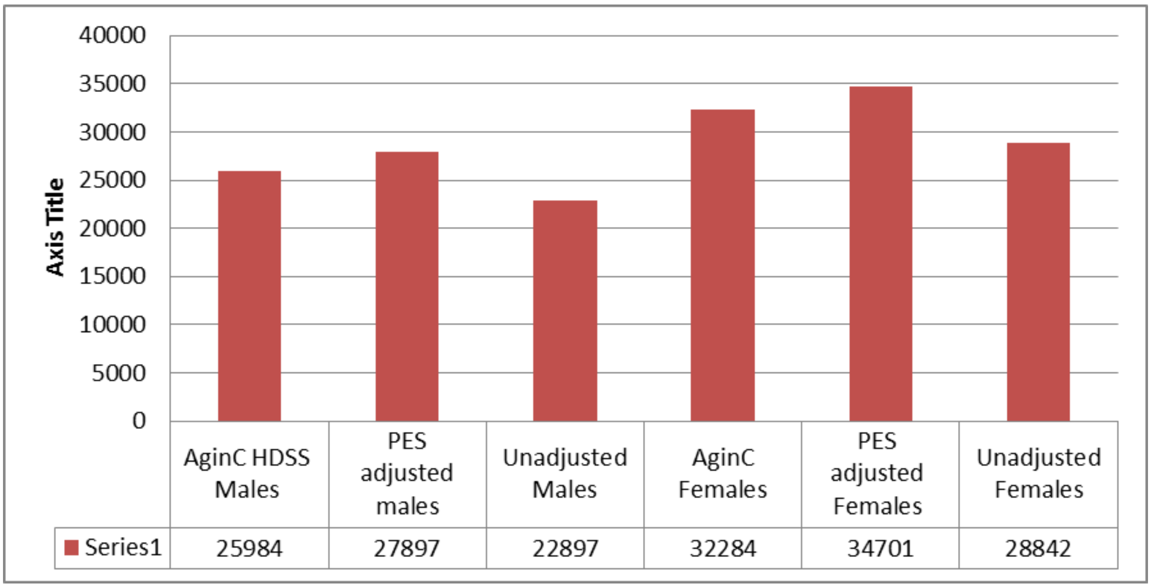
Fig 6.5 Distribution of female counts by age group, 2001



6.3.6 Comparisons of total counts

The total counts for males and for females in Agincourt HDSS data were both less than respective counts in the census. Also, the unadjusted total counts for either sex were far lower than respective counts in Agincourt HDSS data compared to totals in adjusted counts. [Fig 6.6]

Fig 6.6 Comparison of total counts by gender, 2001



6.3.7 Comparison of counts by age: males 2011

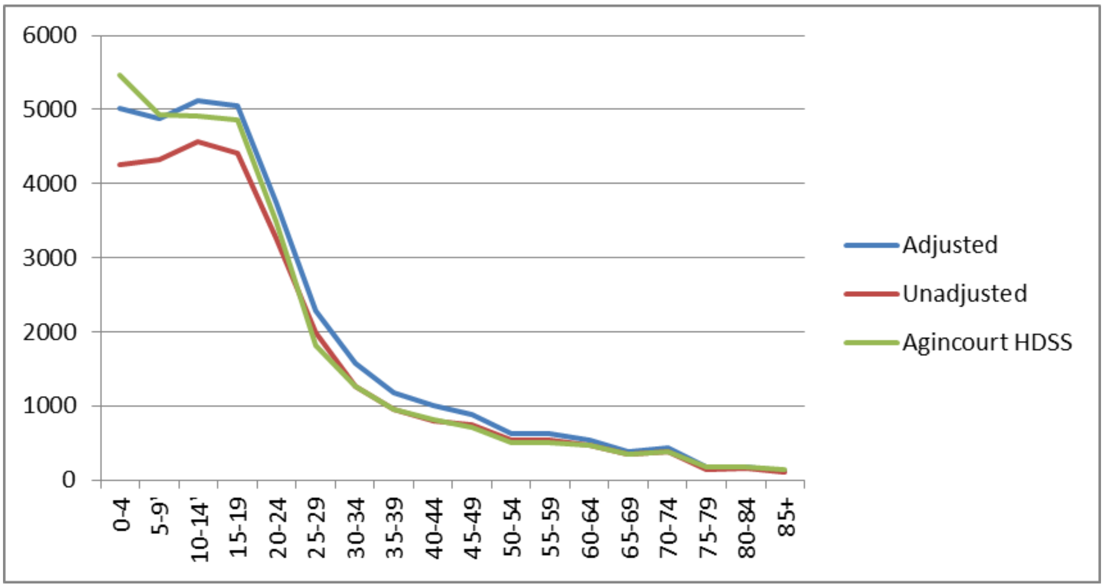
The adjusted census counts of males were consistently higher than Agincourt HDSS counts in most of the age groups. Such a finding was consistent with what was expected. However, the differences between compared counts from the two sources were higher than anticipated as these were generally higher than those observed in the 2001 comparisons. Yet the matching of small areas and village boundaries in 2011 was closer than for either 1996 or 2001. In the 2011 males' comparisons, which was inconsistent with findings in similar analyses for years 1996 and 2001, it was found that the unadjusted counts were generally closer to Agincourt HDSS counts compared to adjusted counts. Out of the 18 age groups whose counts were compared, 12 from the unadjusted data had counts that were closer to Agincourt HDSS compared to respective counts in adjusted data. [Table 6.9]

Table 6.9 Comparison of male counts, 2011

Age grp	Agin HDSS	Agin Adj	Agin Unadj	Differences	
				HDSS-Adj	HDSS-Unadj
0-4	5466	5014	4251.872	-452	-1214.128
5-9	4929	4880	4328.56	-49	-600.44
10-14	4907	5123	4569.716	216	-337.284
15-19	4855	5058	4410.576	203	-444.424
20-24	3501	3754	3273.488	253	-227.512
25-29	1811	2275	1983.8	464	172.8
30-34	1260	1568	1265.376	308	5.376
35-39	948	1171	944.997	223	-3.003
40-44	810	996	803.772	186	-6.228
45-49	706	877	746.327	171	40.327
50-54	507	623	530.173	116	23.173
55-59	504	631	536.981	127	32.981
60-64	471	541	460.391	70	-10.609
65-69	341	383	340.104	42	-0.896
70-74	384	426	378.288	42	-5.712
75-79	168	165	146.52	-3	-21.48
80-84	169	172	152.736	3	-16.264
85+	143	125	111	-18	-32
			0	0	0
Unspecified	480	462	0	-18	-480
Missing	4		0	-4	-4
Total	31868	33782	29234.68	1914	-2633.323

The trend drawn from distribution of counts by age from males' adjusted data matches one in Agincourt HDSS counts better than that in unadjusted counts only in age groups 0-4, 5-9, 10-14, 15-19, 20-24, and 25-29. The unadjusted count trend in the rest of the age groups remained constantly closer to that of Agincourt HDSS counts compared to one in adjusted counts. However, like in previous analyses, trends in adjusted, unadjusted, and Agincourt counts suggested almost the same counts in respective age groups among the old-aged group. [Fig 6.8]

Fig 6.8 Distribution of males counts by age group, 2011



6.3.8 Comparison of counts by age group: females 2011

The comparison of female's counts in 2011 followed the same patterns and trends as those observed in comparisons to males as unadjusted counts were much closer to Agincourt HDSS counts compared to adjusted counts. [Table 6.10]

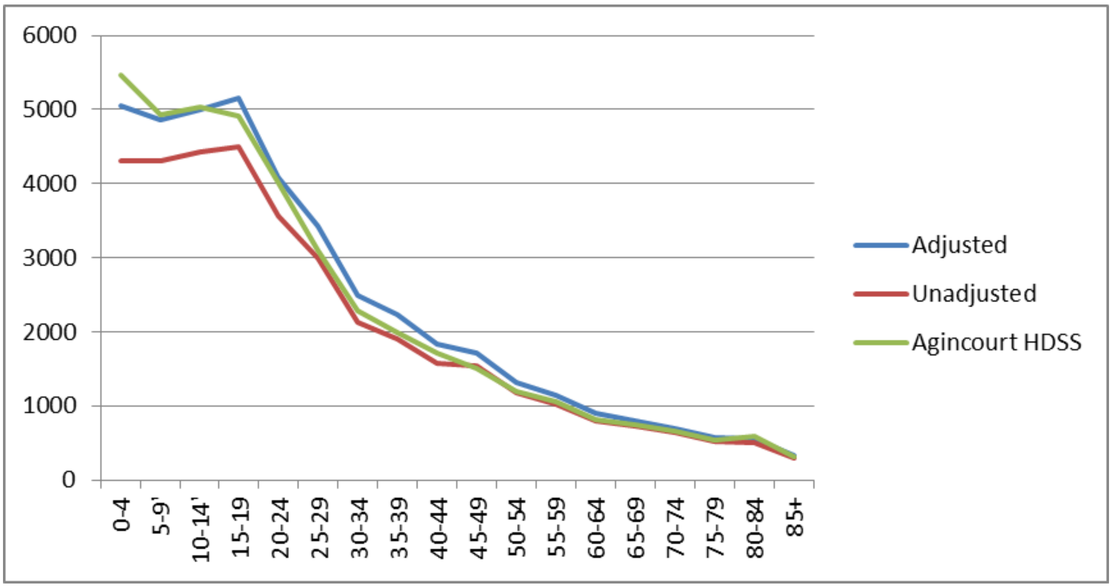
Table 6.10 Comparison of female counts, 2011

Age grp	Agin HDSS	Agin Adj	Agin Unadj	HDSS-Adj	HDSS-Unadj
0-4	5465	5052	4299.252	-413	-1165.748
5-9	4928	4864	4304.64	-64	-623.36
10-14	5037	4996	4431.452	-41	-605.548
15-19	4918	5147	4498.478	229	-419.522
20-24	4012	4075	3561.55	63	-450.45
25-29	3101	3430	2997.82	329	-103.18
30-34	2277	2492	2130.66	215	-146.34
35-39	1996	2223	1900.665	227	-95.335
40-44	1718	1842	1574.91	124	-143.09
45-49	1509	1719	1538.505	210	29.505
50-54	1202	1308	1170.66	106	-31.34
55-59	1053	1147	1026.565	94	-26.435
60-64	811	898	803.71	87	-7.29
65-69	744	794	721.746	50	-22.254
70-74	652	697	633.573	45	-18.427
75-79	528	569	517.221	41	-10.779
80-84	597	559	508.131	-38	-88.869

85+	310	331	300.879	21	-9.121
			0	0	0
Unspecified	1435	1459	0	24	-1435
Missing	4		0	-4	-4
Total	40878	42143	36920.42	1265	-3957.583

The trends from the distribution of female counts by age groups in 2011 comparisons indicated that trends in both adjusted and unadjusted counts were largely identical to Agincourt HDSS counts. Inconsistencies were only observed among the first three age groups. [Fig 6.9]

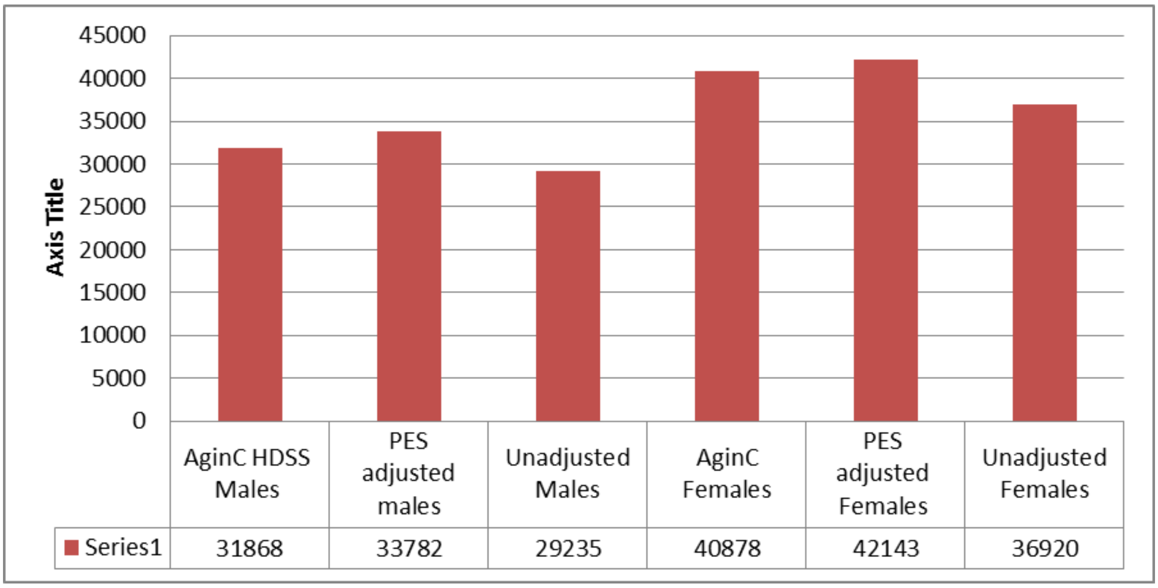
Fig 6.9 Distribution of female counts by age group, 2011



6.3.9 Comparison of total counts, 2011

Totals in adjusted counts of either males or females were higher than respective counts in Agincourt HDSS data. On the other hand, totals in counts of unadjusted data were less than those in Agincourt HDSS counts. The former counts were closer to respective counts in Agincourt HDSS than the latter. [Fig 6.10]

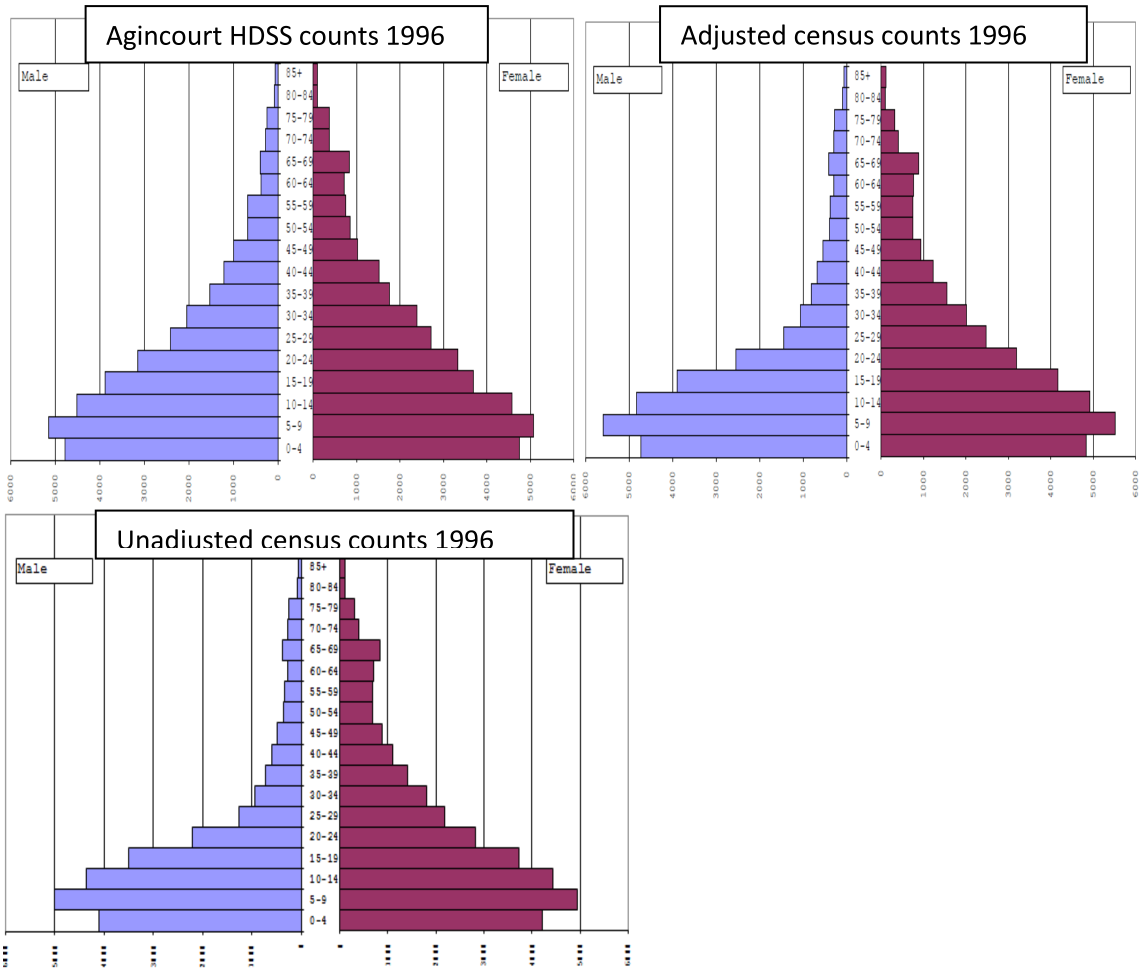
Fig 6.10 Comparison of total counts by gender, 2011



6.3.10 Age-sex structure comparisons

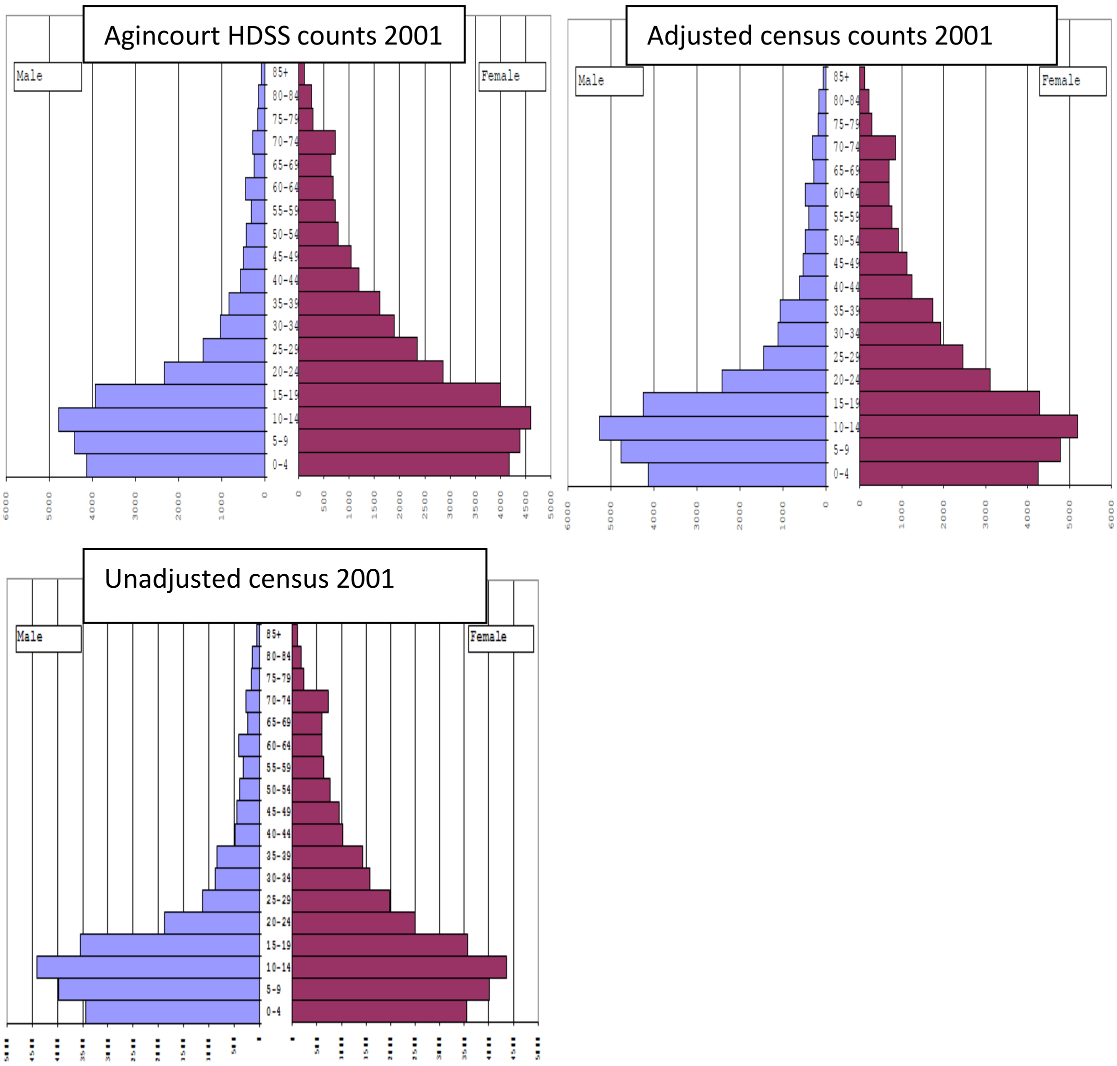
The age-sex pyramids for 1996 comparisons drawn from adjusted, unadjusted, and Agincourt HDSS counts generally resemble each other structurally. For instance, the following was suggested in all the counts: a decline in fertility, a bulge among female population aged 70-74 years, and a largely triangular-shaped population pyramid whose shape is mainly distorted by a base that suggests a decline in fertility. In particular, the most striking similarities are between the population pyramids in the adjusted and unadjusted census counts. A characteristic depicted in the population pyramids of the two census data sets, and not in the Agincourt HDSS population pyramid, is a reduced male population relative to that for females of 20-24 years and above. [Fig 6.11]

Fig 6.11 Age sex pyramids for area covered by Agincourt HDSS, 1996



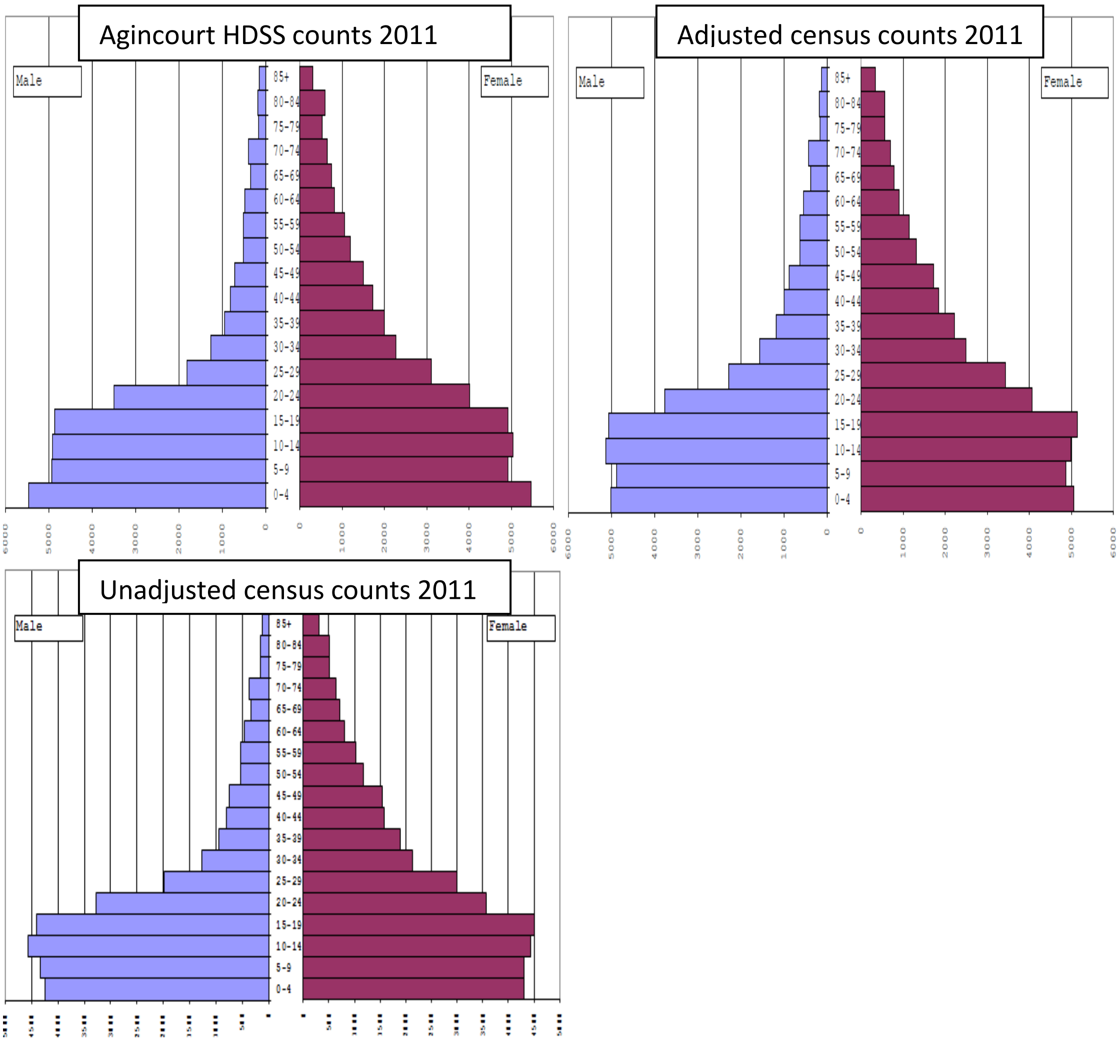
Again, structurally, the adjusted, unadjusted and Agincourt HDSS popuation pyramids of 2001 had striking similarities. They all suggested consistantly declining fertility in the past 15 years. Furthermore, each of the three pyramids suggested population for either males or females aged 75-79 years that was clearly higher than that in the immediately preceding and succeeding age groups. In the three cases, this population was traceable 10 years back to 1996. Between the adjusted and the unadjusted population pyramids, the former however seems to be closer to the Agincourt HDSS population pyramid than the latter. The population pyramid of adjusted data resembled that of Agincourt HDSS more, mainly because both counts were closer to each other than those for uandjusted. [Fig 6.12]

Fig 6.12 Age sex pyramids for area covered by Agincourt HDSS, 2001



The most striking feature in the 2011 comparison of population pyramids is between adjusted and Agincourt HDSS pyramids. Both pyramids indicated fertility upsurge, a characteristic that has been raised by some demographers to partly critique the accuracy of Census 2011 counts. On the other hand, the population pyramid of unadjusted counts contradicted this finding, by suggesting fertility decline, which is more pronounced among females. Furthermore, the population pyramid of unadjusted counts suggested declining fertility in the past 15 years for males, and 20 years for females. Yet on the other hand, both the adjusted and Agincourt HDSS population pyramids do not clearly suggest fertility declines in the same period. [Fig 6.13]

Fig 6.13 Age sex pyramids for area covered by Agincourt HDSS, 2011



6.4 Concluding remarks

The findings in this chapter are expected to have contributed in reducing controversies around the accuracy of the PES as they suggested credibility of the PES and its outcomes as explained below. Comparisons of census counts with respective none census counts at both national and SAs levels indicated that unadjusted census counts widely deviated from respective none census counts; indicating the effect of the undercount error in these counts as estimated by the PES. On the other hand adjusted census counts largely matched respective none census counts they were compared with, indicating the credibility of the PES in adjusting for the undercount error.

CHAPTER 7: PES RELATIVE TO DA UNDERCOUNT ESTIMATES

7.1: Introduction

This chapter investigated how censuses in South Africa should be handled in order to reduce associated controversies by directly comparing the PES undercount estimates against those of DA. The official undercount estimates for the country are those drawn from the former method, and these estimates have also been at the centre of controversy, as critiques argue that they are equally inaccurate. Findings in this investigation are vital in addressing the study's main objective as the similarity of compared undercount estimates drawn from the two methods would be a confirmation of the accuracy of the PES method. On the other hand, if the two methods produce contradicting undercount estimates, this would justify the criticism which has been raised against the PES that the method is largely inaccurate.

Alternative undercount estimates from DA were however not treated as a gold standard for measuring the accuracy of respective PES undercount estimates. This was because the study took cognisance of the fact that both methods were prone to unique sources of error. For this reason there was no justification for treating the DA's estimates as gold standard for evaluating the accuracy of respective PES estimates. Rather the test for accuracy of PES undercount estimates was based on how consistent the respective undercount estimates of the two methods were in relation to each other. DA undercount estimates were obtained using constructed population estimates of MORTPAK projections for the years 2001 and 2011 relative to respective counts of unadjusted data.

7.2 Constructed population estimates and census counts

The two tables below were incorporated to show the three sets of population counts that were derived from both the PES and DA methods. Adjusted and unadjusted census counts were products of the former method, while constructed population estimates were from the latter method. Details of the procedures followed in arriving at these counts were articulated in the sections on data management and analysis in Chapter 3. Each set of population counts is distributed according to age groups. The first table shows population counts for the year 2001 in both males and females, whereas the second table shows 2011 counts. Both tables show that unadjusted census counts were always lower than respective adjusted census and constructed counts. Respective counts from the latter two were generally close to each other [Tables 7.1 and 7.2]

Table 7.1 Constructed Population Estimates and census counts by age group for 2001

Age group	Males 2001			Females 2001		
	Constructed Population estimates	PES adjusted census counts	Unadjusted census counts	Constructed Population estimates	PES adjusted census counts	Unadjusted census counts
0-4	2576368	2214369	1837926	2661248	2215008	1840672
5-9	2121494	2423906	2026385	2152354	2425994	2032983
10-14	2300693	2510361	2103682	2301008	2541811	2135121
15-19	2276139	2454284	2044418	2333691	2527782	2108170
20-24	2011466	2100064	1631749	2109469	2189344	1766801
25-29	1880342	1893200	1471016	2039131	2034172	1641577
30-34	1642258	1596760	1253456	1765191	1741231	1426068
35-39	1435825	1438418	1129158	1591068	1635554	1339519
40-44	1252592	1230423	965882	1346155	1376879	1127664
45-49	993390	962657	787454	1074626	1125861	949101
50-54	770009	770704	630435	831704	870990	734246
55-59	557036	551102	450801	630896	650859	548675
60-64	431450	447549	366095	541570	622622	524871
65-69	297386	305168	256647	473983	483069	415440
70-74	238491	230192	193591	372964	396651	341120
75-79	134067	136967	115189	207343	231978	199501
80+	118299	139578	117385	213367	293171	252127

Table 7.2 Constructed Population Estimates and census counts by age group for 2011

Age group	Males 2001			Females 2001		
	Constructed Population estimates	PES adjusted census counts	Unadjusted census counts	Constructed Population estimates	PES adjusted census counts	Unadjusted census counts

0-4	2703556	2867584.9	2431712	2786569	2817867	2398005
5-9	2423016	2425181	2151136	2523789	2394570	2119194
10-14	2074582	2344275	2091093	2096320	2250611	1996292
15-19	2355025	2498572	2178755	2365610	2504905	2189287
20-24	2473008	2694646	2349731	2506685	2679896	2342229
25-29	2462557	2542681.7	2217218	2517076	2516635	2199539
30-34	2124380	2036206	1643218	2176850	1992804	1703847
35-39	1877162	1709346.5	1379443	1984577	1758420	1503449.
40-44	1544284	1402328	1131679	1669206	1546291	1322079
45-49	1351880	1195740	1017575	1541761	1424543	1274966
50-54	1121600	1011349	860658	1274931	1206940	1080211
55-59	842490	811950	690969	1014969	985459	881985
60-64	633820	612364	521121	751621	773404	692196
65-69	412380	401548	356575	520820	556257	505637
70-74	289910	297145	263864	438134	453343	412089
75-79	159747	163691	145358	275563	317675	288767
80+	133704	174183	154674	137093	402150	365554

7.3 Comparisons of total counts by gender

Comparisons of the total counts for the years 2001 and 2011 indicate that constructed population estimates and adjusted census counts for both males and females were almost the same. However, there is a clear difference between either of the two relative to unadjusted counts. For instance, the total counts for the 2001 comparisons indicate that counts of males drawn from either data were about 21 million, and about 25 million for females, again from either data. The counts of the same compared groups were again equally similar in 2011 comparisons. This is an indication that both the PES and DA undercount estimates for these subpopulation groups are likely to be close to each other as they are computed using the same respective unadjusted counts [Figs 7.1 and 7.2 here]

Fig 7.1 Totals for male counts

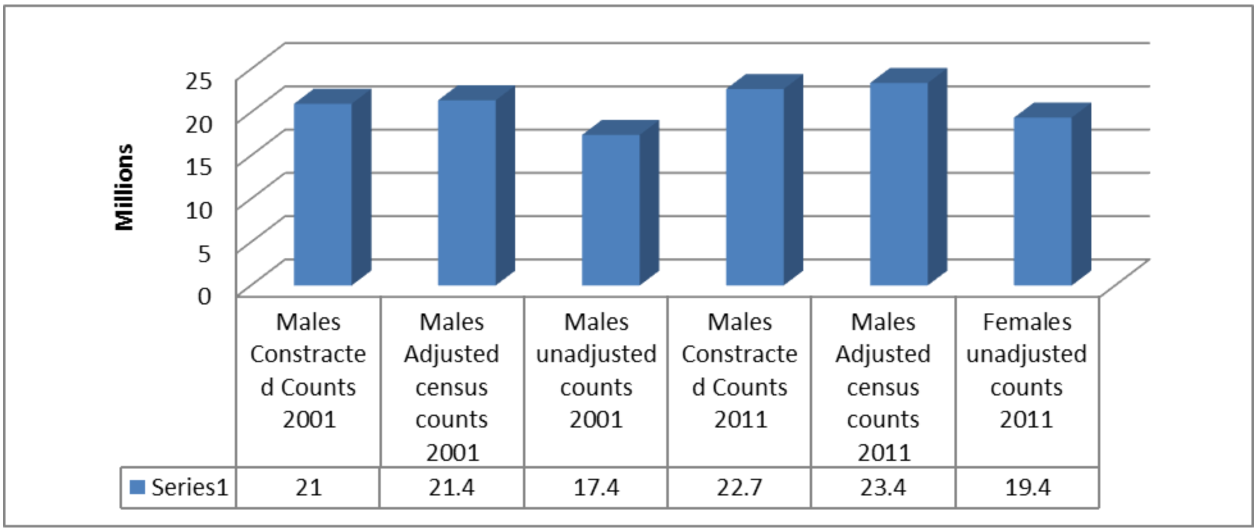
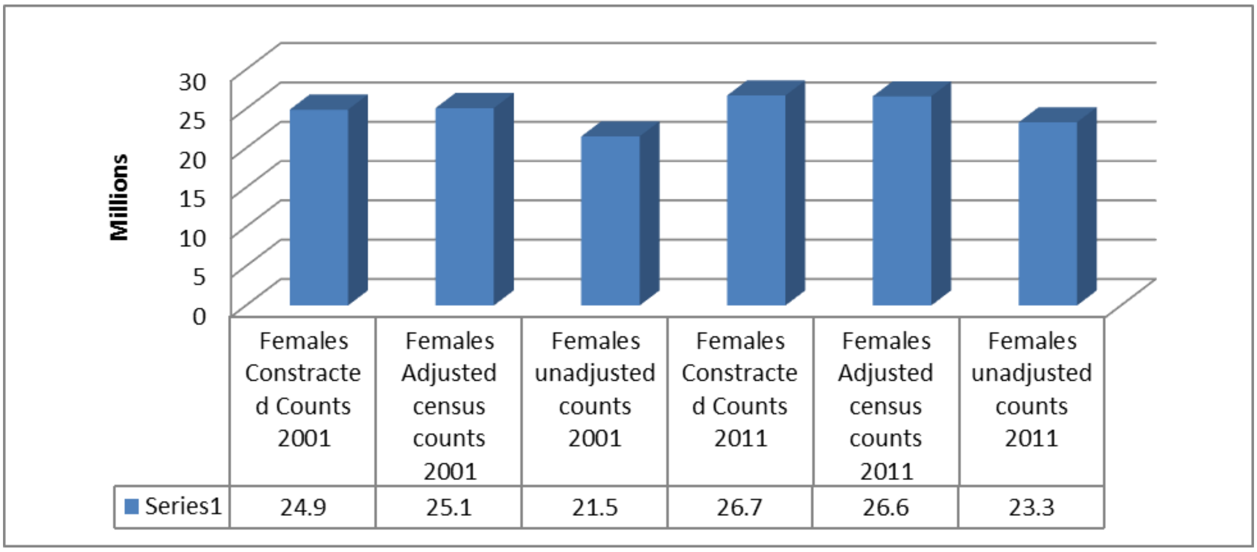


Fig 7.2 Totals for female counts



7.4 Comparison of undercount estimates for totals

The results in Table 7.3 are important in that they directly demonstrate the extent to which respective undercount estimates of the two methods approximate each other as the table summarised overall undercount estimates of males, females and the combined counts of both sexes at national level. A paltry 3.3% difference in undercount estimates of male counts in the 2011 comparison is the worse dissimilarity. In other comparisons undercount estimates from the two methods had percentage differences of as low as e.g. 0.6% for females in 2011, 1.2% males in 2001 and the same percentage difference for the combined counts in 2001. These results therefore indicated that respective undercount estimates from the PES and DA for total counts of either sex as well as for the combined sexes closely matched. Out of the six comparisons summarised in the table, four had percentage differences of around 1% or less. [Tables 7.3]

Table 7.3: PES and DA undercount estimates by gender

Population category	2001			2011		
	PES	DA	Difference	PES	DA	Difference
Males	18.6%	17.4%	1.2%	15.9%	13.6%	3.3%
Females	16.9%	14.4%	2.5%	13.4%	12.8%	0.6%
Overall	17%	15.8%	1.2%	14.7%	13.2%	1.4%

7.5 Comparisons of undercount estimates for age groups

The undercount estimates of the PES and DA for males in 2001 comparisons perfectly matched each other from 45 years onwards. The estimates also followed a consistent trend. In age groups between 20 and 29 years, DA estimates were slightly lower than respective estimates of the PES, and slightly higher than the latter estimates for age groups between 30 to 44 years. However, there was neither consistence of trends nor suggestions of the PES and DA undercount estimates being closer to each other in age groups between 0 and 19 years. The comparisons of the 2011 male undercount estimates were not as perfect as those of the 2001 comparisons. The PES undercount estimates in age groups 15-29 years and 65-69 were slightly lower than respective DA estimates, but the trends were consistent in comparison to each other. In age groups 30-59 years, the PES undercount estimates were however lower than those of DA, and trends remained consistent. There was no consistence of trends as well as no similarities between the PES and DA undercount estimates for age groups 0-4, 5-9, 15-19, 20-24 and 25-29 years.

The PES undercount estimates of females in 2001 comparisons were clearly higher than respective DA estimates in age groups 5-19 years and 70-79 years. The trends in the compared undercount estimates were also inconsistent in comparison. The PES undercount estimates in age group 0-4 were clearly lower than those of DA. The undercount estimates of the two methods in age groups 20-44 years were however similar. As for age groups 45-64 years, the estimates differed slightly, but the trends matched each other. As for 2011 female comparisons, the PES undercount estimates were higher than respective estimates of DA in middle-aged groups, whereas the latter's estimates were also higher than the former in young and older age groups [Fig 7.1 and 7.2]

Fig 7.1 PES and DA undercount estimates for males by age groups

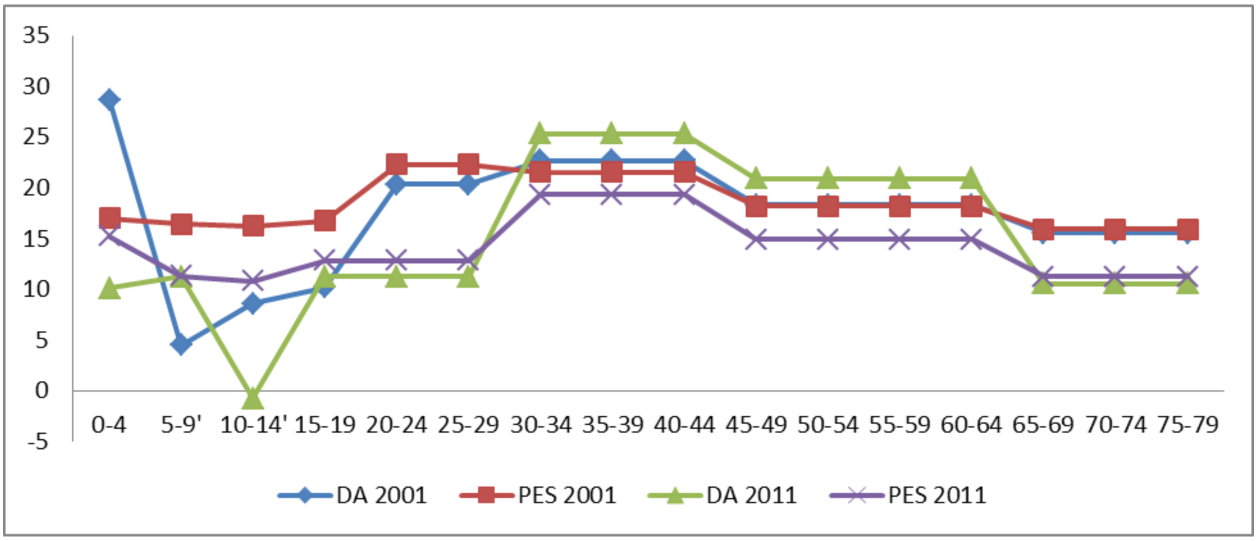
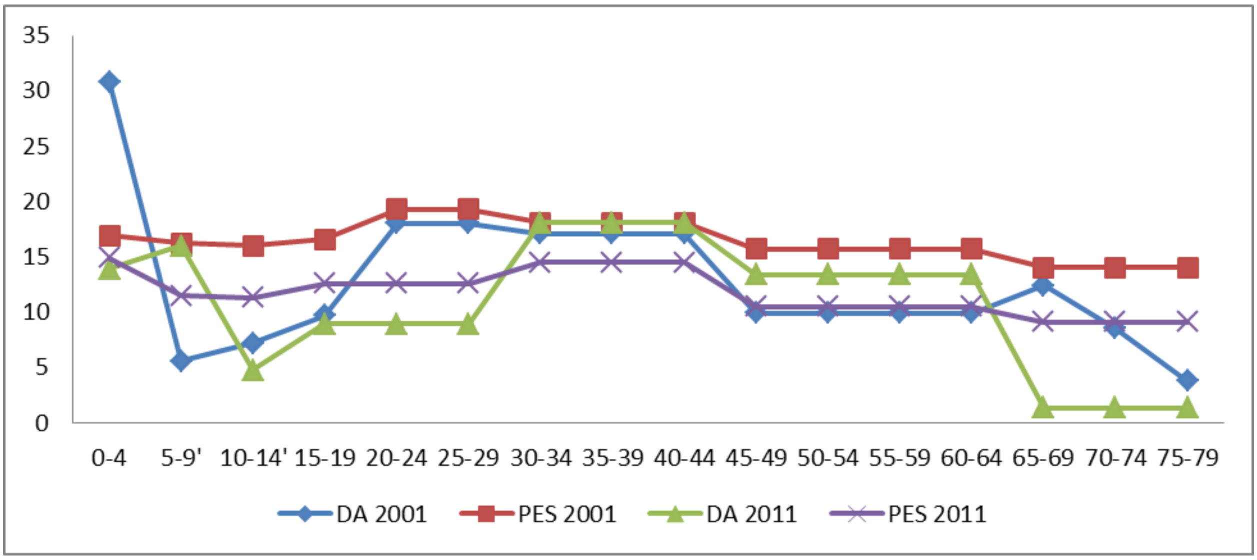


Fig 7.2 PES and DA undercount estimates for females by age groups



7.6 Concluding remarks.

The findings confirmed that respective undercount estimates from PES and DA, for males' total, females' total and for combined sexes were largely similar. The undercount estimates of age groups were not explicit in confirming this as there were some instances where estimates of the two methods indicated clear differences. However, most of the compared age group undercount estimates from the two methods remained largely consistent with each other. Few were even similar. Hence the overall findings suggest that respective undercount estimates from the two methods closely matched. The PES may not be out rightly accurate in estimating and adjusting for the undercount as suggested by some of its undercount estimates which did not match with those from DA. However, the fact that majority of the undercount estimates obtained from PES closely matched respective estimates from DA further confirms credibility of PES in estimating and adjusting for the undercount.

CHAPTER 8: UNDERCOUNT EFFECT ON DEMOGRAPHIC PARAMETERS

8.1 Introduction

This chapter investigated the effects of undercount on demographic parameters. The analysed parameters included crude birth rates (CBR), total fertility rate (TFR), parity progression ratio (PPR), age-specific fertility rates (ASFR), and general fertility rates (GFR). Demographic parameters of adjusted counts were compared with those drawn from respective unadjusted counts. The comparisons were at national level. The unadjusted counts used were those that were reconstructed from adjusted counts drawn from the 10% samples data.

Demographic parameters are used to inform efficiency planning. Inaccurate parameters can therefore compromise effective planning. The purpose of this investigation was to determine if undercount leads to inaccurate demographic parameters. This was achieved by comparing demographic parameters of adjusted and unadjusted counts. The latter counts were not corrected for undercount and hence produced demographic parameters that had the effect of undercount error. The former were corrected for the undercount and hence their demographic parameters were not affected by the undercount error. The difference between demographic parameters in respective adjusted and unadjusted counts quantified the effect of undercounting of these parameters.

8.2 Adjusted and unadjusted counts by age and sex

The proportion of population enumerated for respective male and female age groups were almost the same, suggesting no major differences in undercount estimates by sex. It was indicated that older age groups, specifically those aged 65 years and above, had better enumeration coverage, relative to lower age groups in both sexes. In particular, the age group 0-4 years had the least enumeration coverage. [Table 8.1]

Table 8.1 Adjusted and unadjusted counts, 1996

Age group	PES Adjusted Counts		Proportion of Population enumerated		Unadjusted counts	
	Males	Females	Males	Females	Males	Females
0-4	2215745	2223343	0.868	0.869	1923064.03	1931824.77
09-May	2340363	2332033	0.894	0.898	2092180.73	2094303.53
14-Oct	2309587	2359138	0.903	0.902	2085661.85	2128019.2
15-19	2048208	2135672	0.898	0.898	1839121.2	1917693.95
20-24	1914992	2067653	0.868	0.883	1662025.79	1825466.27
25-29	1661866	1790412	0.868	0.883	1442266.36	1580841.7
30-34	1452053	1617576	0.883	0.899	1282439.97	1454290.11
35-39	1278644	1375399	0.883	0.899	1128494.21	1236095.77
40-44	1026535	1105325	0.883	0.899	906747.312	993803.194
45-49	809067.79	863268.1	0.908	0.923	734430.54	797215.707
50-54	599708.91	665172.11	0.906	0.923	543533.645	614211.342
55-59	482331.66	586841.05	0.908	0.922	437789.158	541276.414
60-64	351752.42	537440.52	0.907	0.923	319115.598	495990.856
65-69	307073.32	455615.11	0.928	0.935	284821.63	426170.597
70-74	195351.08	286534.54	0.929	0.934	181444.732	267665.194
75-79	141217.88	238685.05	0.929	0.934	131243.383	222972.236
80+	104999.16	210474.53	0.93	0.935	97673.6372	196705.168
Missing	252292	236529.873				
Total	19 491 787	21087112.3			17344345.8	18961075.9

There was no big difference between reciprocal adjustment factors in males and females across age groups in Census 2001, except to some extent in age groups 20-29 years in males. Due to the high undercount in the 2001 Census, there are substantial differences between adjusted and 'unadjusted' counts. The counts of male age groups of 20-24 years and 25-29 years were most different when compared to their respective counts in the two data sets, whereas the smallest difference between respective adjusted and unadjusted counts were in age groups above 59 years. The difference between adjusted and unadjusted counts of females, as with males, was also higher due to high undercount. However, the percentage difference between adjusted and ‘unadjusted’ data remained almost the same, as the undercount did not vary much across age groups. [Table 8.2]

Table 8.2 Adjusted and unadjusted counts, 2001

Age group	PES Adjusted Counts		Proportion of Population enumerated		“Unadjusted” counts	
	Males	Females	Males	Females	Males	Females
0-4	2214369	2215008	0.83	0.831	1837926.27	1840671.648
5-9	2423906	2425994	0.836	0.838	2026385.416	2032982.972
10-14	2510361	2541811	0.838	0.84	2103682.518	2135121.24
15-19	2454284	2527782	0.833	0.834	2044418.572	2108170.188
20-24	2100064	2189344	0.777	0.807	1631749.728	1766800.608
25-29	1893200	2034172	0.777	0.807	1471016.4	1641576.804
30-34	1596760	1741231	0.785	0.819	1253456.6	1426068.189
35-39	1438418	1635554	0.785	0.819	1129158.13	1339518.726
40-44	1230423.1	1376879	0.785	0.819	965882.1335	1127663.901
45-49	962657.87	1125861	0.818	0.843	787454.1377	949100.823
50-54	770704.03	870990.936	0.818	0.843	630435.8965	734245.359
55-59	551102.11	650859.782	0.818	0.843	450801.526	548674.7962
60-64	447549.3	622622.77	0.818	0.843	366095.3274	524870.9951
65-69	305168.98	483069.23	0.841	0.86	256647.1122	415439.5378
70-74	230192.45	396651.2	0.841	0.86	193591.8505	341120.032
75-79	136967.29	231978.27	0.841	0.86	115189.4909	199501.3122
80+	139578.25	293171.17	0.841	0.86	117385.3083	252127.2062
Missing	0	0				
Total	21405705	23362979.1			17381276.42	19383654.34

Since undercount was lower in Census 2011 compared to Census 2001, the difference between adjusted and unadjusted counts in respective age groups in the former census was smaller compared to those in Census 2001. The undercount rate in various male age groups did not differ much a fact that meant percentage differences between adjusted and unadjusted counts across age groups did not vary. This was indicated by the reciprocal adjustment factors which did not vary much across age groups. For instance, the least reciprocal adjustment factor was 0.848 in age group 0-4 years, while the highest was 0.892 in age group 10-14 years. The same can be said for females whose reciprocal adjustment factors ranged between 0.851 and 0.909 in age groups 0-4 and age groups above 59 years respectively [Table 8.3]

Table 8.3 Adjusted and unadjusted counts 2011

Age group	PES Adjusted Counts		Proportion of Population enumerated		"Unadjusted" counts	
	Males	Females	Males	Females	Males	Females
0-4	2867584.9	2817867	0.848	0.851	2431711.995	2398004.817
5-9	2425181	2394570	0.887	0.885	2151135.547	2119194.45
10-14	2344275	2250611	0.892	0.887	2091093.3	1996291.957
15-19	2498572	2504905	0.872	0.874	2178754.784	2189286.97
20-24	2694646	2679896	0.872	0.874	2349731.312	2342229.104
25-29	2542681.7	2516635	0.872	0.874	2217218.442	2199538.99
30-34	2036206	1992804	0.807	0.855	1643218.242	1703847.42
35-39	1709346.5	1758420	0.807	0.855	1379442.626	1503449.1
40-44	1402328	1546291	0.807	0.855	1131678.696	1322078.805
45-49	1195740	1424543	0.851	0.895	1017574.74	1274965.985
50-54	1011349	1206940	0.851	0.895	860657.999	1080211.3
55-59	811949.96	985458.39	0.851	0.895	690969.416	881985.2591
60-64	612363.96	773404.01	0.851	0.895	521121.73	692196.589
65-69	401548.2	556256.456	0.888	0.909	356574.8016	505637.1185
70-74	297144.509	453343.42	0.888	0.909	263864.324	412089.1688
75-79	163690.73	317675.03	0.888	0.909	145357.3682	288766.6023
80+	174182.94	402149.69	0.888	0.909	154674.4507	365554.0682
Missing	0	0				
Total	25188790.9	26581769.3			21584779.77	23275327.7

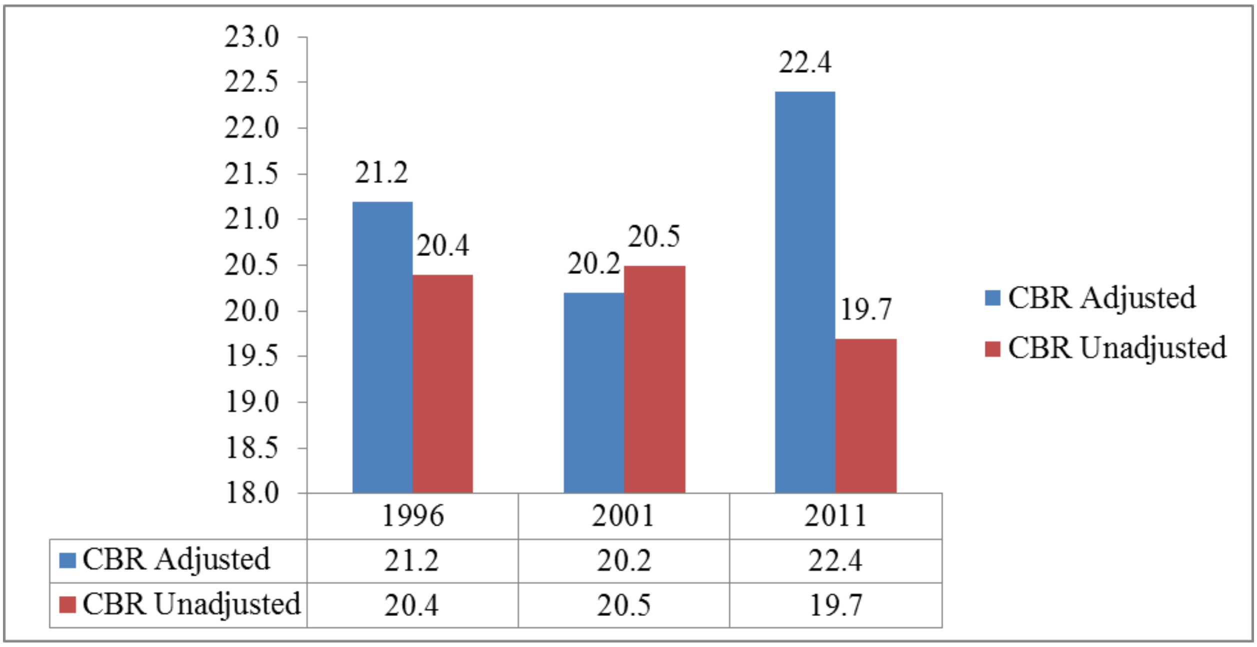
8.3 Demographic parameters of adjusted and unadjusted data sets

8.3.1 Crude birth rate (CBR)

If the undercount data of the 1996 Census had not been adjusted there would have been an underestimation of CBR compared to CBR obtained from adjusted counts. The patterns observed in Census 2001 however contradicted those observed in Census 1996. The unadjusted counts in Census 2001 suggested a higher CBR relative to that of adjusted counts. In the case of the 2001 Census, the high undercount of women of reproductive age could have led to CBR in unadjusted counts. Adjusting this error resulted in lower CBR. Otherwise the undercount adjustment for these women was excessive and hence over-escalated the counts of these women. If so, this could be one of the case examples of the inaccuracy of the PES.

The biggest difference between CBRs in unadjusted and adjusted counts was observed in Census 2011. As in Census 1996, CBR in adjusted counts was higher than that in unadjusted counts. The smallest difference between CBRs in adjusted and unadjusted counts was observed in Census 2001, indicating less under-enumeration of infants in this census compared to the other two censuses. [Fig 8.1]

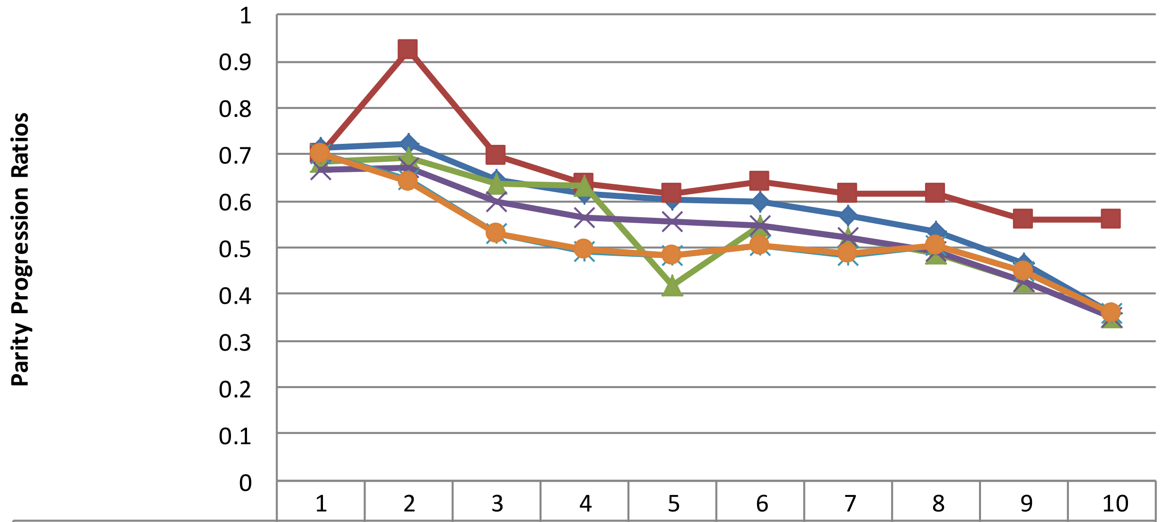
Fig 8.1 CBR for adjusted and unadjusted data



8.3.2 Parity progression ratio analysis

This is a fertility measure that estimates the probability that women in a lower birth order would progress to the next birth order. Women with no child are at birth order 0; whereas those with one child are at birth order one, etc. The high probability of moving to a next birth order among women at a particular parity implies more births are being added to a population. In Census 1996, the probability of women moving from their current parity to the next parity based on unadjusted counts was generally higher across all parity levels than estimates based on respective adjusted counts. As for Census 2001, the proportion of women choosing to increase the number of children at each birth order was higher in adjusted counts relative to unadjusted counts, except for women at child number 5. With regards to Census 2011, the probability of moving from each birth order to the next based on adjusted and unadjusted counts was virtually the same in each birth order [Fig 8.2]. Maybe this is because the ratios of women giving birth at given parities in adjusted counts were the same as those in respective unadjusted counts [Fig 8.2 here]

Fig 8.2 Parity progression ratios for censuses 1996, 2001, & 2011



	1	2	3	4	5	6	7	8	9	10
Adjusted 1996	0.7153	0.72093	0.64459	0.6166	0.60409	0.59803	0.56977	0.5348	0.46546	0.35928
Unadjusted 1996	0.69932	0.92247	0.69819	0.63848	0.61462	0.64117	0.61707	0.61538	0.55852	0.5614
Adjusted 2001	0.68303	0.69172	0.63638	0.632	0.41773	0.54739	0.52243	0.4892	0.42761	0.34993
Unadjusted 2001	0.66618	0.67018	0.59743	0.56434	0.55748	0.54811	0.52307	0.48972	0.42807	0.35007
Adjusted 2011	0.70084	0.64402	0.53002	0.4928	0.48274	0.50481	0.48485	0.50535	0.45047	0.35958
Unadjusted 2011	0.69987	0.64335	0.53031	0.49365	0.48369	0.50556	0.48549	0.50547	0.45038	0.36018

8.3.3 Age-specific fertility rates

The age-specific fertility rates (ASFR) that were computed from adjusted counts in census 1996 were higher than respective ASFRs computed from respective unadjusted counts. This meant undercount underestimated fertility rates for 1996. However, the ASFR counts in Census 2001 from adjusted data were all lower relative to those from respective unadjusted counts. For example, age groups 15-19, 20-24, 25-29 had 37.1 children per thousand women compared to 41/1000; then 12.6/1000 relative to 13.9; and 15.6/1000 relative to 17, 1/1000 respectively. However, there were instances where both data sets would agree. This meant undercount generally underestimated fertility rates in 2001. In Census 2011, as in Census 1996, higher ASFRs were observed in adjusted counts compared to those in 'unadjusted' counts. For instance, among women aged 15-19 years, the adjusted data estimated 39.7 children per 1000 women compared to 35.6 children per 1000 women, in unadjusted data.

However, in all censuses, age groups associated with either high or low fertility were identical. [Table 8.4]

Table 8.4 Age specific fertility rates for censuses 1996, 2001 & 2011

Age groups	1996 adjusted		1996 "unadjusted"	
	Females Pop	ASFR	Females Pop	ASFR
15-19	2 359 138	0.0357	2 118 506	0.0348
20-24	2 135 672	0.1212	1 885 799	0.1182
25-29	2 067 653	0.1506	1 825 737	0.1471
30-34	1 790 413	0.0803	1 609 581	0.0772
35-39	1 697 576	0.0300	1 454 201	0.0276
40-44	1 375 399	0.0085	1 236 484	0.0073
45-49	1 105 325	0.0011	1 020 215	0.0010
TFR		2.1		2.07
GFR		68.8/1000		66.3/1000
	2001 adjusted		2001"unadjusted"	
15-19	2527782	0.0371	2044418.572	0.0411
20-24	2189344	0.1258	1631749.728	0.1394
25-29	2034172	0.1559	1471016.4	0.1717
30-34	1741231	0.0849	1253456.6	0.0986
35-39	1635554	0.0335	1129158.13	0.0441
40-44	1376879	0.0102	965882.1335	0.0155
45-49	1125861	0.0013	787454.1377	0.0018
TFR		2.3		2.6
GFR		71.6 /1000		81.2/1000
	2011 adjusted		2011"unadjusted"	
15-19	2504905	0.0397	2189286.97	0.0356
20-24	2679896	0.1346	2342229.104	0.1207
25-29	2516635	0.1661	2199538.99	0.1501
30-34	1992804	0.0938	1703847.42	0.0799
35-39	1758420	0.0404	1503449.1	0.0296
40-44	1546291	0.0136	1322078.805	0.0083
45-49	1424543	0.0017	1274965.985	0.0011
TFR		2.5		2.1
GFR		80.4/1000		70.5/1000

Pop= population, ASFR= age specific fertility rates, TFR=total fertility rate, GFR= general fertility rate

8.3.4 Total fertility rates

Undercount in Census 1996 had the effect of underestimating total fertility rate (TFR), which was at 2, relative to a TFR of 2.2 in adjusted counts. The two TFRs confirm higher ASFR that were observed in adjusted counts relative to those observed in respective 'unadjusted' counts in Census 1996. In Census 2001, TFR in adjusted counts was at 2.3. This was lower than the 2.6 TFR observed in unadjusted counts. TFR in adjusted counts in Census 2011 was higher than that in unadjusted counts. It can be noted that TFR in either adjusted or unadjusted data in each census was a function of age-specific fertility rates that would have been estimated from the respective data. In cases where age-specific fertility rates in e.g. adjusted data were higher than those in unadjusted data, it would automatically follow that the former would have a higher TFR than the latter [Table 8.4 above].

8.3.5 General fertility rates

General fertility rates (GFR) in adjusted counts in censuses 1996 and 2011 were higher than GFRs in respective unadjusted counts. In Census 2001, the GFR just like TFR was lower in adjusted counts compared to one estimated in unadjusted counts. For example, children expected to be born per 1 000 women of reproductive age group, based on adjusted data, were 10.4 less 9 more relative to those expected to be born based on unadjusted data. [Table 8.4 above]

8.4 Concluding remarks

There were insignificant differences between CBRs computed in adjusted and unadjusted counts, especially for censuses 1996 and 2001. Clear differences were noted in other demographic parameters like parity progression ratio, total fertility rates, general fertility rates and age-specific fertility rates. The overall finding was that undercount led to inaccurate demographic parameters as it either inflated or deflated these parameters. This was demonstrated by the differences between demographic parameters in counts corrected for the undercount (adjusted) and respective counts that had the undercount error (unadjusted).

CHAPTER 9: EFFECT OF UNDERCOUNT ON SERVICE DELIVERY

9.1 Introduction

The chapter investigated the effect of undercount on service delivery with regards to resource allocation. Accurate census counts ensure fair and justified service provision. Allocations of Provincial Equitable Share Funds (PESFs) among South Africa's nine provinces from respective adjusted and unadjusted counts were computed and compared. The latter's allocation relative to the respective fund allocations in the former indicates the effect of undercount on service delivery as the former counts include undercount error. PESF is the total amount of money from various grants which the central government of South Africa allocates to provinces. The grants that form these funds are seven and they are described in Table 9.5.

The allocations are provided yearly, and assuming the same formula is used until the next review, the PESFs were projected for the period between successive reviews. This was done for allocations in both adjusted and unadjusted counts. The first assessment indicated yearly gains or losses due to the undercount in provinces. The second assessment gave the projected loss of funds due to undercount in each province between the successive review periods. Secondly, parliamentary seat allocation estimates of provinces based on adjusted and unadjusted data sets were also examined. This was done for the four elections that were conducted in South Africa since the democratisation of the country.

Both the funds from national grants and parliamentary seat allocations are proxies often used for estimating the quality of service delivery in countries like the United States of America. Fund allocations from national grants are suitable measures of service delivery because the higher the funds received by a subpopulation, the higher the potential for delivering quality services to the respective subpopulations. However, if allocations are affected by undercount some subpopulations are likely to be prejudiced since they may receive less allocation of funds relative to what they would have got if their counts were not affected by the undercount error. There is also a possible scenario that certain subpopulations may get higher funds or parliamentary seat allocations from counts which have not been corrected for undercount error relative to what they would have been allocated based on counts adjusted for this error. This is termed unfair/unjustified and benefits this study

The first table in the chapter describes the official adjusted counts and their respective unadjusted counts, undercount rates, and proportion of enumerated population that were given by StatsSA. This information was only available for Census 1996. In censuses 2001 and 2011, StatsSA only availed adjusted counts for public use. Availability of official unadjusted count for the country's provinces in Census 1996 provided an opportunity for testing the quality of reconstructed unadjusted counts. This explains the inclusion of tables 9.1 and 9.2.

The former table shows the official unadjusted counts from StatsSA for Census 1996 by province, and the latter table shows respective reconstructed unadjusted counts in the same census. A comparison of the respective official and reconstructed unadjusted counts shows that that the latter were good estimates of the former. Hence the study assumes that all other reconstructions of unadjusted counts that were conducted produced credible counts for use in the analyses. For the sake of consistence, all the analyses used the reconstructed unadjusted counts, including analyses in Census 1996 where there was an option to use the official unadjusted counts.

The argument of the chapter is that findings would reduce controversies around accuracy of the method if they indicate that the PES adjustments for undercount ensured fair service delivery among the country's nine provinces. [Table 9.1]

Table 9.1 Census 1996 undercount estimates by Province based actual enumerations

Province	1996 PES Adjusted counts	1996 Unadjusted counts	Estimated Undercount		Proportion of population enumerated
			Numbers	Rate	
1	3956875	3612835	344,040	8.69	0.9131
2	6302525	5636408	666,117	10.57	0.8943
3	840321	709348	130,973	15.59	0.8441
4	2633505	2403009	230,496	8.75	0.9125
5	8417021	7338554	1,078,467	12.81	0.8719
6	3354825	3040607	314,218	9.37	0.9063
7	7348423	6614205	734,218	9.99	0.9001
8	2800711	2518065	282,646	10.09	0.8991
9	4929368	4373560	555,808	11.28	0.8872
	40583574	36246591	4,336,983		

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

Source: Statistics South Africa, 1998

9.2 Adjusted counts relative to unadjusted counts, Census 1996

Undercount rates for all census data used in this study were obtained from various sources from StatsSA, and they publish them after every census when they document the processes and outcomes of respective censuses. Besides suggesting the percentage of people missed in censuses, undercount rates also suggest percentage difference between adjusted relative to unadjusted counts. The largest percentage difference between adjusted and unadjusted counts in Census 1996 was in the Northern Cape. This is at 15.9% and absolute difference in figures was estimated at 131 059. This was followed by KwaZulu-Natal where adjusted counts were 12.8% more than unadjusted counts, implying that in terms of figures adjusted counts were 1 078 271 more than unadjusted counts. Provinces that had the least differences between adjusted and unadjusted counts were the Western Cape and the Free State, recording percentage differences of 8.7% and 8.8% respectively. [Table 9.2]

Table 9.2 Census 1996 undercount estimates by Province

Province	1996 Adjusted count	1996 Unadjusted counts	Estimated Undercount		Proportion of population enumerated
			Numbers	Rate	
1	3957322.3	3613430.992	343,891.308	8.69	0.9131
2	6301972	5635853.56	666,118.44	10.57	0.8943
3	840662.17	709602.9377	131,059.2323	15.59	0.8441
4	2633408.6	2402985.348	230,423.252	8.75	0.9125
5	8417423	7339151.114	1,078,271.886	12.81	0.8719
6	3355012	3040647.376	314,364.624	9.37	0.9063
7	7348071	6613998.707	734,072.293	9.99	0.9001
8	2797692	2515404.877	282,287.123	10.09	0.8991
9	4927336	4371532.499	555,803.501	11.28	0.8872
National	40578899.07	36245189.94	4,333,709.13	10.69	0.8931

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

9.3 Adjusted counts relative to unadjusted counts, Census 2001

Generally, in Census 2001, many provinces recorded high percentage differences between adjusted and unadjusted counts. This is indicated by very high undercount rates across all provinces. For instance, the lowest undercount rate, i.e. 14.4% recorded in Limpopo, was just about 1% less than the province recording the highest undercount in Census 2001. This again implied that unadjusted counts were 14.4% lower than adjusted counts. Provinces with the highest differences between the two counts in Census 2001 were KwaZulu-Natal, Gauteng and the Free State. KwaZulu-Natal had an undercount rate of 22.5%, which meant that unadjusted counts had the same percentage to tally with adjusted counts. In terms of absolute figures, the latter were about 2 119 717 more than the former. Such a high percentage difference between adjusted and unadjusted data across all provinces meant that the overall difference between adjusted and unadjusted counts was equally high at national level, i.e. 17%. [Table 9.3]

Table 9.3 Census 2001 undercount estimates by Province

Province	2001 PES Adjusted counts	2001 unadjusted counts	Estimated Undercount		Proportion of population enumerated
			Numbers	Rate	
1	4513206	3777553.422	735,652.578	16.30	0.837
2	6415451.7	5472380.3	943,071.4	14.70	0.853
3	823429.16	678505.6278	144,923.5322	17.60	0.824
4	2715587	2207772.231	507,814.769	18.70	0.813
5	9420961	7301244.775	2,119,716.225	22.50	0.775
6	3662194	3076242.96	585,951.04	16	0.84
7	8830155	7178916.015	1,651,238.985	18.70	0.813
8	3125664	2622432.096	503,231.904	16	0.839
9	5262037	4504303.672	757,733.328	14.40	0.856
Total	44,768,685	36,819,351.10	7,949,333.761	17%	0.83

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

9.4 Adjusted counts relative to unadjusted counts, Census 2011

Though slightly improved, the percentage differences of adjusted counts relative to unadjusted counts remained high in Census 2011. The worst-affected province was the Western Cape where the percentage difference between adjusted and unadjusted data was 18.5%. This culminated in a difference of about 1 077 203 people between the two data sets for the province. Two other provinces that had high percentage difference between adjusted and unadjusted counts were KwaZulu-Natal, 16.7%, and Mpumalanga, 15.5%. This resulted in absolute differences between adjusted and unadjusted counts of 1 714 639 and 626 191 respectively for the two provinces. However, what can be noted is that for all the censuses, even in case of provinces reporting lowest percentage difference between adjusted and unadjusted counts, the differences in absolute figures represented substantial populations as they all ran in hundreds of thousands or more. [Table 9.4]

Table 9.4 Census 2011 undercount estimates by Province

Province	2011 PES Adjusted counts	2011 unadjusted counts	Estimated Undercount		Proportion enumerated
			Numbers	Rate	
1	5822734	4745528.21	1077205.79	18.50	0.815
2	6562053	5715548.163	846504.837	12.90	0.871
3	1145861	992315.626	153545.374	13.40	0.866
4	2745590	2468285.41	277304.59	10.10	0.899
5	10267300.4	8552661.233	1714639.167	16.70	0.833
6	3509953	2986970.003	522982.997	14.90	0.851
7	12272262.9	10468240.25	1804022.65	14.70	0.853
8	4039939	3413748.455	626190.545	15.50	0.845
9	5404868	4864381.2	540486.8	10	0.9
	51,770,561	44,207,678.55	7562882.75	14.6	0.854

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

9.5 The Provincial Equitable Share Fund

In line with promoting service delivery, the national government of South Africa allocates funds to provinces that in turn are expected to plan their utilisation to the benefit of their populations. These funds are distributed under the umbrella of the Provincial Equitable Share Fund (PESF). The main equitable share components are Education, Health, Social Development, Economic Activity, Backlog, Basic Services Grant, and Institutional. In the table below, the various equitable share components were shown together with their corresponding funds allocations for the specified years 1999, 2004, 2013. These years mark points in time at which increments were reviewed. Until a new review of allocated funds to a share component is done, the last review remains in force for successive years. Education and Health had claimed the largest shares at each review point, where the former received a whooping R162 034 758 in the latest review in 2013. Furthermore, it can be noted that the different equitable share components use different formulas for sharing the allocated funds to provinces. All grants except the BSGF use formulas that are not necessarily based on census counts. These are Education, Health, Social Development, Economic Activity, Backlog, and Institutional Grants, and their formulas are stated adjacent to them. All these were listed in Table 9.5, but they were irrelevant for this analysis. Allocations for the Education Grant partly use census counts in their formula. The grant became relevant for the analysis as it also incorporated total enrolments of children in the formula and these were not derived from census counts.

The BSG grant was the only grant where the fund allocations were based entirely on census counts. In the years 1999, 2004 and 2013, the allocations for this share component from the National Treasury were: R7 578 180, R11 197 998 and R54 011 586 respectively. This would then be shared among provinces according to the proportions of their respective population sizes in relation to total national population as drawn from the census. [Table 9.5]

Table 9.5 Provincial Equitable Share in South Africa

Equitable share component	**Data used**	**Formula**	**Year**	**Amount in Rands (rounded off)**
Education	Total enrolment numbers (Ai)	$Ai+2\ [(Pi_{6\text{-}16})]/+\sum i2\ [(Pi_{6\text{-}16})]$	1999 2004 2013	**33 680 800** **65 588 275** **162 034 758**
	School age cohort (6-17 years) ($Pi_{6\text{-}16}$)]			
Health	Population with and without Medical	$(Phi+4Pwi)/\sum i(Phi+4Pwi)$	1999 2004	**15 156 360** **30 394 566**

	AID		2013	**91 144 551**
Social Development	Target population for each grand type	Sum of grants (weighted 75) and Provincial population income quintiles (weighted 25). Grant values are sum of population eligible for old age grant (65), childcare grant (10) and population	1999 2004 2013	**14 314 340** **28 794 852** **10 127 172**
Economic activity	Gross geographic product (replaced by Remuneration data in 2000)		1999 2004 2013	**6 736 160** **11 197 998** **3 375 724**
Backlog	School Survey of Needs Health Sectorial Report		1999 2004 2013	**2 526 060** **4 799 142**
Basic Services Grant	Census	Provincial share of Population	1999 2004 2013	**7 578 180** **11 197 998** **54 011 586**
Institutional	Independent data		1999 2004 2013	**4 210 100** **7 998 570** **16 878 621**

Source: (South Africa) National Treasury. IGFR, 2014

9.6 Yearly and projected funds allocations based on census 1996

The baseline for the Basic Service Grant was given in 1999, and the allocations were based on the latest census counts of 1996. It was indicated that certain provinces would have lost revenue from the grant due to undercount, yet other provinces would have benefitted at the expense of those losing. Three provinces that would have been prejudiced by undercount are KwaZulu-Natal, the Northern Cape and Limpopo. The biggest loser among the three provinces in terms of the absolute amounts would have been KwaZulu-Natal – a fact that should be explained by both the high population size and undercount rate recorded in the province compared to the other two provinces. The other six provinces would have gained revenue above what was due for them, at the expense of the above three. The biggest beneficiary from an undercount would have been the Western Cape, followed by the Free

State, Gauteng, North-West, Mpumalanga and the Eastern Cape respectively. Results confirm that the magnitude of either the loss or gain from undercount is a function of population size and level of undercount.

Assuming that the allocations of the Basic Services Grant funds relied on the Census 1996 counts from baseline until next in 2004, a five-year projection of the provinces' respective losses or gains from undercount was estimated. KwaZulu-Natal would have been the worst prejudiced province, with a projected loss of around R187 456 followed by the Northern Cape. Among those gaining from undercount, the highest projected gains would have been for the Western Cape Province, i.e. around R82 314. Gauteng would have gained R52 981.87. [Table 9.6]

Table 9.6 Basic Services Grant allocation based on 1996 census

Provinces	Basic Services Grant funds allocations for 1999. Based on census 1996 counts.			Projected funds allocation for period 1999-2003		
	Adjusted (A)	Reconstructed (B)	Difference (A - B)	Adjusted (D)	Unadjusted (E)	Difference (D -E)
1	739 036.83	755 499.71	-1 6462.94	3 695 184.1	3777498.55	-82314.702
2	117 6904.39	1 178 349.90	-1445.66	588 451.95	5891749.51	-7228.324
3	156 995.12	148 364.49	8630.60	784 975.6	741822.47	43152.9867
4	491 793.64	502 418.68	-10624.97	2 458 968.2	2512093.39	-53124.824
5	1 571 968.4	1 534 477.12	37491.27	7 859 842	7672385.62	187456.342
6	626 554.33	635 741.68	-9187.35	3 132 771.65	3178708.38	-45936.735
7	1 372 265	1 382 861.42	-10596.38	6 861 325	6914307.10	-52981.871
8	522 473.85	525 923.35	-3449.50	2 612 369.25	2629616.77	-17247.524
9	920 188.57	914 004.21	6184.36	4 600 942.85	4570021.07	30921.791

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

9.7 Yearly and projected funds allocations based on census 2001

Negative difference indicates gains and positive indicates losses in provinces due to undercount. If a province got less under adjusted count allocations (A) compared to what would be allocated from unadjusted counts (B) this would imply a negative total after subtracting allocation of unadjusted counts from allocations of respective adjusted counts (A-B) as shown in tables 9.6; 9.7 and 9.8

The reviewed allocation for BSGF in 2004 raised the grant's allocation from the previous amount of R33 680 800 to R65 588 275. This annual allocation remained in force until the new grant review in 2013. Again three provinces were bound to be prejudiced by undercount, whereas six provinces would have benefited from this. KwaZulu-Natal would have lost an amount of R135 513.88 due to undercount. Projecting this loss to year 2012, it meant the province would have lost R1 223 219.08. This again would have been the biggest loss of all the provinces. Gauteng was bound to lose R25 338.23, and when projected the amount would be R228 049.91. The Free State Province would have lost R7 792.77 yearly and R70 135.93 after projection.

The three provinces that were going to be the highest beneficiaries from the prejudice due to undercount were the Eastern Cape, the Western Cape, and Mpumalanga. The Eastern Cape, which was bound to be the biggest beneficiary, would have received an extra R59 636.50, and when projected the gain would have risen to R536 727.01. [Table 9.7]

Table 9.7 Estimates of funds allocation based on 2001 census counts by Province

Provinces	Basic Services Grant funds allocations based on census 2001 counts			Projected funds allocation for period 2004-2012		
	Adjusted (A)	Reconstructed (B)	Difference (A - B)	Adjusted (D)	Unadjusted (E)	Difference (D -E)

1	1 128 888.9	1148880.41	-19991.50	10 160 000.1	10339923.66	-179923.51
2	1 604 698	1664334.07	-59636.33	14 442 282	14979006.63	-536727.01
3	205 964.47	206356.71	-392.29	1 853 680.23	1857210.364	-3530.57
4	679 250.21	671457.42	7792.77	6 113 251.89	6043116.783	70134.93
5	2 356 466.5	2220553.18	135913.23	21 208 198.5	19984978.61	1223219.08
6	916 025.17	935588.54	-19563.40	8 244 226.53	8420296.888	-176070.60
7	2 208 688.1	2183349.14	25338.88	19 878 192.9	19650142.23	228049.91
8	781 822.84	797569.72	-15746.91	7 036 405.56	7178127.485	-141722.17
9	1 316 194.2	1369909.73	-53715.61	11 845 747.8	12329187.57	-483440.52

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo. NB: Negative difference indicates loss and positive indicates gains by provinces, due to undercounting

8.8 Yearly and projected funds allocations based on census 2011

The Basic Services Grant review of 2013 doubled from R65 588 275 to R162 034 758. For this reason each province’s allocations from the grant were above R1 000 000. For instance, based on the 2011 adjusted census counts, the province with the least population counts was the Northern Cape, which received R1 195 462.6 in 2013. If counts were not adjusted for undercount, the province could have received R1 212 381.92, implying that the province was among those that would have gained from undercount. Other provinces that would have gained from undercount include the Eastern Cape, which would have been the biggest beneficiary, the Free State and Limpopo.

Five provinces would have been prejudiced by undercount. As in previous findings, KwaZulu-Natal would have been the biggest loser. For a single year, the province would have lost R262 368.06, translating to R2 361 312.50 after projection. Other provinces that would have been prejudiced by the Census 2011 undercount were the Western Cape, North-West, Gauteng and Mpumalanga. [Table 9.8]

Table 9.8 Estimates of funds allocation based on 2011 census counts by Province

Provinces	Basic Services Grant funds allocations based on census 2011 counts			Projected funds allocation for period 2013-2022		
	Adjusted (A)	Reconstructed (B)	Difference (A - B)	Adjusted (D)	Unadjusted (E)	Difference (D -E)
1	6 074 786.2	5797940.658	276845.71	54 673 075.8	52181465.93	2491611.38
2	6 846 108.5	6983081.492	-136972.63	61 614 976.5	62847733.43	-1232753.66
3	1 195 462.6	1212381.319	-16918.67	10 759 163.4	10911431.87	-152267.99
4	2 864 440	3015675.127	-151234.88	25 779 960	27141076.14	-1361113.90
5	10 711 747	10449379.84	262368.06	96 405 723	94044418.58	2361312.51
6	3 661 890.4	3649388.732	12502.16	32 957 013.6	32844498.59	112519.45
7	12 803 500	12789776.2	13724.39	115 231 500	115107985.8	123519.49
8	4 214 818.3	4170814.436	44004.50	37 933 364.7	37537329.93	396040.53
9	5 638 831.8	5943152.491	-304319.72	50 749 486.2	53488372.42	-2738877.46

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo. NB: Negative difference indicates gain and positive indicates loss by provinces, due to undercounting

9. 9: Population based parliamentary seats distribution by province

The Electoral Act of South Africa defines an eligible voter as an individual aged 18 years and above, who is registered to vote. The country has a maximum of 400 parliamentary seats that are shared among provinces. From this total, 200 seats are shared among provinces according to population registered as voters. The formula is based on a parliamentary seat that is allocated to each 100 000 eligible voters. The analysis below is based on the assumption that all people aged 18 years and above in each province according to the last census before the respective election would have registered as voters. Each census includes two sets of counts, i.e. adjusted and 'unadjusted'. Therefore the result tables that follow described parliamentary seat allocation drawn from three scenarios. The first scenario is the official seat allocation based on actual people who registered to vote as obtained from the Independent Electoral

Commission (IEC) of South Africa. The second scenario is the expected seat allocation based on adjusted census counts. The assumption that was made is that all people above 18 years were registered to vote. The third scenario is the expected seat allocation based on 'unadjusted' census counts, assuming all people above 18 years also registered to vote.

9.10 Election 1999 parliamentary seat allocation

According to the IEC of South Africa's parliamentary seat allocation, Gauteng received the highest number of seats, followed by KwaZulu-Natal, i.e. 43 and 40 seats respectively. The Eastern Cape and Limpopo were third and fourth. All other provinces were allocated parliamentary seats that were above 10, except for the Northern Cape which was allocated 4. However, if all people in the voting age range in census 1996 adjusted counts registered to vote, parliamentary seats for Gauteng would be about 52, and for KwaZulu-Natal about 49. The Northern Cape would have added one to make it 5. Comparing these with parliamentary seats expected from 'unadjusted' data in the same conditions, Gauteng would have been allocated about 47 seats and KwaZulu-Natal 42. This would be 4 and 6 seats less than those expected from adjusted data, respectively. Based on adjusted counts, the Northern Cape would have been allocated 5 seats, which is one more than 4 seats expected from 'unadjusted' counts, but the latter allocation would be the same as one by the IEC of South Africa. [Table 9.10]

Table 9.10 Official and Expected Provencal seat allocation for Election 1999

Province	Official Seats (Based on Voter registration)	10% sample Adjusted 1996 census count (18 years & above)	Expected seats (Based on 10% 1996 adjusted count)	Unadjusted (18 year & above)	Expected seats (based on 1996 reconstructed)
1	21	2604325.7	26.0	2378009.797	23.8
2	26	3364049.7	33.6	3008469.647	30.1
3	4	508123.6	5.08	428907.1308	4.3
4	15	1646063.5	16.5	1502032.944	15.0
5	40	4870770.4	48.7	4246824.712	42.5

6	17	2003052.5	20.0	1815366.481	18.2
7	43	5174424	51.7	4657499.042	46.6
8	14	1589657.3	15.9	1429260.878	14.3
9	20	2471035.3	24.7	2192302.518	21.9
Total	200	24,231,502	242.3	21641154.44	216.4

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo.

9.11 Elections 2004

Elections 2009 had the same official parliamentary seat allocation by province as Election 2004. Also the census counts that applied for both elections were from Census 2001. For this reason only, one election, i.e. Election 2004, was analysed, because the two would still have the same results. In the official seat allocations by IEC for Election 2004, Gauteng remained the main beneficiary receiving 47 seats, 5 more than 43, in Election 1999. KwaZulu-Natal also remained the second highest beneficiary, but its seat allocations fell from 40 to 39 seats. Provinces that gained were the Western Cape from 21 in Election 1999 to 23 in Election 2004. The Northern Cape gained one more, whereas Limpopo and the Eastern Cape each lost one seat. Mpumalanga and the Free State also gained.

However, comparing the expected parliamentary seat allocations by provinces based on Census 2001 counts, using the same assumptions as those stated above, the expected seat allocation for e.g. the Western Cape based on unadjusted counts was 25 relative to 30 from adjusted counts. In the Eastern Cape unadjusted counts would lead to 30 seats relative to about 36 from adjusted counts; in the Northern Cape, 4 relative to 5; the Free State, 13 relative to 1; and North-West, 19 relative to 23 seats. [Table 9.11]

Table 9.11 Official and Expected Provincial seat allocation for Election 2004/9

Province	Official Seats (Based on Voter registration)	10% sample Adjusted 2001 census count (18	Expected seats (Based on 10% sample 2001	Unadjusted (18 year & above)	Expected seats (based on 2001 reconstructed)

		years & above)	adjusted count)		
1	23	3013024	30.1	2521901.088	25.2
2	26	3550922.9	35.5	3028937.234	30.3
3	5	519206	5.2	427825.744	4.3
4	12	1697750.4	17	1380271.075	13.8
5	39	5470630.9	54.7	4239738.948	42.4
6	14	2285648.1	22.9	1919944.404	19.2
7	47	6314155.9	63.1	5133408.747	51.3
8	15	1803704.7	18.0	1513308.243	15.1
9	19	2763913.9	27.6	2365910.298	23.7
Total	200	27,418,957	274.2	22757734.31	227.6

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo

9.12 Election 2014

In Election 2014, assuming all people aged 18 years and above in Census 2011 registered to vote, the proportion of seat allocation to each province based on adjusted counts would have been substantially high. For example, provinces with the largest population sizes, like Gauteng and KwaZulu-Natal, would have been allocated 87 and 63 seats respectively, but from unadjusted counts the same provinces would have been allocated 76 and 53 seats respectively. From adjusted counts the Western Cape would have been allocated about 41, the Eastern Cape 39 and the Northern Cape 7. From unadjusted data the three provinces would have experienced a drop to 33, 34 and 6 seats respectively. Just as in the case of provinces noted above, other provinces like the Free State, Mpumalanga and Limpopo, the 'unadjusted' counts would have resulted in lower parliamentary seat allocation to these respective provinces relative to adjusted counts. However, the effect of undercount would vary. For instance, in the Free State the difference between the two data sets would have been one seat, whereas in Mpumalanga it would have been 4 seats and in Limpopo 3 seats. [Table 9.12]

Table 9.12 Official and Expected Provincial seat allocation for Election 2014

Province	Official Seats (Based on Voter registration)	10% sample Adjusted 2011 census count (18 years & above)	Expected seats (Based on 10% sample 2011 adjusted count)	Unadjusted (18 year & above)	Expected seats (based on 2011 reconstructed)
1	26	4082578.7	40.8	3327301.641	33.3
2	18	3940532.6	39.4	3432203.895	34.3
3	16	735225.4	7.4	636705.1964	6.4
4	17	1793776.9	17.9	1612605.433	16.1
5	18	6309020.5	63.1	5255414.077	52.6
6	16	2280852.6	22.8	1941005.563	19.4
7	22	8848775.7	88.5	7548005.672	75.5
8	16	2524094.1	25.2	2132859.515	21.3
9	20	3190339.1	31.9	2871305.19	28.7
Total	186	33,705,196	337.1	28784237.38	287.8

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo.

9.13 Proportional distribution of seats

The allocation of seats based on adjusted and unadjusted counts analysed above meant that the number of seat allocated from the latter would naturally be less than those allocated from the former. This is because the latter has fewer counts than the former. Therefore a further analysis was to compare the proportional seat allocations to provinces from adjusted counts relative to unadjusted counts. In Election 1999, the Western Cape would have gained from the undercount as unadjusted counts would have resulted in the province getting 11% of the total seats allocated to all provinces. This was relative to 10.7% from adjusted counts. From the official allocations the province received 10.5% of the total seats. Other provinces that would have gained from the undercount during this election are the Free State and Gauteng. The Northern Cape, KwaZulu-Natal and Limpopo would have been prejudiced while Mpumalanga and the Eastern Cape would have retained the same proportion of seats from either adjusted or unadjusted counts.

Allocations for elections 2004 and 2009 were both analysed using Census 2001 counts since there was no census in 2006. The Western Cape, Eastern Cape, Mpumalanga and Limpopo would have gained from the undercount. The other five provinces would have been prejudiced by the undercount. As for Election 2014, provinces that would have received less proportional allocations of parliamentary seats due to undercount are the Western Cape and KwaZulu-Natal. Only one province, i.e. Mpumalanga, would have neither lost nor benefited from the undercount since the proportion of seats the province would have obtained based on either adjusted or unadjusted counts remained the same. Provinces that would have benefited from the undercount are the Eastern Cape, Gauteng, the Northern Cape, North-West, the Free State and Limpopo. [Table 9.13]

Table 9.13 Proportional allocation of seats

Province	Proportional allocations (%) Election 1999			Proportional allocations (%) Election 2004 & 2009			Proportional allocations (%) Election 2014		
	Official	Adjusted	Unadjusted	Official	Adjusted	Unadjusted	Official	Adjusted	Unadjusted
1	10.5	10.7	11	11.5	11	11.1	13.9	11.9	11.5
2	13	13.9	13.9	13	12.9	13.3	9.6	11.6	11.9
3	2	2.1	2	2.5	1.9	1.89	8.6	2.1	2.2
4	7.5	6.8	6.9	6	6.2	6.06	9.1	5	5.6
5	20	20.1	19.9	19.5	19.9	18.6	9.6	18.7	18.3
6	8.5	8.3	8.4	7	8.35	8.44	8.6	6.5	6.7
7	21.5	21.3	21.5	23.5	23	22.5	11.8	26.1	26.2
8	7	6.6	6.6	7.5	6.56	6.63	8.6	7.4	7.4
9	10	10.2	10.1	9.5	10.1	10.4	10.8	9.2	10
Total	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**

8.14 Concluding remarks

Findings from the study's prior analyses have already suggested the credibility of the PES and its adjustment outcomes. This implies that allocations based on adjusted counts are more credible than those based on unadjusted data. Hence the allocations from the BSGF among the country's nine provinces were based on credible census counts. Allocations from unadjusted counts would have led to unfair distribution of resources. Official allocation of resources in South Africa has been based on adjusted counts from these censuses. By indicating that allocations based on adjusted counts were fair compared to those based on unadjusted, these findings significantly contribute to reducing controversies around the accuracy of the PES.

CHAPTER 10: DISCUSSION

10.1 Introduction

The outline of this chapter is as follows. After the introduction, which stated the main objective of the study, followed a brief outline of specific objectives and how their findings addressed the main objective. Discussions of findings from analyses of specific objectives were then presented. The respective findings were discussed in relation to findings in similar studies by explaining similarities and differences noted, the current discourse around the subject, and limitations, were applicable. The findings that were discussed first were derived from the exploration of systematic patterns of undercount in South African censuses. This was followed by a discussion of findings on the test for content and coverage errors, which is better: to adjust undercount or not, comparisons of the PES and DA undercount estimate and the effect of undercount on demographic parameters and service delivery, respectively.

10.2 Summary of study objectives

The main purpose of this study was to reduce controversies associated with the country's last three censuses. All the analyses discussed in the study were therefore mainly directed at addressing this objective. The controversies centred on: the accuracy of the PES as a method for undercount estimation and adjustment. For this reason the controversies regarding the PES extended to contestations around the accuracy of both the undercount estimates and adjusted counts. Controversies characterising South Africa's censuses were a result of high undercount estimates that were recorded in these censuses. The PES, undercount estimates, and adjusted counts are directly linked in that adjusted counts are estimated from respective undercount estimates, and the undercount estimates are determined by the PES. Hence, an outcome from an investigation on either of the three has the same implication on the other two. The accuracy of the PES was at the centre of controversies in the country's censuses, largely because the method was used to determine final census counts, and these counts are used for planning and resource allocation among subpopulation groups.

Five specific objectives were set to achieve the study's main objectives and a description of how each addressed the main objective is provided below. The first objective explored whether there were systematic patterns of undercount in South African censuses. Prior researches around the discourse were explored for the purpose of determining if there were recurrent undercount patterns observed in other censuses across the globe. Such evidence was vital for this study because the observed systematic patterns of undercount from censuses conducted elsewhere across the globe were compared to those observed in South African censuses. The consistence of undercount patterns between those in South Africa's census and other censuses across the globe partly reduces the controversies as it would indicate that South Africa's undercount estimates as estimated by the PES were credible.

Findings in analyses that addressed the second objective were vital in achieving the main objective of the study in that the indications of insignificant content error in the adjusted counts of these censuses confirmed the accuracy of the PES in correcting this error. Yet the presence of content error indicated the inaccuracy of the PES in adjusting the data for this error. In case of the former being the outcome, this would confirm that critiques of the PES have no substantial grounds to contest the accuracy of the method. In the event of the latter outcome, this would be a confirmation that contestations that propagate the use of the PES have no basis. This is because the PES would have produced counts that are characterised by internal inconsistency

The findings from the analyses that determined the level of coverage error were vital in reducing controversies associated with the country's last censuses in the following ways. The findings determined whether adjusted counts were accurate or not. As noted already, adjusted counts are drawn directly from the PES and the results of the findings confirm that adjusted counts were accurate, and such findings largely confirm the accuracy of the PES and equally emphasise that there is no basis for contesting the accuracy of the PES. If it is indicated that adjusted counts have coverage errors, then this vindicates critiques of the method.

Findings in the analyses that addressed the third objective were equally significant in reducing controversies associated with South African censuses. These findings determined which census counts between adjusted and unadjusted were good estimates of the country's actual population. Agincourt HDSS counts were used as the gold standard against which respective adjusted and unadjusted counts were compared. Unadjusted counts were expected to deviate widely from respective counts of Agincourt HDSS, to indicate the effect of the undercount error which they carried. On the other hand, adjusted counts would closely estimate respective counts from the gold standard since they had been corrected for coverage error. Such findings would strongly indicate the accuracy of the PES, and would contribute greatly towards reducing controversies associated with South Africa's last three censuses. On the other hand, a reverse of the above outcome would imply that the PES was inaccurate, which would also indicate no basis for arguing in favour of the PES.

The findings from the fourth objective contribute significantly to reducing the controversies around the accuracy of the PES. The analyses addressing this objective involved comparing undercount estimates from the PES and DA. Similarities of respective undercount estimates from the two methods would directly confirm the accuracy of the PES – findings which should contribute strongly to reducing the controversies associated with South African censuses. On the other hand, if the compared undercount estimates differ, this may escalate the controversies rather than reduce them as none of the methods was a gold standard for evaluating the other. Rather the investigation was based on checking for consistency between the computations of the two undercount estimates.

The fifth objective determined the effect of undercount on demographic parameters and service delivery. This was achieved by computing and comparing demographic parameters, Basic Service Grant Funds, and parliamentary seats from adjusted and unadjusted counts. The difference between the allocations from adjusted and unadjusted counts would signify the negative implications of undercount. This would be demonstrating the importance of the PES adjustments for the undercount, a fact that should reduce controversies around the PES.

10.3 Undercount patterns

Findings confirmed a few systematic patterns of undercount in the three censuses. Males consistently had higher undercount estimates compared to their female counterparts in all three censuses. This pattern was further confirmed in the comparisons of undercount estimates for race by sex in these censuses. Other systematic patterns of undercount that characterised the three censuses were high undercount of male adults, especially those aged 20-44 years, and children below five years. Lower undercount rates were generally observed among the elderly population aged 65 years and above and to some extent among older children, particularly those aged 10-14 years.

The above patterns of undercount that largely characterised the three censuses were generally consistent with those observed in the majority of prior researches. Prior researches confirmed that censuses across the world were often characterised by high undercounts among young children and young adults, especially males (Anderson, 2004; Steffey, 1997). Under counting of children of five years and below in South African censuses can be partly attributed to a lack of information on the purpose of censuses among participants responding, e.g. on behalf of these children. This view is supported by patterns of undercount in these censuses indicating that there was consistently less undercount of children of five years and below among the white race compared to other races. It is suggested that the race is generally highly informed. From this perspective, there is a need for aggressive marketing emphasising the inclusiveness of censuses.

Child preference is a factor that researchers have also identified which partly explains why children are often undercounted compared to other subpopulation groups (Anderson, 2004). Certain societies prefer boys above girls, and the deliberate under-reporting of female children results in an overall undercount of children. Such a practice is prevalent and tolerated in some Asian countries. However, the practice is not well manifested in South Africa. For this reason the difference between undercount rates in female relative to male children was not pronounced in South African censuses.

As observed in the majority of prior studies, the most difficult subpopulation group to capture was young adults. The same subpopulation group consistently recorded the highest undercount rates in South African censuses. Young adults aged 24-34 from either sex always recorded the worst undercounts. Between the two, males had the higher undercount rate. The main reasons for high undercount among young adults, especially males, in South African censuses should be the same as those suggested in prior researches. Young people are more mobile than any other subpopulation group. They travel regularly searching for jobs, they relocate regularly in pursuit of better working conditions, and overall they are at the peak of their health which affords them the capacity to move. High mobility behaviours increase the risk of individuals being missed in censuses as mobility reduces the chances of being captured in a census.

However, there some undercount patterns were noticed in the studied censuses that were not consistent with findings in related studies. For instance, Census 1996 indicated a higher undercount in Indian/Asian females over males. This was not consistent with prior findings which largely concurred that males were more often missed than females. Furthermore, the same censuses indicated that the overall undercount estimate for black people who were the majority was higher than the estimates for other races who were minorities. Prior studies, particularly those in the USA confirmed that minority groups like black Americans and migrants often had higher undercount estimates than whites who are the majority.

In the same regard, Census 2011 also indicated some inconsistence with undercount patterns observed in the majority of prior studies, indicating that children under the age of 5 were one of the two age groups that had the lowest undercount rate. Findings in the majority of studies have indicated that this is one of the most difficult age groups to capture in censuses. The inconsistencies outlined above may possibly be indicative of errors in the process of estimating the undercount. The possibility of such an explanation is supported by the fact that the data on undercount estimates for black people and children under five years in coloured people is missing. These are possible indicators of bad data, which critiques of these censuses have often suggested (Moultrie and Timaeus, 2002; Moultrie and Dorrington, 2004; Moultrie and Dorrington, 2012).

In summary, there were systematic patterns of undercount that were observed in the three censuses. These were largely consistent with undercount patterns observed in censuses across the globe, suggesting that the PES and its outcomes in these censuses are reliable. However, some undercount patterns in the three censuses were inconsistent with those observed in prior studies. Such inconsistence of undercount patterns in South Africa's three last censuses against undercount patterns observed in the majority of studies suggests some inaccuracies in the data of these censuses. The findings therefore suggest that the PES has credibility, though it may not necessarily be that accurate.

10.4 Content error estimation

Adjusted counts should have minimal or no internal inconsistencies because the process of adjusting coverage error also involves correcting the counts regarding content error. The test for content error was therefore applied on adjusted data to check if the adjusted counts were clear of content error as expected. The importance of this test was that the findings would confirm whether the PES accurately adjusted the data of this type of error or not. Such findings significantly contribute to reducing controversies associated with these censuses. Also, content error can significantly confound findings in methods used in this study to investigate the presence of coverage error in the studied censuses. As a result, the first step when investigating coverage error is to control for content error. This was done by determining its magnitude before testing for the coverage error. Findings in investigations on coverage error and the discussion on their relevance to the study are covered in the next section

Findings in the UN Joint Age and Sex Score which investigated errors in age and sex reporting suggested that the level of content error in these censuses was minimal. The above findings were complemented by findings in other indices that also test internal errors when reporting on age, i.e. the Whipple, Meyers, and Bachi's indices. The findings suggest that there was insignificant content error regarding age and sex reporting in the three censuses. This confirmation was emphatic as the findings in the three censuses remained consistent.

These findings were consistent with what should be expected from adjusted data as the counts in adjusted are corrected for both content and coverage error. Hence any suggestion of the presence of content error would have been interpreted as indicating inaccuracy of the PES. Tests on content error that particularly investigate quality of age reporting are important in that age is a characteristic that affects other characteristics recorded about an individual in a census (Swanson and Siegel, 2004). Hence errors in age reporting are an immediate indicator of the quality of data with regard to content errors.

There were contradictions with regard to findings in prior studies as related studies have suggested deficiencies in both content and coverage in these censuses (Moultrie and Timaeus, 2002; Moultrie and Dorrington, 2004). A possible explanation for the difference in findings could be that this study used the indices that are already noted as restricting regarding testing for content errors that relate to age and sex. However, a content error in census data is not only confined to the two characteristics. Internal inconsistencies in census data can be noted in the reporting of other census information, e.g. events on births, deaths, and migrations. The indices used do not test such content errors.

Also these indices have associated limitations. For instance, methods used to test content errors are often based on linearity assumptions which can rarely be tested to confirm if they hold, before the methods are used. For such reasons it would be inaccurate to conclude that the findings in these indices suggested the counts from the three censuses were free of content errors. What can be said with certainty is that it is suggested in these findings that errors on age and sex reporting in these censuses are insignificant.

In conclusion, the tests carried out to investigate levels of content errors suggested that the data had insignificant levels of content error related to age and sex reporting. The tests used may not be sufficient to lead to the conclusion that the PES adjusted counts in these censuses were accurate since they may not have been that exhaustive. But on their part they have demonstrated the credibility of the PES in adjusting the respective types of content error.

10.5 Coverage error estimation

Findings in the comparison of adjusted counts against Agincourt HDSS counts indicated that the former counts were largely matched respective counts from the latter data. For instance, in 1996 comparisons, the boundaries of coinciding EAs and villages from Agincourt HDSS area generally matched. Hence, if adjusted census counts were accurate they were expected to match respective counts from Agincourt HDSS. The total counts in females were the same, though in males the difference was clear.

Stronger suggestions of the accuracy of adjusted counts came from the 2001 and 2011 matchings because, as expected, the adjusted counts for both males and females were slightly above respective counts from Agincourt HDSS data. This is what was expected as the SAs boundaries that coincided with village boundaries during the GIS mapping of these censuses counts for area covered by Agincourt HDSS, which slightly overlapped the latter. This meant counts for censuses 2001 and 2011from area covered by Agincourt HDSS would be slightly more than respective counts from Agincourt HDSS. There were however a few instances where Agincourt HDSS counts in certain age groups were more than respective counts in these censuses. Furthermore, respective counts by age groups in adjusted and Agincourt HDSS data were further confirmed credible with regard to census counts. The trends followed by counts in the two sources were consistent, but the matching was not as precise in 1996 comparisons, particularly for males.

Findings in growth rate analyses for age groups by sex, race and province also indicated that adjusted counts were credible as each analysis had only a few instances where growth rates were outside the expected range, which is between 0% and 3.5%. For instance, in the analysis of race, all other growth rates were within the expected range except for white people, which was below 0 for the inter-census period of 1996-2001. Researchers have noted that overlaps beyond this range suggest inaccuracy of census counts in the absence of excessive migrations (Moultrie, 2013; Keane et al., 1985). Indications of a negative growth rate in white people during inter-census period 1996-2001was consistent with the migration theory suggested by the Centre for Development Enterprise (1998). According to this theory, a substantial number of the white population left the country during this inter-census period. Based on this theory, the negative growth rate for white people during this period is not an understanding of counts for this subpopulation, but rather a reality. This gives credibility to the PES in adjusting the undercount in these censuses.

Findings in growth rate analyses indicate negative growth rates in the Northern Cape and North-West between 1996-2001 and 2001-2011 respectively. The negative growth rates suggest substantial losses of population in the two provinces during the respective inter-census periods. This could be due to high out-migrations from the respective provinces to other provinces. This would be in contrast with the low volumes of immigrants received by the two provinces in return. High out-migrations from these provinces may be due to less employment opportunities.

Findings in other tests for coverage error were however inconsistent with those in the first two tests. Outcomes from graphical cohort analyses, age ratio analyses and sex ratios indicated inaccuracy of adjusted data.

The order of population distributions in the three censuses depicted from graphical cohort analyses for both males and females were not consistent with what is expected from accurate counts. Due to mortality effect the population distributions in accurate census counts were expected to follow an order where the latest censuses should be at the bottom and the earliest censuses at the top. Population distributions in the country's three censuses followed a reverse order where the latest censuses were at the top – an outcome that suggested inaccurate census counts. However, such an interpretation of findings in these analyses must be treated with caution. As noted in previous researches, high immigrations may distort the order that is consistent with accurate counts (Keane et al., 1985). It is possible that South Africa has experienced substantial volumes of immigration.

A further point that complements the possibility of migration effect in distorting the expected patterns in the three censuses was that in older age groups the patterns were close to those expected in accurate counts. Such an outcome was consistent with the findings in related studies which suggested that migration behaviours decline with age. This implied less distortive effect of immigration in old age groups.

A further test of the accuracy of adjusted census counts was based on age ratios as all three censuses that investigated the patterns of age ratios in the conducted analyses were inconsistent with accurate counts. However, there are certain conditions assumed to have prevailed for age ratios to follow the expected patterns. These include assumptions that mortality and fertility have remained constant over the period covered by the censuses investigated (Kean et al., 1985). Furthermore, migration effect in the population investigated should be insignificant. These assumptions may not hold in the context of South Africa.

The deviation of age ratios from expected patterns may have been the result of distortive effects of inconsistent fertility and mortality as well as high volumes of migration. Mortality may not have been constant in South Africa mainly due to high HIV prevalence. Access to ARVs was not guaranteed to the poor who in most cases are the main victims of infections from the pandemic (Hattori et al., 2006). The role of anti-retroviral drugs issued by the government and its partners, especially after 2000, should also have led to a sudden decline in mortality, violating the assumption of constant mortality.

It is also indicated that fertility has been inconsistent. In fact, evidence gathered indicated that fertility trends have been changing. It has followed a declining trend in the past decades until around the mid-2000s and started increasing thereafter, as indicated in South African censuses. Moreover, the migration effect in South Africa's population change cannot be assumed to be minimal as indications are that there are high volumes of immigrants. These patterns of age ratios that are not expected in accurate census counts may also be due to such distortive factors.

Sex ratio patterns produced in the three censuses were largely consistent with inaccurate census counts. For instance, sex ratios at birth were often too low compared to high sex ratios of 105-107 that are often expected. The low sex ratios at birth were however consistent with findings in a study by Garenne (2004). The study's findings confirmed low sex ratios at birth in Eastern and Southern African countries. Moreover, there were some instances were findings in sex ratio analyses produced patterns that were consistent with accurate counts. For instance, there was evidence of declining trends of sex ratios in the age group 15-19 years to the oldest age group in the three censuses. The trend would however be distorted among a few age groups. The overall declining trend in sex ratios is consistent with suggestions in prior researches where findings confirmed that mortality differentials by sex favours the declining trend in sex ratios with increase in age (Keane et al., 19985) as mortality risk in males is often higher than in females after childhood.

In summary, findings in the comparison of census counts against respective counts of Agincourt HDSS largely indicated close matching. These findings are vital in the investigation of coverage errors in the three censuses because their analyses are based on no assumption and involve direct comparison of exact counts. They were complemented by findings in growth rate analyses. Such outcomes suggested the credibility of the PES in undercount adjustment in the census counts. However, findings in three other analyses indicated that the counts of these censuses had notable coverage errors.

10.6 Census undercount: To adjust or not

Determining which is better, to adjust the undercount or not, was vital in achieving the main objective of the study, i.e. to reduce controversies around the accuracy of the PES. The approach used to carry out this investigation was the comparison of adjusted and unadjusted census counts at both national and SAs levels in respective non-census counts. At SAs level counts for the area covered by Agincourt HDSS drawn from SUPERCROSS were compared with respective counts in Agincourt HDSS data. The latter were the gold standard since the data is regularly updated, and also collected at small areas level where coverage error is expected to be less compared to counts collected at national level. Analyses at national level compared census counts from the 10% samples against respective counts in MORTPAK PROJCT programme's constructed population estimates.

The argument is that adjusted counts should closely match respective non-census counts compared to respective unadjusted counts. The rationale is that the former corrected the undercount yet the latter still contain the error, hence the latter should have wider deviations from compared counts relative to the former. The main findings were that at national level, the majority of the matchings indicated that adjusted counts closely matched respective counts in constructed population estimates relative to respective unadjusted counts. There were however some isolated cases were the pattern was the opposite. Such cases probably indicated that the PES would also have errors in its adjustment process in some instances. At SAs level again, unadjusted counts generally had greater deviation from the gold standard's respective counts. On the one hand, respective adjusted counts closely approximated the gold standard's respective counts. This implied that the percentage error between unadjusted and respective Agincourt HDSS counts was higher compared to the percentage error between respective adjusted and Agincourt HDSS counts.

These findings largely suggest the credibility of the PES in estimating and adjusting the undercount in the following ways. As expected, unadjusted counts had wider deviations from the gold standard's respective counts relative to respective adjusted counts. This should be an indication that the former had the undercount error, as estimated by the PES. On the other hand, the adjusted counts that had corrected this error using the PES method closely matched the respective counts of the gold standard for most of the comparisons. However, like comparison at national level, there were also cases when adjusted counts were closer to respective counts from the gold standard compared to adjusted counts. The fact that the findings confirm that there were certain instances when PES adjusted and unadjusted counts did not confirm the expected patterns indicates some shortfalls of the PES. Although in most cases the patterns were as expected, the few inconsistencies imply that the PES cannot be prescribed as accurate, though largely indicate its credibility.

The findings were inconsistent with those in related studies. Previous studies have demonstrated the inaccuracy of PES adjusted counts, e.g. they noted the underestimation of children under 5 and young adult men in Census 1996 (Dorrington, 1999). Other studies suggested the underestimation of children under 5 and overestimation of children aged 10-19 years in Census 2001 (Dorrington et al., 2003), and overestimation of children under 5 in Census 2011 (CARe, 2012). These studies were articulate in that the PESs that were conducted in South African censuses were inaccurate, leading to the inaccuracy of both the undercount estimates and adjusted counts.

A possible explanation for the difference between this study's findings and those of related studies could be the methods that were used. Findings in related studies have been largely drawn from the comparison of ASSA models' projected counts relative to census counts. Yet findings in this study were drawn firstly from comparisons of MORTPAK PROJCT's constructed population estimates against respective counts in the three censuses. The second comparisons were for Agincourt HDSS counts relative to respective census counts in the area covered by Agincourt HDSS

The major limitation of the former findings is that the constructed population counts could not be treated as gold standard. Rather the comparison largely tested the consistence between respective counts from census and projections. There was no basis for treating the counts from constructed population estimates as gold standard, mainly because their construction was based on assumptions which the study was aware may not necessarily be met.

The strength of findings in this section is that Agincourt HDSS counts were a credible gold standard for evaluating census counts. The reasons being that; Agincourt HDSS counts are regularly updated, and the area covered is small which reduces chances for coverage errors.

10.7 Comparison of PES and DA undercount estimates

The findings in this section are crucial in achieving the ultimate goal of the study, which was to reduce controversies associated with South Africa's last three censuses. As already noted, the controversies were about the accuracy of the PES in estimating and adjusting the undercount. Among the main criticisms of the PES was that the method's undercount estimates were inaccurate, leading to inaccurate adjusted counts. The argument for this investigation is that the similarities between the PES undercount estimates and respective undercount estimates from alternative methods for estimating the undercount confirm the accuracy of the PES.

Such findings should put to rest these census controversies as this would be sufficient evidence that the PES has been accurate in estimating and adjusting the undercount. The controversies can equally be put to rest if it is found that the PES undercount estimates are completely different from respective undercount estimates obtained from alternative methods, in particular if the respective undercount estimates from alternative methods matched each other. This would equally be sufficient evidence that the PES is not accurate; hence there is a need for alternative methods that estimate and adjust the undercount in these censuses.

Out of the two alternative methods for estimating the undercount, DA was the only one feasible for the scope of this study. Regarding the other method, i.e. Bayesian model, the main limitations for its application in this study were that firstly, this method cannot be used until it obtains unadjusted counts from main countries that receive its emigrants (Redfern, 2001). This data can only be obtained after negotiation with respective countries' statistical agencies since it is rarely available for public consumption because countries often publish adjusted data. Moreover, the countries from which the data will be drawn are too many, e.g. Britain drew its emigrants from 224 countries (Redfern, 2001). The scope of the method is therefore only feasible if conducted by a country rather than an individual's research. Other complications which also make the method inapplicable to the study were discussed in Chapter 2, Section 2.3 of this study. Hence the only alternative method that was used to estimate alternative undercount estimates to those from the PES was DA.

The PES undercount estimates were obtained from StatsSA, the custodian of censuses in South Africa. Therefore the study only computed alternative undercount estimates using the DA method. The compared undercount estimates were calculated by sex and age group in the respective censuses. Respective overall undercounts estimates in males, females, and in combined sexes from the two methods were largely similar. The confirmation was strong that in other comparisons respective PES and DA undercount estimates only had a percentage variance of as little as 0.4%. These findings were inconsistent with suggestions in related studies that suggested the inaccuracy of the PES undercount estimates (Dorrington, 1999; Moultrie and Dorrington, 2012). The difference may again be explained by difference in methods of analyses that were used. This study computed alternative undercount estimates using counts drawn from a different technique for computing constructed population estimates. This is a direct approach for investigating the accuracy of the PES and the approach has rarely been use in related studies. Rather previous studies mainly used ASSA models to project population counts which were compared against respective census counts. Mismatch between the counts in respective sources were mostly used as the basis for suggesting the inaccuracy of the PES.

There were also some instances where findings in some analyses indicated differences between respective PES and DA undercount estimates. In particular, the undercount estimates for age groups were not as explicit as those for subpopulation groups described above in confirming similarities between respective PES and DA undercount estimates. They were certain age groups that indicated clear differences between respective undercount estimates from the PES and DA. There were also age groups whose respective undercount estimates in the two methods matched closely. In the majority of age groups the deviation between respective undercount estimates in the two methods was moderate.

In the case of inconsistence between respective undercount estimates in the two methods, there were uncertainties as to which method's estimates were credible. This is because the tests were based on the consistence of undercount estimates in the two methods rather than on e.g. treating the estimates from DA as the gold standard. Overall, the findings could not confirm that the PES was accurate since there were some cases where undercount estimates in the two methods were inconsistent. However, since most of the respective undercount estimates in the two methods suggested consistence with each other, the PES can be confirmed to be largely credible.

In relation to other findings in this study, the findings in this investigation were largely complementary. Findings from comparison of undercount patterns in the country's three censuses and those in censuses across the globe largely indicated similarities. The test for content errors related to age and sex that were conducted suggested that the PES was credible in adjusting these errors as various findings suggested insignificant presence of these errors. Findings in tests for coverage errors also indicted that overall the PES was credible, despite certain instances where questionable findings were noted. The findings in a test on which was better, to adjust or not adjust the undercount in the country's last three censuses, were also consistent with these other findings in the study. This is because as expected, adjusted counts had slight deviations from respective non-censuses counts to which they were compared, relative to unadjusted data which were expected to be affected by the undercount.

10.8 Undercount effect on demographic parameters

The following two sections are important in that they quantified negative outcomes associated with census undercount. The context of these controversies is that census counts in South Africa are used for estimating demographic parameters that guide planning. They are also used for ensuring fair allocation of resources and efficient service delivery. The effect of undercount on demographic parameters and service delivery was achieved by comparing demographic rates, the Basic Services Grant Fund and parliamentary seat allocations in respective adjusted and unadjusted counts. The effect of undercount on demographic parameters is measured by determining the difference between estimates in unadjusted counts which carry the undercount error relative to estimates in respective adjusted counts.

The study first examined the accuracy of the reconstructed unadjusted counts before the analyses, and the results were recorded in Chapter 9. The published census data for 1996 had both adjusted and unadjusted counts for provinces and this presented an opportunity to compare reconstructed unadjusted counts with official unadjusted counts. As expected in the study, the reconstructed unadjusted counts in Census 1996 closely matched the official unadjusted counts of StatsSA. This implied that the reconstructed unadjusted counts used in this study were credible to stand in for official unadjusted counts which were not released by StatsSA in the case of the other two censuses.

The main finding was that undercount could have compromised the accuracy of demographic parameters by either inflating or deflating them. For instance, undercount could have compromised crude birth rate, total fertility rate and general fertility rate for censuses 1996 and 2011 by deflating the rates. Yet in Census 2001, undercount could have inflated the same rates. These findings were consistent with those in prior studies. For instance, such studies had noted that undercount often varies by subpopulation group. This implies that an undercount of women in the reproductive age group can inflate crude death rates, age-specific rates, total fertility rates and general fertility rates if this is not reciprocated by the same error in birth counts (Rindfuss, 1974). The undercount of children born, relative to completeness in coverage of women of reproductive age deflates the fertility rates.

10.9 The effect of undercount on service delivery

Service delivery was measured in this study using the Basic Service Grant Funds (BSGF) and parliamentary seats allocations. To estimate the effect of undercount on service delivery, allocations of BSGF and parliamentary seats in the country's nine provinces based on their respective adjusted and unadjusted counts were computed. The selection of these proxies was consistent with what was observed in prior studies. In particular, the studies conducted in the USA have used the same proxies for measuring service delivery.

There are seven main grants under the Provincial Equitable Share Fund (PESF), whose funds are distributed to provinces but using various formulas. The grants whose funds are allocated using a formula that partly includes census counts are: the Education Grant, Health Grant, and the Basic Services Grant. Among the three grants it is only the latter grant whose funds are allocated entirely using census counts. Hence this was the only grant relevant for this investigation. The first PESF allocations coinciding with censuses investigated in this study were provided in 1999. Since then, there have been two reviews of the amounts for each grant in 2004 and 2013 and both reviews have been upwards. The biggest allocations have been consistently given to Education, Health and BSGF respectively.

The findings in investigations on BSGF allocations were that undercount was associated with unjustified resource allocations. For instance, results based on allocations using counts from Census 1996 indicated that the three provinces would have been prejudiced by undercount, yet the other six would have gained unfairly at the expense of the three provinces. This would have meant compromised service delivery among provinces that would have been prejudiced by undercount. The prejudice suffered by undercounting provinces would worsen if yearly losses were projected over respective periods until the next review. Yet provinces that benefited unfairly from the undercount would have continued to benefit unfairly until the next review.

The study’s findings complemented the findings in prior related studies that were conducted in the USA. Undercount compromised service delivery in states like California (Steffey, 1997). The Consistence of findings may be partly explained by the similarity in method used to estimate the undercount. Both studies used PRICEWATERHOUSECOOPERS and the same formula, i.e. both adjusted and unadjusted census counts to partly measure the effect of undercount on service delivery.

To estimate the effect of undercount on service delivery through parliamentary seat allocations, the study made an assumption that all persons legible to vote in South Africa had registered to vote by the date of each respective election. As for parliamentary seat allocations, if all individuals aged 18 years and above, as obtained from respective censuses, were to register as voters, seat allocations from unadjusted counts would be less compared to seat allocations based on adjusted counts, for each province. This is expected since adjusted counts were higher than unadjusted counts after the former undercount was corrected.

However, the effect of undercount on parliamentary seat allocations drawn from adjusted and unadjusted counts would not follow the same patterns as above if allocations were based on proportions of populations in provinces. As in the allocation of funds from the BSG, other provinces would have been prejudiced regarding parliamentary seats, while others would have gained at the expense of those prejudiced. These findings are also consistent with those obtained in related studies conducted in the USA, as referenced above.

10.10 The study findings in relation to the discourse of census undercount

This study is nested in the discourse of census controversies in South Africa. The controversies were triggered by high undercount estimates recorded in the country's last three censuses. At the centre of the controversies is the PES whose accuracy in determining and adjusting the undercount has been contested.

In relation to the discourse of census controversies in South African censuses, the study findings largely make a paradigm shift from the perception that exists regarding these censuses. The paradigm shift is not that the study's findings confirmed the accuracy of the

PES. In that regard none of the findings in all analyses conducted in this study were sufficient enough to confirm this. Rather, the paradigm shift is that the various findings in the study concurred that to a large extent the PES and its adjustments were credible and hence reliable. This is how the findings in this study contribute to reducing the controversies around the accuracy of the PES in estimating and adjusting the undercount in the three censuses.

CHAPTER 11: CONCLUSIONS AND RECOMMENDATIONS

11.1 Conclusions

The study's research question was: How the controversies associated with South African censuses can be reduced? The controversies are centred round the accuracy of the PES in testing and adjusting the undercount. This study had five specific objectives whose findings addressed this research question. Findings in the analyses for specific objectives were not sufficient to confirm that the PES and its adjustment outcomes were exactly accurate, but they provided reasonable evidence that suggested that the method and its outcomes were credible. The findings confirming the accuracy of the PES and its adjustment outcomes would have put these controversies to rest. Sufficient evidence indicating the credibility of the method and its outcomes reduces controversies in these censuses because the evidence indicates that these censuses are reliable.

The study's findings indicated the credibility of the PES and its adjustment outcomes, as explained below. The exploration of undercount patterns in related studies across the globe observed that children under the age of 5 years, young adults, males relative to females, and minority groups often have high undercount rates. On the other hand, the elderly population had a lower undercount. South Africa's last three censuses had systematic undercount patterns which largely confirmed those noted in related studies.

The findings in tests for content error suggested that the three censuses were largely free from content error that is related to age and sex misreporting. The indices used can only test the two types of content error. The findings were consistent with what is expected from adjusted counts, as the latter will have been corrected for such types of content error. The main findings in tests for coverage error also indicated that the three census counts were reliable as comparison of census counts in the area covered by Agincourt HDSS with respective counts in the surveillance site generally matched. This was complemented by findings regarding growth rate analysis. However, findings in the other three analyses suggested the presence of some coverage errors in the counts. A possible explanation for this was that the distorted trends and patterns observed in these findings may also have been a result of other distortive factors, like excess migrations and inconsistent mortality and fertility.

The findings on whether to adjust census undercount or not largely conformed to what was expected. The latter counts closely matched respective counts from non-census sources. Yet respective unadjusted counts in most cases widely deviated from respective counts from the same non-census sources. This was a strong indication of the credibility of the PES in estimating and adjusting the undercount in the three censuses as the wider deviations of unadjusted counts from respective counts they were matched against confirmed the reliability of the PES in estimating the undercount. The deviations reflected the undercount error in unadjusted counts as opposed to compared counts. The adjusted counts' close matching with respective counts from constructed population estimates and Agincourt HDSS data indicated the credibility of the PES in adjusting the undercount in these censuses. For these counts were corrected for the error, and hence were expected to closely match respective counts to which they were compared.

The outcomes from the comparison of the PES and DA undercount estimates also suggested the credibility of the PES and its adjustment outcomes. The findings were that the respective overall undercount estimates for males, females and for combined sexes in the two methods were largely similar. The respective undercount estimates for age groups in the two methods did not consistently maintain as close a match as the former. There were a few cases where the undercount estimates in the two methods clearly contradicted each other and there were also a few cases where they matched closely. Regarding the majority of age groups, the variations were slightly noticeable.

The findings in these analyses were central to reducing controversies round the accuracy of the PES and its outcomes as undercount estimates in the investigated censuses are direct outcomes of the PES, and their contradictions with undercount estimates in other methods would hinder any suggestion of the credibility of the PES. The findings in most of the subpopulation groups whose undercount estimates were investigated indicated that the respective outcomes in both the PES and DA matched closely. In particular, this was observed in the overall undercount estimates of males, females, and both sexes. Among the undercount estimates that could not match these were those in the comparisons of age groups.

Tests on the effect of undercount on both demographic parameters and service delivery indicated that the error leads to inaccuracy of the former and also to unfair distribution of resources in the latter case. As the other findings in the study have suggested that the PES and its outcomes are credible, these findings confirmed that it was necessary to adjust the undercount in the last three censuses. For example, failure to adjust the undercount would have prejudiced certain provinces regarding deserved funds earmarked for development while unfairly benefiting other provinces. Failure to adjust the undercount would have also compromised efficient planning in instances where demographic parameters in census data were relied upon. These findings reduce controversies about these censuses in that the PES adjustments avoided unfair resource allocation, and the reliance of planners on inaccurate demographic parameters.

11.2 Recommendations

Based on the study's findings recommendations are that firstly, studies on South African censuses should direct more effort in investigating whether alternative methods for testing for undercount would yield better results. This study was the first to make an effort in that direction. Although this study can be commended for having initiated research in this direction its findings may not have exhausted the issue. There is need for more researches on this issue in order to verify the findings from this study.

Further suggestions from some of the study's findings were that the PES may also be ushering some errors in South Africa's census data during the process of estimating and adjusting for the undercount. Such findings seem to complement those from prior studies which had suggested that the PES sample size used in South Africa's last three censuses was too small and likely to introduce errors. Recommendation is that StatsSA should consider increasing the PES sample size to test if this improves quality of future censuses.

Other researches particularly that are qualitative in approach are also recommended on the following issues that may turn out to be effective in improving quality of the country's censuses. Firstly, the marketing of the country's census brand by relevant authorities. StatsSA may have already initiated this, but what may matter are the strategies employed and message conveyed. Such researches may seek to verify the public's perception, understanding and knowledge of purpose censuses. It may be undocumented, but information gathered from street talk revealed that some people see no benefit in putting aside activities that directly benefit their families for the sake of participating in a census.

Researches can also prioritise on investigating on the effect of providing feedback of census outcomes to member of the public by census authorities through various platforms that reach out to all the subpopulation groups. In most cases, workshops and conferences are the only platforms used for this. Such channels of information dissemination just reach out to certain sections of the population, especially those who are invited to attend. Such researches may further investigate on people's perceptions and preferences on methods that can be used for dissemination census outcome to the general public.

Research that also investigates effect of qualifications of census enumerators is also encouraged. Enumerators used during the country's censuses are often unqualified and unemployed individuals, and their capacity to perform as expected may be questionable. Such enumerators have greater potential of contributing to the poor quality of data that is collected. In most countries in Sub-Saharan Africa, enumerators are selected from professionals in the civil service, like teachers and police officers. The logistics with regard to ensuring efficient census enumeration are complex and the use of such professionals, e.g. those whose communication skills have been tested, largely contribute to quality census data. Zimbabwe uses such professionals in census taking and the quality of their census data is said to be of a good standard. However, South Africa uses individuals who are unemployed as enumerators. This is rational in that census taking becomes a source of employment for the unemployed. However, this is often at the expense of obtaining quality data.

Further researches may also be carried out to determine the effect census authorities working together with gate keepers of respective local communities on improving success of census taking. Such studies may be target areas that often record high undercount rates. Gate keepers often have influence over their communities. Such studies may further investigate on strategizes that census authorities can use for building cordial relations with the gatekeepers.

Other recommendations that can improve quality of censuses that was gathered from review of literature on census taking include reducing the length of questionnaires used in South African. Current questionnaires are too long, and they can be shortened by e.g. including only basic questions. The country conducts annual national surveys and these can be targeted on complementing census data by packing them with more questions. Long questionnaires often result in unwillingness to participate, and the increasing number of people requesting to keep questionnaires to complete and send later at a convenient time partly confirms this. Even those participating may end up forging information just for the sake of completing the form.

Furthermore census authorities are encouraged to build cordial relations with members of the public. There have been suggestions that census authorities can achieve this by distributing their officers to attend various local public gatherings like weddings, funerals and religious events. This can be complemented by media coverage, so that the message about such cordial relations between census authorities and communities spreads widely. This would encourage more people to associate with events like census taking partly as a way of showing solidarity with census authorities who participate at community events.

Census authorities in South Africa are also encouraged to seek extra funding for census taking. The government is largely the sole sponsor of South Africa's censuses, and the funds allocated are often inadequate as there are other pressing issues calling for the government's immediate attention. Censuses are known to be highly expensive to conduct, especially when there is need to carry out extra pre-census activities like marketing of the census brand. However, well-funded censuses often meet expected standards and outcomes. Additional funding can be sourced from both the private sector as well as from non-governmental organisations that are known to often fund censuses.

11.3 Overall study limitations

Findings in objective 4 may be anticipated to be compromised by the unavailability of unadjusted data, where a proxy data that was reconstructed was used. However, this limitation is not expected to have any significant effect on the study's thrust as the idea with this objective was to just get an estimate of how undercounted data can produce varying outcomes to those from adjusted data, as an alert call to census authorities.

Furthermore, the fact that this thesis was entirely based on the use of secondary data limited its flexibility as the researcher was not be able to capture key information from the original source. Some of the key information that is vital for this study may not have been collected, e.g. because relevant questions that could have been asked when gathering such information were not included in the questionnaire. However, since the thrust of this research was not so much based on the nature and quality of data from specific variables, this limitation became very trivial. The research was rather concerned with examining the accuracy of population counts in South African censuses.

REFERENCES

Anderson, B, A. (2004): “Undercount in China’s 2000 census in comparative perspective.” PSC Research Report, No. 04-565 University of Michigan 1-14

Australian Bureau of Statistics (2013): “Census of Population and Housing” http://www.abs.gov.au/websitedbs/censushome.nsf/home/Census?opendocument&ref=topBar 15-03-2015

Banda, P, J., Navarro, A., Singh, R, P and Whitford, D, C. (2013): “Post Enumeration Surveys in Africa.” AMSTATNEWS http://magazine.amstat.org/blog/2013/03/01/post-enumeration/ 20-07-2014

BBC News Wales (2011): “Census 2011. Cardiff council fears 22 000 Undercount.” http://www.bbc.com/news/uk-wales-12868199 11-09-204

Berkowitz, P (2012): “The (incomplete) (probably inaccurate) sum of us.” Daily Maverick http://www.dailymaverick.co.za/article/2012-10-31-census-2011-the-incomplete-probably-inaccurate-sum-of-us#.VvreU2Pd6dk 17-09-2013

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

Brown, L, D,. Eaton, M, L,. Freedman, D, A,. Klein, S, P,. Olshen, R, A,. Watcher, K, W,. Wells, M, T and Ylvisaker, D. (1999): “Statistical controversies in census 2000.” Jurimetrics, Vol. 39, no. 4, 347-375.

Census, 1865-G20/66 (Cape Town, Government Printer, 1866) Cape of Good Hope House of Assembly, Colonial Office, 1866

Centre for Disease Control and Prevention, (1999): “Quality of death rates by race and Hispanic origin: A summary of current research, 1999.” Vital and Health Statistics, Series 2. No. 18, 1-13

Central Statistical Service (1991): “Population census 1991. Adjustment for Undercount.” No. 03-01-26 (1991) ‘N SSD-VERSLAG. A CSS REPORT

City Press Weekly Newspaper (2013): “2011 census row grows.” www.citypress.co.za/news/2011.census-row-grows, 04-08-2015

Chuks, J, M. (1996) “Assessing the reliability of the 1986 and 1996 Lesotho census data” JOURNAL OF SOCIAL DEVELOPMENT IN AFRICA, Vol 18 No 1, 111-128

Courbage, Y. (2000): “South Africa and its populations.” *POPULATION ET SOCIETIES* (354): 1-4

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

Dahiru, T and Dikko H, G. (2013): "Digit preference in Nigerian censuses data of 1991 and 2006." Epidemiology Biostatistics and Public Health Volume 10, Number; 8843-1-8843-5

De Wet, P (2012) **"**Is Census 2011 accurate? Still depends on who you ask" Mail & Guardian http://mg.co.za/article/2012-11-01-is-census-2011-accurate-still-depends-who-you-ask 28-09-2015

Dorrington, R., Moultrie, T, A and Timaeus, I, M. (2004): "Estimation of Mortality using the South African census 2001 data. Centre for Actuarial Research, Monograph No.11

Dorrington, R., Budlender, D and Moultrie, T. (2003): "Census 2001 Group 3 Report: Are the numbers what we expect? Statistics Council, Pretoria

Dorrington, R (2002): "Did they jump or were they pushed? An investigation into the apparent undercount of whites in 1996 South African census." SA Journal of Demography, 8 (1); 37-46

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

Ewbank, D, C. (1981): "Misreporting Age-Selective Remunerations: Sources, Patterns, and Consequences for Demographic Analysis." Committee on Population and Demography, Report No. 4; National Academy Press.

Garenne, M. (2004): "Sex ratios at birth in populations of Eastern and Southern Africa." Southern African Journal of Demography 9 (1): 91-96

Gernertzky, K (2012): "Census: Undercount still double digit." Mail & Guardian

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

IUSSP Tools for Demographic Estimation http://demographicestimation.iussp.org/, 08-06 2015

Johnson, J, E. (2012): "Estimates and Implications of U.S. census undercount of native-born population." Harris School of Public Health, University of Chicago; 1-40

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

Lalasz R (2006) "In the News: Nigerian Census." http://www.prb.org/Publications/Articles/2006/inNewsTheNigerianCensus.aspx, 15-05-2015

Libreton, J.D. (1992): "Modelling survival and testing biological hypothesis using market animals: A unified approach with case studies." Ecological Monographs 62; 67-118

Lemieux, O, A. “Problems in Census Taking” The Canadian Journal of Economics and Political Science Vol. 14, No. 4; 469-280

Liang, Z and Ma, Z. (2004): “China’s floating population: New evidence from the 2000 census.” *Population and Development Review* 30 (3); 467-488

MacLennan, W. (1996): “Information paper: Census of [population and housing: Data quality-undercount.” Australian Bureau of Statistics; 1-29

Mail & Guardian Weekly (2013): “Census 2011 irregularities; Top officials pressured.” http://mg.co.za/article/2013-02-02-census-2011-irregularities-staff-were-bullied 26-06-2013

Moultrie, T, A and Timaeus, I, M (2002): “An Analyses of the 1996 Census and 1998 Demographic and Health Survey.” Burden of Disease Research Unit, Medical Research Council, Cape Town

Moultrie, T and Dorrington R. (2012): “The census: Some more questions.” http://www.politicsweb.co.za/news-and-analysis/the-census-some-more-questions 27-05-2013

Moultrie, T (2013): “General assessment of age and sex data” Tools for Demographic Estimation; UNFPA; 3-21

Murdock, S, H and Hoque, M, N. (1995): “The effect of undercount on accuracy of small-area population estimates: Implications for the use of administrative data for improving population enumeration.” Population Research and Policy Review, Vol. 14, No.2; 251-271

Nyirenda, M, Hosegood, V and Moultrie, T, A (2007): “Population modelling for a small area: a comparative analysis of census and demographic surveillance system data in South Africa [Private citation]

Office of National Statistics UK, 2011) http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/about-censuses/methods-of-census-taking/index.html, 17-07- 2014

O’ Hare, W, P. (2009): “Why are young children missed so often in the census?” The Annie E. Casey Foundation, Kids count working paper; 1-15

Onsembe, J, O and Ntozi, J, P, M. (2006): “The 2000 round of censuses in Africa: Achievements and challenges.” *The African Statistical Journal*, Vol 3; 11-28

Palamuleni, M, E. (1995): “Age misreporting in Malawian censuses and sample surveys: An application of the United Nations joint age sex score.” *Southern African Journal of Demography* 5 (1); 11-17

Patton, (2003): “Utilization-Focus based Evaluation.” *Kluwer International Handbooks of Education*, Volume 9; 223-242

PRICEWATERHOUSECOOPERS (2000): "Effect of census 2000 undercount on federal funding to states and selected counties, 2002-2012." US Census monitoring Board; 114-174

Raley, R, K (2002): "The Effects of the differential undercount on Survey estimates of race differences in marriage." Journal of Marriage and Family, 64; 774-779.

Rossouw, L., Burger, R and Burger, R. (2012): "The fertility transition in South Africa: A retrospective panel data analysis" Working paper of the Bureau for Economic Research at the University of Stellenbosch, Matieland

Rowland, D, T (2012): "Demographic methods and concepts." Oxford University Press, New York

Schirm, A, L. (1991): "The effects of census undercount adjustment on congregational apportionment." The Journal of the American Statistical Association, Vol. 86, No. 414; 526-541.

Schultz, D (2013): "Was census 2011 successful?" Mail & Guardian http://mg.co.za/article/2013-03-22-was-census-2011-sucessful 21-08-2013

Siegel, J, S and Swanson, D, A (2004): "The methods and Materials of Demography." Second Edition, *Genus* 60 (2)

South Africa National Treasury ***Intergovernmental*** Fiscal Reviews (IGFR, 2014): http://www.treasury.gov.za/publications/igfr/, 16-08-2014

Statistics Canada (2006) http://www12.statcan.gc.ca/census-recensement/2006/ref/dict/overview-apercu/po7-eng.cfm, 13-04-2015

Orkin, F, M. (1998): "THE PEOPLE OF SOUTH AFRICA: Population census, 1996: Calculating the undercount in census' 96" http://www.statssa.gov.za, 12-03-2014

StatsSA, 1998: "The count and how it was done."**http://www.statssa.gov.za/census01/census96/html/metadata/docs/count/chapter_5.htm**, 12-03-2014

Census 2001: "Post Enumeration Survey: Results and Methodology" Statistics South Africa Report No. 03-02-2014

Census 2011: "Census undercount and Strategies: Version No 6." Statistics South Africa.https://www.statssa.gov.za/census2011/documents/Census_2011_UndercountStrategies-version_6a.pdf , 12-03-2014

South African News Agency (2012): "Census undercount of 14.5% a concern." SA news.gov.za http://www.sanews.gov.za/south-africa/census-undercount-146-concern 11-04-2014

Steffey, D, l. (1997): "A Review of census undercount issue." Department of Mathematical and Computer Sciences, San Diego State University; 1-25

TNA Reporter (2012): "Big errors claimed in census" The New Age www.thenewage.co.za

Think Africa Press (2011): "The Impact of service delivery on South Africa's Local Government Elections." http://thinkafricapress.com/south-africa/impact-service-delivery-south-africa%E2%80%99s-local-government-elections-0

US Bureau of Census (1992): "Assessment of accuracy of adjusted versus unadjusted 1992 census base for use in intercensal estimates." Report of the committee on adjustment of post censal estimates: Washington DC, Bureau of Statistics.

Vijayan, K, P and Barton, T, R. (1992): "Marketing the census: Toward a social strategy." International Review of modern sociology Vol.22; 55-64

Vincent, R. (1992): "The use of administrative data to improve survey estimates: Proceedings of the US Bureau of the census annual research conference." Washington DC: US Bureau of Census; 597-606.

Appendix A: Proof Reader's certificate

25/03/2016

To whom it may concern

I hereby confirm that I edited Jeremy Dickson Gumbo's doctoral thesis titled: 'Undercount Controversies in South African Censuses'.

Mr Aré van Schalkwyk
112 Hof Park
28 Hof Street
Gardens 8001

Tel: 082 898 7218

Regards

Aré van Schalkwyk

Appendix B: Turnitin Report

UndercountingcontroversiesinSouthAfricanCensusFinal.doc

ORIGINALITY REPORT

5%	3%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.un.org Internet Source	1%
2	www.statssa.gov.za Internet Source	1%
3	www.ijcm.org.in Internet Source	<1%
4	Submitted to Cranfield University Student Paper	<1%
5	Submitted to University of Witwatersrand Student Paper	<1%
6	Submitted to University of Leeds Student Paper	<1%
7	Akbarzadeh, V., and A. N. Hrymak. "Coupled fluid-particle modeling of a slot die coating system : Particle Technology and Fluidization", AIChE Journal, 2016. Publication	<1%
8	www.statsghana.gov.gh Internet Source	<1%
9	Submitted to Massey University	

Appendix C: Literature Matrix

	Title	Author/s & year	Data Source	Method	Level of Analyses	Findings	Missing gap
1	Undercount in China's 2000 census in comparative perspective	Anderson, B, A. (2004).	Census	Comparison of counts	Individual	Patterns of undercount in China were consistent with those from other studies	Sources of undercount not addressed
2	The census adjustment: Undercount or bad data	Breiman, L (1994)	Census	Compared undercount estimates from PES & Dual Systems Estimate	Individual	There were some errors in PES undercount estimate	Does not address undercount estimates at macro level
3	Statistical controversies in census 2000	Brown, L, D et al (1999)	Desk reviews	Evaluations of PES in estimating undercount and legal implications of adjusting for undercount	National	Accuracy of PES is determined by a range of errors that it is susceptible to and also to arbitrary decisions	There isn't evaluation of alternative methods of undercount's susceptibility to same challenges
4	Assessing the reliability of the 1986 and 1996 Lesotho census data	Chuks, J, M (1996)	Census	Computing various Indices	Macro	Data is fairly free of content error those there are also indications of misreporting of age and sex	Paper did not address other types of content error e.g. birth misreporting
5	South Africa and its populations	Courbage, Y. (2000)	Census	Review of all censuses in South Africa by time of the study	Micro and macro	High undercount estimates have been consistently recorded in the country's censuses	Accuracy of PES in estimating undercounting in these censuses not determined
6	Comparing census 1996 a	Cronje, M	Census	Compare	National	There are strength	No determination of

	with 2001: An operational perspective	and Budlender, D. (2004)		procedures in censuses 1996 and 2001		weaknesses in changes of procedures of census taking for the two censuses	which census' outcomes predicted better the respective populations of the country
7	Did they jump or were they pushed? An investigation into the apparent undercount of whites in 1996 South African census.	Dorrington, R (2002)	Census and Population Projections	Comparisons of respective counts and Hill's Indirect technique of estimating migration flows	Micro and macro	PES was inaccurate in estimating white population. Migration could only account for 250 000 whites out of 800 000	No investigations on which counts between PES adjusted and those from projection were reliable
8	To count or to model that is not the question: Some possible deficiencies with 1996 census results.	Dorrington, R. (1999)	Census	Matching of PES adjusted and Projected counts	Micro	Adjusted counts underestimated e.g. children under 5 years	No investigation on accuracy of the PES undercount estimates that led to these adjusted counts by comparing them with undercount estimates from other methods
9	Misreporting Age-Selective Remunerations: Sources, Patterns, and Consequences for Demographic Analysis.	Ewbank, D, C. (1981)	Censuses	Indices for detecting content error, PES, Brass growth balance technique	Individual		No test for coverage errors
10	Sex ratios at birth in populations of Eastern and Southern Africa	Garenne, M. (2004).	DHS	Chi square tests	Individual	Eastern and Southern African region has low sex ratios relative to other regions	No determination on whether sex ratios from DHS are the same with respective sex ratios from census data.

11	China's floating population: New evidence from the 2000 census.	Liang, Z and Ma, Z. (2004)	Census	Cross tabulations	Micro	Floating population in China is on the increase	Study is only descriptive based and no inferential analyses.
12	The census: Some more questions	Moultrie, T and Dorrington R. (2012)					
13	General assessment of age and sex data	Moultrie, T. (2013).	Census, and United Nations World Population Prospects	Evaluations on single and multiple census/s analyses, and PES	Micro	The various methods evaluated have strengths and limitations that eventual determine their accuracy	Other methods of estimating undercounting like DSE, and DA were not evaluated
14	Why are young children missed so often in the census	O' Hare, W, P. (2009)	Census, and other various source of data on children e.g. hospital records,	Evaluation of sources of children data and reasons why children are missed	Micro	Reasons for missing out on children many and also vary with sub population groups.	No investigations on implications of children undercount
15	The 2000 round of censuses in Africa: Achievements and challenges."	Onsembe, J, O and Ntozi, J, P, M. (2006)	Census	National	macro	Implementation of second round of censuses were generally successful, though many countries had challenges	No evaluations on what should be done to ensure successful implementation of censes in Sub Sahara region in the 3rd round
16	Age misreporting in Malawian censuses and	Palamuleni, M, E.	Census and DHS	UN Joint age and sex index	Micro	Age reporting has improved in Malawi,	The study does not address other types

	sample surveys: An application of the United Nations joint age sex score	(1995)				though there are indications that significant errors are still recorded	of content errors besides age misreporting
17	The effects of census undercount adjustment on congregational apportionment	Schirm, A, L. (1991)	censuses	Regression analyses	macro	Adjustment for the undercount lead to reallocation of number of parliamentary seats among states	The study did not determine which counts between adjusted and unadjusted were more reliable to use for allocating parliamentary seats
18	A Review of census undercount issue	Steffey, D, l. (1997)	Census	Follows the PRICEWATER HOUSECOOPERS Method's procedures	macro	Failure to adjust would have prejudiced certain states like Californian, of their deserved Grants' funds	The study did not determine which counts between adjusted and unadjusted would be more appropriate for allocating funds from national grants fund to provinces

Appendix D: Publication Plan

Title	**Thesis Chapter & title**	**abstract**	**Status**	**Journal**	Authors	submit
Implications of Undercounting on Service Delivery in Sub Sahara Africa	Chapter 9 Undercounting effect on service delivery	*Paper used the PRICEWATERHOUSECOOPERS method to investigate effect of undercounting on allocations of funds from national grants and parliamentary seats among the country's provinces*	*Reviewers' comments were addressed and the corrected version resubmitted together with a cope of how comments were addressed.*	Southern African Journal of Demography	Clifford Odimegwu and Jeremy Gumbo	June 2015
Coverage evaluation of South Africa's census 2011	Chapter 6 Estimation of coverage error	*Paper tested for completeness of South Africa's last three censuses.*	*Submitted to journal and awaiting reviewers comments*	Southern African Journal of Demography	Jeremy Gumbo and Clifford Odimegwu	Sept 2015
Undercount Patterns in South African censuses: Systematic or Inconsistent?	Chapter 2	*The paper investigated if: undercount patterns in South Africa's last three censuses were systematic and matching with those observed from related studies.*	*Submitted to journal and awaiting reviewers comments*	African Population studies Journal (APS)	Jeremy Gumbo, Garikai Chemhaka ,Pedzisai Ndagurwa ,Nyasha Chadoka, Emmanuel Olaijuman and Baruwa	April 2016

					Ololade	
Undercounting controversies in South African censuses	Chapters 6 and 7	*The paper compared PES undercount estimates against respective estimates from other methods. Furthermore it compared unadjusted and adjusted census counts for area covered by the surveillance site against respective counts from Agincourt HDSS.*	*The paper is expected to be complete by end of August*	Expected journal for submission: Genus Italia	Jeremy Gumbo and Clifford Odimegwu	Expected to be submitted for journal review by end of August 2016
Other papers still to come from chapters 5 and 8	Chapter 5 Estimations of levels of content error					TBA
	Chapter 8 Undercount effect on demographic parameters					TBA

Appendix E: Policy brief

Undercounting Controversies in South African Censuses

Background

Censuses are vital tools for directing policy formulation that lead to fair and efficient resource allocations in a given country. Therefore an accurate count is vital in census taking. As a result governments often allocate huge budgets for planning and implementations of censuses (Cronje and Budlender, 2004). This is because the benefits from an accurate census are believed to outweigh the costs (Cronje and Budlender, 2004; Steffey, 1997). However a major challenge in achieving accurate census counts is that the process is susceptible to none sampling errors of content and coverage. The latter is more common than the former and can either be in form of undercounting or over counting. South Africa's last three censuses recorded high undercount estimates i.e. 10.6%, 17%, and 14.6% for censuses 1996, 2001, and 2011 respectively.

Problem Statement

High undercount in the country's last three censuses led to contests on the reliability of processes and outcomes of these censuses. In particular some researchers contested the accuracy of the Post Enumeration Survey (PES) in estimating and adjusting for the undercount. For instance, they argued that the sample sizes used to conduct the PES were too small and may have introduced further bias in counts (Moultrie and Dorrington, 2012). For example, they noted that the censuses underestimated children below the age of years, young male adults, white population, while overestimating adult women (Dorrington, 1998, Dorrington, 2002). Conversely, other researchers noted that counts from these censuses are better estimates of the country's population than outcomes from models that have been proved incorrect (De Wet, 2012). Such controversies negatively affect the general public's trust on policies, service delivery plans, and research outcomes that are based on information drawn from these censuses. For, inaccurate census data leads to; research findings that are not credible, and policies that are inefficient and inaccurate. This may lead to civil unrest as members of the public protest against unfair and inefficient resource provision. Besides, inaccurate census data also affect quality of surveys that rely on this data to draw their sampling frames.

Findings

Main finding was that; data from South Africa's last three censuses may not be precisely accurate, but were credible. This is because; undercount patterns drawn from the three censuses were systematic and largely consistent with those observed from other censuses across the globe. Second, indications were that data from the country's three censuses had less content and coverage errors. Third, the adjusted census counts closely matched respective counts from none census data. Fourth, comparison of both adjusted and unadjusted census counts against respective counts from none census data indicated that; adjusted counts were closer to none censuses counts relative to unadjusted counts. Fifth, PES undercount estimates that were used to adjust for coverage error were largely similar to respective undercount estimates drawn from other methods. Finally, the findings also indicated that undercounting compromised demographic parameters and service delivery.

Policy Implementation and Implications

Study's findings suggested that the implementation of policies that relate to achieving quality of census data were reasonably successful. This is because the study found that quality of data from the investigated censuses was credible. The policy implication is that; it is better to sustain the same policy direction in ensuring quality census data in future rather than instituting a policy shift. This should be complemented by identifying and rectifying possible loopholes in the current policies to ensure further reduction of errors in the country's future censuses. The latter, which implies an overhaul of such policies, may have been an ideal option if the study's findings suggested largely inaccurate data from the three censuses.

For policies relating to poverty alleviation like the National Development Plan: Vision 2030, the study's findings indicated that data from the country's last three censuses are reliable in informing planning and implementation of such policies. Policy implications are that: the monitoring and evaluation of progress made in achieving goals set in such policies is credible.

Potential users of Thesis findings

StatsSA are among main users of the study's findings. For instance, they are mandated by the government of South Africa to provide indicators that track progress in poverty reduction. Census data is a vital source of such information because it is collected from national down to the lowest level. StatsSA can therefore use the study's finding to justify credibility of the respective indicators they draw from the country's census data. Furthermore, they are the custodians of censuses in South Africa, and hence they can use the findings as a spring board to strategise for better census outcomes in future.

The government is the main stakeholder in addressing needs of the South African population. They are expected to fairly and efficiently allocate resources across the country's sub-population groups. This mandate can be effectively accomplished if guided by information drawn from credible data.

Researchers and research institutes are also potential users of the study's findings. This is because validity of their researches is primarily based on quality of data they used to arrive at their findings. Data from the country's last three censuses were found to be reliable.

Appendix F: Revised Manuscript submitted SAJ of Demography

Service delivery implications of census undercount in Sub Sahara Africa

Authors:

Clifford Odimegwu

Demography & Population Studies Programme, Schools of Public Health and Social Sciences, University of The Witwatersrand

*Jeremy Gumbo

RMPRU, Chris Hani Baragwaneth Hospital, Johannesburg, South Africa

Demography and Population Studies Programme, Schools of Public Health and Social Sciences, University of Witwatersrand, Johannesburg, South Africa

*Corresponding author
jeremy.d.gumbo@gmail.com

Abstract

Studies carried out in United States of America have confirmed that census undercounting compromises fair distribution of resources. None of such studies have been done in Sub Sahara Africa where most countries use census data to achieve the same goal. This study estimated effect of undercounting on service delivery in the region, using South Africa's 10% census samples. The data provided adjusted counts and these were also used for recreating unadjusted counts. We then applied the PRICEWATERHOUSECOOPERS method to estimate undercounting effect on the allocation of the Basic Service Grant among the country’s nine provinces using both adjusted and unadjusted counts. We further estimated the effects of undercounting on parliamentary seat allocations using the same procedure. We found undercounting to unfairly benefit certain provinces at the expense of others.

Introduction

Authorities largely rely on national data collected from each individual to implement efficient service delivery (Onsembe and Ntozi, 2006). In most cases censuses are the only integral source of such data (Cronje and Budlender, 2004). Availability and quality of such data is often better among developed countries than developing countries especially those from Sub Sahara Africa (SSA) (Onsembe and Ntozi, 2006). In this region census taking processes among majority of countries are relatively substandard (Onsembe and Ntozi, 2006). Therefore in most cases data from these censuses are of poor quality (Anderson, 2004). Census taking is also a massive, highly costly, as well as largely complex process (Cronje and Budlenger). For this reason the capacity to consistently handle censuses among majority of countries in this region has remained a challenge (Sembajwe, 2001).

A persistent challenge in census taking is undercounting (Keane *et., al,* 1995). The challenge is greater among majority of countries from SSA, largely due to budget constraints (Onsembe and Ntozi, 2006). Therefore few countries like South Africa, Nigeria, Tanzania, Mozambique, Zambia and recently Ghana and Zimbabwe tested for census undercount. Even among these countries only South Africa has consistently tested and adjusted the enumerated counts for undercounting. Methods often used to investigate undercount effect on service delivery require information from such data. Hence only South Africa's census data apply to these methods.

Unavailability of such data partly explains why this subject remains unexplored in this region, further contributing to a compromise on quality of service delivery. Empirical evidence on such a subject matter is necessary for planning authorities as well as members of the public. The latter need to comprehend the importance of participating in such data collection processes. For instance, in Nigeria censuses counts are used in deciding on both the distribution of federal funds and political power (Population Reference Bureau, 2006). In Zimbabwe the distribution of the Parliamentary constituencies among the 10 Provinces is based on census counts (Ministry of Justice, Legal and Parliamentary Affairs). Each Parliamentary constituency receives a yearly allocation of 50 000 United States of American dollars, under the Constituency Development Fund (Ministry of Justice, Legal and Parliamentary Affairs). In South Africa census counts are used for allocating funds as well as the 200 Parliamentary seats that are shared among the country's nine provinces.

Compromise on quality of service delivery should further be worsened by irregular census conduction which is characteristic among most countries in this region (Osembe and Ntozi, 2006). This implies that available census data may be repeatedly used beyond recommendable duration. A maximum of 10 year intercensal period is recommended. Research conducted among developed countries particularly in the United States of America (USA) have confirmed the negative effects of census undercounting on service delivery (PRICEWATERHOUSECOOPERS, 2000). The same source further confirmed that the effect becomes greater in cases where the same data is repeatedly relied upon for many years; a practice that is common in SSA.

For such reasons there have often been calls for census adjustment in the USA (Schirm, 1991, Breiman (1994). For example the decision not to adjust for census undercounting in 1990 was met with a lot of disappointment by many States governors (Murray, 1992). Such decisions often provoke controversies as undercounting is rarely homogeneous across sub populations. As Steffey (1997) observed, undercounting poses a serious problem, particularly in that not all areas and sub population groups are undercounted at the same rate. This implies that population groups with higher undercount rates tend to lose more in case of e.g. state funding being distributed according to population counts. For instance in United States of America's census 2000 the implications were likely to be enormous for such populous states recording high undercount estimates like California (Steffey, 1997).

The objective of our paper was therefore; to estimate the effect of undercounting on service delivery in SSA. To achieve this objective in a region where most countries had inappropriate data for our method, we used data from South Africa's last three censuses. We argue that results from the country's nine provinces investigated were largely generalizable across SSA. Of course we acknowledge that using findings from a single country's censuses and generalizing them on an entire region has its own limitations. However, we were contented that given the challenges of inadequate census data from other countries' in the region; findings arrived at using South African censuses provided an eye opener of what to expect from the region. From this perspective, our study is important as it provided a starting point that should stirrer further researches that may be more exhaustive in investigating this subject matter in this region.

Data

South Africa's 10% sample data for censuses 1996, 2001, and 2011 were used in this study. For all censuses the adjustments for undercounting were based on calculations from the Post Enumeration Surveys (PES). For each analysis we applied the respective weighting variable for the purpose of weighting the samples' counts so that they estimate actual census counts. For the purpose of this study we also needed unadjusted data. Since Statistics South Africa (StatsSA), the custodians of this data do not avail unadjusted census data for public consumption, we recreated the data. We achieved this by multiplying the reciprocal of the adjustment factor for each province on the respective adjusted counts. StatsSA arrived at adjustment factors using the following procedure: for instance a 10% undercounting rate; the adjustment factor would be; 1/(1-0.1) = <u>1.1</u>. For a 17% undercount rate; adjustment factor would be; 1/(1-0.17) = <u>1.2</u>, and for an undercount rate of 14.5% adjustment factor would be; 1/(1-0.145) = <u>1.17</u>. The reciprocals for above adjustment factors would therefore be: (1-0.1)/1 = <u>0.9</u> for the first example; then (1-0.17)/1 = <u>0.83</u> for the second example, and (1-0.145)/1 = <u>0.855</u> for the last example (StatsSA Census, 2011). We termed the recreated counts "unadjusted", to differentiate them from enumerated counts termed above as unadjusted.

Measurement of service delivery and Procedures

We adopted funds and parliamentary seats allocations as our proxies for service delivery. This selection was guided by literature which suggest that amount of funds allocated to local authorities as well as number of parliamentary seats are vital in achieving quality service delivery (Steffey, 1997; PRICEWATERHOUSECOOPERS, 2000). According to the South Africa National Treasury IGFR (2008) there are seven main grants under the Provincial Equitable Share funds in South Africa, we chose the Basic Services Grant, whose allocation of funds is entirely based on census counts. The formula for allocating these funds is built on the premise that funding received by a province is proportional to its total population. Hence the formula is given as: **(Provincial Pop/National Pop)*Total Funds Budgeted for particular year from the Grant.** Where **Pop** = population counts.

To estimate the effect of undercounting, the PRICEWATERHOUSECOOPERS method was applied. The method was first used to estimate undercount effect on service delivery by a PRICEWATERHOUSECOOPERS a USA's consultancy company, in census 2000. The company was tasked by the USA government to investigate the effect of census undercounting among the country's States and Counties before the next census in 2000. Their procedure was to first identify key governments grants that were distributed from the USA's central government to the respective States and Counties using a formula that was entirely based on population counts. They identified eight grants. Each of the grants was allocated to the respective State and County based on population counts from both their unadjusted and adjusted census data. The difference in funds allocations based on the two data for each respective State and County was the estimated effect of undercounting on service delivery. Since the census counts were used in allocating these grants for the entire 10 year intercensal period in USA, the difference in funds allocations based on the two data sets for each State and County were projected for 10 years.

In the case of South Africa we used the country's nine Provinces as a replacement of the States and Counties in the case of the USA. Just like South Africa all other SSA countries are dissected into Provinces which also receive various allocations from respective central governments, except for Nigeria which is divided into Federal States like the USA. Using the same logic behind the method's application in the USA scenario, we computed funds allocation for each of the nine provinces in South Africa; firstly based on counts from "unadjusted" and then unadjusted data. The difference in funds allocations based on the counts from the two comparative data sets for each respective Province became the estimated undercounting effect on service delivery, based on this proxy. We did this for each of the three censuses.

We further projected the effect that would accrue over the entire intercensal period, assuming the same formula remained in effect over that period. We did this with clear knowledge that South Africa does not rely on census data in allocating funds to the country's Provinces for the entire intercensal period. The country only uses census data for only one year immediately after the full census results are declared. For other subsequent years before the declaration of next census's results they rely on population estimates from yearly conducted General Household Survey. Rather we did this analysis in the interest of the majority of SSA countries that largely rely on a single census' counts for many years, when allocating resources. For instance as recommended by the United Nations some countries in this region conduct censuses after the lapse of each 10 year period. Yet some have no capacity to conduct the census within this regulated period. The projection of undercounting effect on service delivery over the years was meant to give the estimates for such countries, assuming that they relied on census data for implementing service delivery.

For the other proxy; the Independent Electoral Commission of South Africa (IEC, 1993) states that; out of the 200 parliamentary seats designated to the country's nine provinces, 100 000 eligible voters constitute a seat (IEC, 1993). The same source defined eligible voters as individuals eighteen years and above who registered to vote. We assumed that all individuals who are eighteen years and above from each province as obtained from census counts would have registered to vote. We therefore computed estimated parliamentary seat allocations based on adjusted and "unadjusted" data for each province. The difference was also interpreted as undercounting effect on service delivery based on this proxy. The same logic in allocating parliamentary seats applies elsewhere in SSA, except that the thresholds for allocating such seats vary with countries. For example Zimbabwe's threshold is 30 000 registered voters (Ministry of Justice, Legal and Parliamentary Affairs).

Results

Adjusted and "unadjusted" counts

Generally there were substantial differences between adjusted and "unadjusted" counts, running into hundreds of thousands or more for each province for all the three censuses. As for census 1996, the biggest difference between the counts from the two data sets was from KwaZulu Natal, i.e. 1 078271, followed by Gauteng with 734 072, and Eastern Cape with 666 118. The lowest was from Northern Cape which had 131 059. Census 2001 recorded bigger differences compared to those for census 1996, for the respective provinces. Again, KwaZulu Natal had the biggest followed by Gauteng, and Northern Cape having the smallest. There were some changes in census 2011, where the biggest difference of counts between the two data sets were from Gauteng with 1 804 023, followed by KwaZulu Natal with 1 714 639, with Western Cape ranking third. Northern Cape still had the smallest difference.

[Table 1 here]

Provincial Equitable Share Funds

There are seven main grants that fall under the Provincial Equitable Share (PESF), whose funds are distributed to provinces using various formulas, shown in Table 2. The grants whose funds are allocated using a formula that at least relies on census counts are education, Health, and the Basic Services Grant. The former two are however not 100% dependent on census counts, but also depend on counts sourced from other data sources. The Basic Services Grant is the only one whose funds are allocated entirely based on census counts. However, these census counts are only used during the first year in which the respective census' counts are officially availed. Thereafter, yearly population estimates are used. The first PESF allocations coinciding with investigated censuses were in 1999 and these remained in force for the next 5 years. In 2004 allocations were reviewed upwards as should be expected, and these remained in force for ten years, i.e. until 2013 when the third reviews which are currently in effect were carried out. In terms of amounts allocated, Education, followed by Health, and Basic Services had been receiving the highest shares during each allocation.

[Table 2 here]

Basic Services Grant Funds allocations

The provincial allocations from Basic Services Grant (BSG) based on counts from census 1996 were effected in the year 1999. Three provinces were going to lose if the adjustment for undercount was not carried out. KwaZulu Natal was going to lose about R37 491.27, Northern Cape, R 3 7491, and Limpopo R6 184.36. The other six provinces would have benefited from the undercount. In terms of absolute figures the biggest beneficiary would have been Western Cape. Allocations based on census 2001 were effected in year 2004. Again three provinces would have been prejudiced by the undercount. The provinces would have been Free State, Gauteng, and again KwaZulu Natal. Among the three provinces, the biggest loser in terms of absolute figures would have been KwaZulu Natal, which could have lost about R 135 913.23. Eastern Cape Province would have been the biggest beneficiary from the undercount. Allocations based on census 2011's counts were effected in year 2013. Five out of the nine provinces were bound to lose if the last census counts were not adjusted. The provinces are Gauteng, KwaZulu Natal, Mpumalanga, North West, and Western Cape. In terms of absolute loses Western Cape would have incurred the biggest lose, compared to other provinces, i.e. R 276 845.71. On the other hand Eastern Cape would have been the biggest beneficiary from the undercount.

Projecting the effect of undercounting until the next review, naturally a province would have maintained the status earlier obtained. Meaning that, provinces which recorded loses would have escalated loses after projection, and those with gains also having escalated gains.

The only difference would be that the respective projected losses or gains will be larger relative to those for a single year, as they indicate the cumulative effect of undercounting over the projected years. For instance were as KwaZulu Natal would have lost about R

37491.27 from allocations for 1999 using census 1996's unadjusted counts, loses would have escalated to R 187 456.34 after a 5 year projection period.

[Tables 3-5 here]

Estimated Parliamentary seat allocations

Parliamentary seat allocations in Tables 6, 7 and 8 denoted as official seats are the actual seat allocations from the EIC of South Africa, for respective elections. These were allocated to provinces based on people aged eighteen years and above who had registered to vote per province. The other two sets of parliamentary seat allocations from each of the three tables are estimates based on adjusted and "unadjusted" census counts respectively; assuming everyone who is eighteen years and above registered was to vote. In election 1999, from the actual seat allocations, Gauteng received the highest allocation, with 43 seats. KwaZulu Natal followed, then Eastern Cape, and Western Cape respectively. Northern Cape which got 4 seats received the least number of parliamentary seats. If all people aged eighteen years and above had registered to vote, Gauteng would have obtained about 52 seats, followed by KwaZulu Natal getting about 49, Eastern Cape about 33, Western Cape about 26. However the parliamentary seat allocation would reduce if based on "unadjusted" counts. Gauteng would get about 47, KwaZulu Natal 43, Eastern Cape 30, and Western Cape 24. Northern Cape would have obtained 5 seats based on adjusted counts but about 4 seats if based on "unadjusted" counts.

As for elections 2004 and 2009 the allocations maintained the status quo observed in election 1999 in terms of which provinces would get more seats under the three categories of allocation. The only difference was that expected number of seats to be allocated to each province based on adjusted and "unadjusted" counts during elections 2004 and 2009 would be more compared to those for election 1999. For instance Gauteng would then get 63 seats from adjusted counts, or 51 seats if based on "unadjusted" counts. Allocations of seats based on both adjusted and "unadjusted" counts for election 2014 would have further seen all provinces again gaining more seats. However, like in previous cases estimated seat allocations based on adjusted counts would be more compared to those allocated based on "unadjusted" counts, for each province.

[Tables 6-8 here]

Discussion

The purpose of this study was to estimate the effect of undercounting on service delivery in SSA. We used Grants as well as Parliamentary seats allocations to measure service delivery because access to both is expected to contribute towards improving welfare of a population. Access to funds equips authorities with necessary revenue that can be used to plan for the improvement of the welfare of the population concerned. Equally the same adequate parliamentary representation should contribute towards the same goal. For parliamentarians are vehicles that are constitutionally mandated to spearhead development among communities they represent. These proxies that we used to measure service delivery have often been used in the same capacity in other related studies in the USA.

We were aware of the fact that use of censuses from only one country i.e. South Africa, and generalizing the findings to the entire region of SSA has its own limitations. A much preferred approach would have been to use censuses that were sampled across the region. However, we took cognoscente of the fact that the methods we used to achieve our objective only applied to cases where respective countries' censuses were tested and adjusted for undercounting. Then both adjusted and unadjusted data should be accessible, which was not the case among all SSA countries except for South Africa. These countries only publish enumerated data, without adjusting for undercounting. Our attempt to construct adjusted data for these countries proved difficult since we were not privy to the procedures they would employ in adjusting for undercounting. We could only proceed with the research using South African censuses. As for South Africa we took advantage of the fact that, we were aware of the procedures used in arriving at the country's adjusted census data for the last three censuses. This meant we had information that was necessary for recreating the estimates of the respective censuses' unadjusted data, since only adjusted data was accessible. Furthermore not only one census was investigated, but rather three successive censuses. This gave some fair representation with regards to time.

We were also aware of the fact that the recreation of unadjusted data had its own limitations as this had the potential of introducing further bias in the data. However, we realized that the error was very much unlikely to have any significant effect on our findings. StatsSA made available the official unadjusted counts for the nine provinces for census 1996 (Statistics South Africa, 1998). Based on this we noted that our "unadjusted" counts relative to official unadjusted counts for census 1996, for the respective Provinces only differed slightly. Such a realization led to our assumption that equally the same the “unadjusted” counts we constructed for the other two censuses should not have any significant differences with respective counts from the official unadjusted data.

Our findings from investigations on the Basic Service Grant allocations were that undercounting was associated with unfair and unjustified resources allocations. For instance results based on allocations using counts from census 1996 indicated that the three provinces prejudiced by undercounting would have lost a total revenue of R 52 306.26. This amount would have been unfairly distributed among the other six provinces. Allocations based on census 2001 would have also resulted in the three prejudiced provinces losing a total revenue of R 169 044.88. As for the allocations based on census 2011 counts the five prejudiced provinces would have lost a total revenue of R 605 444.82. This would have meant compromised service delivery among provinces which were prejudiced by undercounting. Provinces unfairly benefitting from undercounting would have received more funds and parliamentary seats relative to what they eventually got after adjustment for undercounting. Naturally the losses incurred by prejudiced provinces would worsen if projected over respective intercensal periods. As for parliamentary seats allocations, if all individuals aged 18 years and above, as obtained from respective censuses, were to register as voters, seat allocations from unadjusted counts would be lesser compared to seat allocations based on adjusted counts, for each province. These findings indicated that undercounting compromised service delivery.

As noted most countries in SSA rely on census counts for funds distribution among sub populations groups for entire intercensal period, largely because of data scarcity. The intercensal periods may at times exceed beyond the standard period of 10 years for certain countries, mainly due to lack of capacity to conduct another census. Such countries may have no alternative besides to rely on this data until another census is eventually conducted. For this reason we found it necessary to also project the effect of undercounting on Basic Services Grand and Parliamentary seat allocations in South Africa, for each intercensal period. Yet these projections may not apply to South Africa’s context, since the country only relies on census counts for only one year; the projections gave an insight on the long term effect of undercounting on service delivery among countries which rely on same census counts for longer durations. Findings from the projections were that undercounting would lead to arithmetic escalated revenue lose for each province over time. The longer the intercensal period would be, the bigger the revenue loss for a particular province. Again the effects would be worse for populous provinces with high undercount rates.

Conclusions

Conclusions are that undercounting often compromises service delivery as partly evidenced by some of the findings from Basic Service Grant allocations. There were certain provinces that would have been prejudiced of deserved funds because of undercounting. Further evidence to back up this conclusion also came from the findings from Parliamentary seat allocations. Undercounting would have always decreased the number of parliamentary seats allocated to each province, compared to seats that would have been obtained using adjusted counts. However, further findings from the Basic Service grant allocations indicated that besides compromising service delivery among certain sub population groups, undercounting would also benefit other sub population groups. This was evidenced by the fact that certain provinces were bound to benefit extra funds at the expense of prejudiced provinces. Overall, we are convinced that these findings highlight possible effects of undercounting on service delivery across SSA at large.

From our findings; firstly, we recommend that all census authorities from SSA countries should consistently implement the measurement and adjustment of census undercount. The benefits of this are wide. Secondly, we recommend more researches on this matter.

References

Anderson, B, A. (2004). “Undercount in China’s 2000 census in comparative perspective.” PSC Research Report, No. 04-565 University of Michigan; 1-14

Berkowitz, P (2012) “The (incomplete) (probably inaccurate) sum of us.” Daily Maverick http://www.dailymaverick.co.za/article/2012-10-31-census-2011-the-incomplete-probably-inaccurate-sum-of-us#.VvreU2Pd6dk 17-09-2013

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

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

Independent Electoral Commission of South Africa (IEC) www.elections.org.za 20/08/2014

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

Ministry of Justice, legal and Parliamentary Affairs http://www.justice.gov.zw/index.php/2013-11-11-14-19-27?showall=1&limitstart= 03/04/2015

Murray, M.P. (1992), "Census Adjustment and the Distribution of Federal Spending," Demography, 29; 319-332.

Onsembe, J, O and Ntozi, J, P, M. "The 2000 round of censuses in Africa: Achievements and challenges." *The African Statistical Journal*, Vol 3; 11-28

PRINCEWATERHOUSECOOPERS (2000) "Effect of census 2000 undercount on federal funding to states and selected counties, 2002-2012." Final Report to Congress

Schirm, A, L. (1991) "The effects of census undercount adjustment on congregational apportionment." The Journal of the American Statistical Association, Vol. 86, No. 414; 526-541.

Schultz, D (2013) "Was census 2011 successful?" Mail & Guardian http://mg.co.za/article/2013-03-22-was-census-2011-sucessful 21-08-2013

Sembajwe, I. (2001) "Making Censuses in Africa Cost Effective: Suggestions for Reflection." 1-8

South Africa National Treasury ***Intergovernmental*** Fiscal Reviews (IGFR) http://www.treasury.gov.za/publications/igfr/ 15/09/2014

Statistics South Africa (1998) "THE PEOPLE OF SOUTH AFRICA: Population census, 1996: Calculating the undercount in census' 96" Dr. FM Orkin Head: http//www.statssa.gov.za

StatsSA Census (2011). "Census undercount and Strategies: Version No 6." Statistics South Africa.https://www.statssa.gov.za/census2011/documents/Census_2011_Undercount%20Strategies-version_6a.pdf 20/03/2015

Steffey, D, l. (1997) "A Review of census undercount issue." Department of mathematical and computer sciences, San Diego State University; 1-25

Tables and graphs

Table 1: Adjusted and "unadjusted" counts from the last three censuses

Province	1996			2001			2011		
	Adjusted	Unadjusted	Difference	Adjusted	Unadjusted	Difference	Adjusted	Unadjusted	Difference
1	3957322	3613431	343891	4513206	3777553	735653	5822734	4745528	1077206
2	6301972	5635854	666118	6415451	5472380	943071	6562053	5715548	846505
3	840662	709603	131059	823429	678507	144922	1145861	992316	153546
4	2633409	2402986	230423	2715587	2207772	507815	2745590	2468285	277305
5	8417423	7339152	1078271	9420961	7301245	2119716	10267300	8552661	1714639

6	3355012	3040648	314364	3662194	3076243	585951	3509953	2986970	522983
7	7348071	6613999	734072	8830155	7178917	1651238	12272262	10468240	1804023
8	2797692	2515405	282287	3125664	2622433	503231	4039939	3413749	626191
9	4927336	4371532	555804	5262037	4504304	757733	5404868	4864381	540487
	40578899	36245190	4333709	44768685	36819351	7949334	51770561	44207678	7562883

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo. NB Figures in the table were rounded

Table 2: Provincial Equitable Share in South Africa

Equitable share component	**Data used**	**Formula**	**Year**	**Amount in Rands (rounded off)**
Education	Total enrolment numbers (Ai) School age cohort (6-17 years) (Pi_{6-16})]	Ai+2 [(Pi_{6-16})]/+∑i2 [(Pi_{6-16})]	1999 2004 2013	**33 680 800** **65 588 275** **162 034 758**
Health	Population with and without Medical AID	(Phi+4Pwi)/∑i(Phi+4 Pwi)	1999 2004 2013	**15 156 360** **30 394 566** **91 144 551**
Social Development	Target population for each grand type	Sum of grants (weighted 75) and Provincial population income quintiles (weighted 25). Grant values are sum of population eligible for old age grant (65), childcare grant (10) and population	1999 2004 2013	**14 314 340** **28 794 852** **10 127 172**
Economic activity	Gross geographic product (replaced by Remuneration data in 2000)		1999 2004 2013	**6 736 160** **11 197 998** **3 375 724**
Backlog	School Survey of Needs Health Sectorial Report		1999 2004 2013	**2 526 060** **4 799 142**
Basic Services Grant	Census	Provincial share of Population	1999 2004 2013	**7 578 180** **11 197 998** **54 011 586**
Institutional	Independent data		1999 2004 2013	**4 210 100** **7 998 570** **16 878 621**

Source: (South Africa) National Treasury. IGFR

Table 3: Basic Services Grant allocation based on 1996 census

Provinces	Basic Services Grant funds allocations for 1999. Based on census 1996 counts (Rands)			Projected funds allocation for period 1999-2003 (Rands)		
	Adjusted	Reconstructed	Gain/loss	Adjusted	Reconstructed	Difference
1	739 036.83	755 499.71	-1 6462.94	3695183.848	3777498.55	-82314.702
2	117 6904.34	1 178 349.90	-1445.66	5884521.186	5891749.51	-7228.324
3	156 995.12	148 364.49	8630.60	784975.4567	741822.47	43152.9867
4	491 793.64	502 418.68	-10624.97	2458968.566	2512093.39	-53124.824
5	1 571 968.4	1 534 477.12	37491.27	7859841.962	7672385.62	187456.342
6	626 554.33	635 741.68	-9187.35	3132771.645	3178708.38	-45936.735
7	1 372 265	1 382 861.42	-10596.38	6861325.229	6914307.10	-52981.871
8	522 473.85	525 923.35	-3449.50	2612369.246	2629616.77	-17247.524
9	920 188.57	914 004.21	6184.36	4600942.861	4570021.07	30921.791
Total	7578180	7577640.571		37890900	37888202.9	

Table 4: Estimates of funds allocation based on 2001 census counts by Province

Provinces	Basic Services Grant funds allocations based on census 2001 counts (Rands)			Projected funds allocation for period 2004-2012 (Rands)		
	Adjusted	Reconstructed	Difference	Adjusted	Reconstructed	Difference
1	1 128 888.9	1148880.41	-19991.50	10 160 000.1	10339923.66	-179923.51
2	1 604 698	1664334.07	-59636.33	14 442 282	14979006.63	-536727.01
3	205 964.47	206356.71	-392.29	1 853 680.23	1857210.364	-3530.57
4	679 250.21	671457.42	7792.77	6 113 251.89	6043116.783	70134.93
5	2 356 466.5	2220553.18	135913.23	21 208 198.5	19984978.61	1223219.08
6	916 025.17	935588.54	-19563.40	8 244 226.53	8420296.888	-176070.60
7	2 208 688.1	2183349.14	25338.88	19 878 192.9	19650142.23	228049.91
8	781 822.84	797569.72	-15746.91	7 036 405.56	7178127.485	-141722.17
9	1 316 194.2	1369909.73	-53715.61	11 845 747.8	12329187.57	-483440.52
Total	11197998.0	11197998.92		100781985.5	100781990.2	

Table 5: Estimates of funds allocation based on 2011 census counts by Province

Provinces	Basic Services Grant funds allocations based on census 2011 counts (Rands)			Projected funds allocation for period 2013-2022 (Rands)		
	Adjusted	Reconstructed	Difference	Adjusted	Reconstructed	Difference
1	6 074 786.2	5797940.658	276845.71	54673075.8	52181465.93	2491611.38
2	6 846 108.5	6983081.492	-136972.63	61614976.5	62847733.43	-1232753.66
3	1 195 462.6	1212381.319	-16918.67	10759163.4	10911431.87	-152267.99
4	2 864 440	3015675.127	-151234.88	25779960	27141076.14	-1361113.90
5	10 711 747	10449379.84	262368.06	96405723	94044418.58	2361312.51

6	3 661 890.4	3649388.732	12502.16	32957013.6	32844498.59	112519.45
7	12 803 500	12789776.2	13724.39	115231500	115107985.8	123519.49
8	4 214 818.3	4170814.436	44004.50	37933364.7	37537329.93	396040.53
9	5 638 831.8	5943152.491	-304319.72	50749486.2	53488372.42	-2738877.46
Total	54011586.0	54011590.30		486104263.2	486104312.7	

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo.

Table 6: Official and Expected Provincial l seat allocation for Election 1999

Province	Official Seats (Based on Voter registration)	10% sample Adjusted 1996 census count (18 years & above)	Expected seats (Based on 10% 1996 adjusted count)	Reconstructed 1996 count (18 year & above)	Expected seats (based on 1996 reconstructed)
1	21	2604325.7	26.0	2378009.797	23.8
2	26	3364049.7	33.6	3008469.647	30.1
3	4	508123.6	5.08	428907.1308	4.3
4	15	1646063.5	16.5	1502032.944	15.0
5	40	4870770.4	48.7	4246824.712	42.5
6	17	2003052.5	20.0	1815366.481	18.2
7	43	5174424	51.7	4657499.042	46.6
8	14	1589657.3	15.9	1429260.878	14.3
9	20	2471035.3	24.7	2192302.518	21.9
Total	200	24,231,502	242.3	21641154.44	216.4

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo.

Table 7: Official and Expected Provincial seat allocation for Election 2004/9

Province	Official Seats (Based on Voter registration)	10% sample Adjusted 2001 census count (18 years & above)	Expected seats (Based on 10% sample 2001 adjusted count)	Reconstructed 2001 count (18 year & above)	Expected seats (based on 2001 reconstructed)
1	23	3013024	30.1	2521901.088	25.2
2	26	3550922.9	35.5	3028937.234	30.3
3	5	519206	5.2	427825.744	4.3
4	12	1697750.4	17	1380271.075	13.8
5	39	5470630.9	54.7	4239738.948	42.4
6	14	2285648.1	22.9	1919944.404	19.2
7	47	6314155.9	63.1	5133408.747	51.3
8	15	1803704.7	18.0	1513308.243	15.1
9	19	2763913.9	27.6	2365910.298	23.7
Total	197	27,418,957	274.2	22757734.31	227.6

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo.

Table 8: Official and Expected Provincial l seat allocation for Election 2014

Province	Official Seats (Based on Voter registration)	10% sample Adjusted 2011 census count (18 years & above)	Expected seats (Based on 10% sample 2011 adjusted count)	Reconstructed 2011 count (18 year & above)	Expected seats (based on 2011 reconstructed)
1	26	4082578.7	40.8	3327301.641	33.3
2	18	3940532.6	39.4	3432203.895	34.3
3	16	735225.4	7.4	636705.1964	6.4
4	17	1793776.9	17.9	1612605.433	16.1
5	18	6309020.5	63.1	5255414.077	52.6
6	16	2280852.6	22.8	1941005.563	19.4
7	22	8848775.7	88.5	7548005.672	75.5
8	16	2524094.1	25.2	2132859.515	21.3
9	20	3190339.1	31.9	2871305.19	28.7
Total	186	33,705,196	337.1	28784237.38	287.8

1=Western Cape; 2=Eastern Cape; 3= Northern Cape; 4= Free State; 5= KwaZulu Natal; 6= North West; 7= Gauteng; 8= Mpumalanga; 9= Limpopo.

Appendix G: Address of comments for the above manuscript

Service delivery implications of census undercount in Sub Sahara Africa

<u>Address of Comments</u>

Reviewer 1

Introduction

1. Year for article by Ensembe and Ntozi was corrected as 2006, not 2004.

2. The suggested rephrasing of second sentence in paragraph 6 may alter our way of thinking; as a result we rephrased this sentence to make it clearer and flowing.

Data

1. A complementary comment, hence no need to address.

2. A sentence that highlight on how the adjusted counts were arrived at was added in the second sentence of the section titled Data, as recommended.

3. Data was weighted using the weighting factor variable, to make the counts from the 10% samples estimate actual population counts. This was also explained under the section on Data.

Measurement of service delivery

1. The general comment made is again a complement of the value of this paper, and hence need no address.

Basic Services Grant Funds Allocation

1. Findings reported in this section have been readjusted after the correction of some minor computation errors that were noted by reviewer 2. Hence the reporting of the results in this section has also been readjusted, partly taking into consideration this comment. The main finding coming from this correction was that undercounting prejudiced certain provinces at the same time unfairly benefitting other provinces.

Parliamentary seat allocations

1. The missing figure was 4 and this has been added to the sentence.

Discussion

1. The incorrect word that was noted by the reviewer had been taken care of during the processing of realigning the discussion with slight adjustments made in the results section.

Conclusion

1. The suggested recommendation from the reviewer has been added in this section.

Reviewer 2

Relevance

1. The comment is a complement that emphasises on the great value of the subject addressed by the paper. This work was also orally presented at the last year's UAPS Conference and the discussant Professor Modise from University of Southampton also emphasised that point.

Scope

1. Firstly, we noted it as a limitation of this paper that our study could only rely on census data from South Africa due to data limitations from other countries within the region. We therefore used what is there as a way of initiating research in this area, and hopefully our findings will motivate census authorities from these countries to implement the test and adjustment for census undercount as we recommended. This way other researches using census data from various countries could be initiated. Also the nine provinces of South Africa included in this investigation should fairly estimate how undercounting could be affecting service delivery elsewhere in SSA.

Introduction

1. The introduction has been revised and made short and compact by; restructuring, shortening, removing, and combining some sentences. Hence even after increasing spacing, the section could not go beyond 2 pages.

Data

1. The comment is a complement of the route we took to recreate "unadjusted" counts after failing to get the official unadjusted counts. Just as we felt; the reviewer also noted that failure to avail the officially unadjusted data by census authorities is in fact odd.

Methodology

2. Again the comment is a complement. It's a mere extension of the above complement. The reviewers revisited our methods of reconstructing "unadjusted" counts, and agree that we were able to recreate counts that closely estimated actual unadjusted counts.

Results

1. Firstly, we make it clear that we did not derive our own estimates of grant allocations based on the PRICEWATERHOUSECOOPERS, as opposed to those from the National Treasury, as perceived by the reviewer. Our method follows the same principle used by the National Treasury in allocating Basic Service Grant through which they used the provincial percentage share of population. It is only that in the United States of America where the method has so far been applied they used States instead of provinces as here in South Africa.

In fact we started by applying the method using South Africa's official adjusted counts to check if the allocations per province remained the same as those obtained from the National Treasury records. We reproduced exactly the same results as those arrived at by the National Treasury. This confirms accuracy of our computation since we both used adjusted data to arrive at these allocations across provinces. We then applied the same method to our recreated "unadjusted" counts so as to examine the difference between allocations based on adjusted and those based on our recreated "unadjusted "counts. The latter counts carried the undercount whereas the former had been corrected for the undercount. This way we estimated the effect of undercounting on service delivery using this proxy of Basic Grant Service allocations.

2. However, we agree with the reviewers on that there was an error on the amount of funds allocated to provinces based on our recreated "unadjusted" counts. For the proportions from provinces did not add up to 100%; hence we revisited our calculations and discovered where the error came from. It was a slight error in setting up the formula, and after correcting it our results matched what the reviewer suggested. Tables 3, 4, and 5, show these corrections under column for "unadjusted" grant allocations, and under the column for differences.

Conclusions

1. Yes, our initial conclusions though not whole affected by errors from our initial computations, agreeably were to some extent affected. So after incorporating corrected results, we revisited, and accordingly corrected all affected sections, in particular the Results, Discussion, and Conclusion sections.

General Remark

1. True as noted by the reviewer after correcting for our error in the initial computation, findings indicate that Western Cape would have been prejudiced by the undercount for census 2011, yet Eastern Cape would have benefitted from the undercount.

Appendix H: Reviewers comments for above manuscript

Service delivery implications of census undercount in SSA

Reviewers' comments

Review #1

Title

The acronym SSA/abbreviation should be in   be in full in the title.

Introduction

1. Verify year with reference to Onsembe and Ntozi, should it not be 2006? Refer to paragraph 4.
2. In paragraph 6 in the second sentence**--- effect** should be replace by **under allocation of resources to service delivery will be** escalated**-------------**

Data

1. *General comment*: the samples of 10 percent are   adequate to ensure reliable estimates**.**
2. It is suggested that you indicate upfront that the adjusted figures are based on the adjusting calculated from the Post Enumeration Survey, results, which preceded the 2011 census.
3. Clarify what you mean by "we weighted each data   before using it. A brief elaboration will help the reader. I take that this pertains to the estimation of population values.

Measurement of service delivery and procedures

General comment:
Undercoverage/undercount is one of the limitations of undertaking censuses in most sub Saharan African countries. Despite the above most of them do not adjust the census figures as obtained during enumeration.   Thus there are inherent undercounts during the census year and the intercensal years. This paper clearly brings out the impediment of objective service delivery in terms of allocation of resources and parliamentary seats to provinces /regions.   For some countries allocation of resources for service are not strictly based on census figures but partly on arbitrary assumptions.

Basic Services Grant Funds allocation

In the last paragraph it is not clear to which year the projected deficit of R 1,006,846   refers to.

Parliamentary seat allocation

In the first paragraph and the last line a figure is missing before the word **if ----**

Discussion

In paragraph 5 in the third sentence the word mathematical should be replaced by mathematically.

Conclusion

In addition to recommending for more research, it is advisable that SSA countries should consistently carry out objective tests after each census. The results in the discrepancy between the adjusted and unadjusted figure would facilitate evidence based decision making in the allocation of resource to service delivery in Provinces.

Review #2

Relevance – The paper entitled "**Service delivery implications of census undercount in SSA**" deals with an interesting and important question about the quality of census data and its use as a distribution key for a just and equitable resource allocation within a country. In the absence of good administrative data, the census data becomes almost the only source on which governments will rely for planning and decision making. The undercount in a census is an important indicator of the quality of a census. High undercount rates mean that a large portion of the census population has been missed, whereas low undercount rates show that census takers were able to reach a significant portion of the census population. The paper has questioned the accuracy of the allocation of resources on the basis of census data to provincial spheres in South Africa. Given the controversies that one has seen over the years around the accuracy of South African censuses, the issue that the authors have attempted to address is nothing new; however, it is relevant.

Scope – The intended scope of the paper – Sub-Sahara Africa – is too ambitious but commendable; however, the authors have failed to produce any analysis beyond South Africa that would justify the generalization of the results to the rest of the region (SSA).

Introduction – the introduction is too long and winding. It should be short and precise.

Data – the data used is based on samples of census data of 1996, 2001 and 2011 conducted by Statistics South Africa. Use has also been made of data from the National Treasury and the Independent Election Commission (IEC). The respective sample data is weighted to give provincial person-counts. The weights that are supplied with the data would seem to be a combination of sample weights, adjusted for undercount as is easily seen by upward aggregation to provincial levels. Therefore various sample weighted aggregates coincide with published official figures. The authors seem to be interested in the **unadjusted raw counts of the census data**, which would naturally give an indication of the extent of undercount in the census. The natural and direct route to obtain that information is through the statistical agency. Apparently they have been unsuccessful in obtaining counts from Stats SA, except for 1996. This seems to be odd as there ought not to be an issue of disclosure of sensitive personal information at this high level of aggregation (provincial). Furthermore, why make it available for one census and not for all three.

Methodology – At the centre of their analysis, the authors intended to use both the unadjusted counts as well as the final adjusted counts. Having been denied access to the official unadjusted counts, the authors have attempted to reverse-engineer the sample weights to estimate the adjustment factors, and hence could estimate the raw and unadjusted census counts. A comparison with Stats SA unadjusted counts from Census 2011 shows that they have succeeded getting almost the same counts as those they have been denied access to.

The 'hidden' unadjusted census counts can easily be derived by using the provincial undercount rates (see table on page 8) published by Stats SA together with the corresponding adjusted census totals (Table 3.1, p. 14)[1] :

[1] http://www.statssa.gov.za/publications/P03014/P030142011.pdf

Suppose X denotes the adjusted census count, X1 the PES adjusted count and X2 the out-of-scope count; let U denote the undercount rate and f the adjustment factor. Then

X=X1+X2; f= 1/(1-U) = [(X-X2)/X1] = [X/X1-X2/X1].
If X2 is small relative to X1, then 1/(1-U) =X/X1 approximately or X1=X*(1-U)

When these 'reverse-engineered' unadjusted estimates were compared to the actual census raw counts there hardly differences as is apparent from the following table:

Table 1: Recreated census raw data

					Percentage differences to census counts	
Province	Undercount rate	unadjusted Census counts	Recreated unadjusted counts by authors	Unadjusted recreated counts by Arrow	Authors	Arrow
WC	0.185	4,810,694	4,745,528	4,745,528	1.4	1.4
EC	0.129	5,744,083	5,715,548	5,715,548	0.5	0.5
NC	0.144	969,682	992,316	980,857	-2.3	-1.2
FS	0.101	2,471,749	2,468,285	2,468,285	0.1	0.1
KZN	0.167	8,561,465	8,552,661	8,552,661	0.1	0.1
NW	0.149	2,949,287	2,986,970	2,986,970	-1.3	-1.3
GP	0.147	10,491,375	10,468,240	10,468,240	0.2	0.2
MP	0.155	3,393,848	3,413,749	3,413,748	-0.6	-0.6
LP	0.100	4,828,214	4,864,381	4,864,381	-0.7	-0.7
RSA	0.146	44,220,397	44,207,678	44,196,220	0.0	0.1

The small differences could be attributed to exclusion of out-of-scope units.

Results – the paper claims that BSG allocations, which are based on the PES adjusted census 2011 results, have disadvantaged some provinces, notably KZN and GP. However, they have made a mistake by comparing their estimates of the BSG with those from the National Treasury which are based on adjusted census data. The BSG is allocated using the provincial percentage share of the population. This is obvious from the following table:

Province	adj. Census 2011	BSG NT	%share of pop	BSG NT %share
WC	5,822,734	6,074,786	11.2	11.2
EC	6,562,053	6,846,109	12.7	12.7
NC	1,145,861	1,195,463	2.2	2.2
FS	2,745,590	2,864,440	5.3	5.3
KZN	10,267,300	10,711,747	19.8	19.8
NW	3,509,953	3,661,890	6.8	6.8
GP	12,272,263	12,803,500	23.7	23.7
MP	4,039,939	4,214,818	7.8	7.8
LP	5,404,868	5,638,832	10.4	10.4
RSA	51,770,561	54,011,585	100.0	100.0

The authors have derived their own estimates of the BSG based on what they call the PriceWatersHouseCoopers. Part of their table is shown below:

Province	BSG PwC reconstructed	BSG NT	difference	% difference	
WC	5,962,961	6,074,786	-111,826	-1.9	
EC	4,950,951	6,846,109	-1,895,158	-38.3	
NC	1,035,271	1,195,463	-160,192	-15.5	
FS	2,575,132	2,864,440	-289,308	-11.2	
KZN	8,922,885	10,711,747	-1,788,862	-20.0	

NW	3,116,269	3,661,890	-545,622	-17.5	
GP	10,921,385	12,803,500	-1,882,115	-17.2	
MP	3,561,521	4,214,818	-653,297	-18.3	
LP	5,074,949	5,638,832	-563,883	-11.1	
RSA	46,121,323	54,011,585	-7,890,262	-17.1	

It is mainly on the strength of this table that the authors have come to the conclusion that the distribution of the BSG has disadvantaged some provinces, especially KZN and GP.

There are problems with such a conclusion. First, the re-calculated BSG does not add up to the total NT grant – falling short by 17%. Since this amount is fixed, any re-distribution of the BSG must be constrained by it. Secondly, the authors seemed to have used their re-created and unadjusted census data. This approach would have been permissible if they had shown a compelling argument why adjustment for under coverage was undesirable. This would have been a different ball game altogether, amounting to a rejection of the census results.

A slightly different approach would have been insightful, and would not result in the throwing out of the baby with the tube water. I have used adjusted PwC census counts and applied this to derive BSG PwC estimates. In so doing, I have maintained to constrain the total allocation of the BSG. The results are shown in the next table:

Province	BSG PwC reconstructed	BSG NT	diff	%diff	BSG PwC adj	BSG NT- BSG PwC adj	% diff in BSG
WC	5,962,961	6,074,786	- 111,826	-1.8	7,316,516	- 1,241,730	-16.97
EC	4,950,951	6,846,109	- 1,895,158	-27.7	5,684,214	1,161,894	20.44
NC	1,035,271	1,195,463	- 160,192	-13.4	1,209,428	- 13,966	-1.15
FS	2,575,132	2,864,440	- 289,308	-10.1	2,864,440	-0	-0.00
KZN	8,922,885	10,711,747	- 1,788,862	-16.7	10,711,747	-0	-0.00
NW	3,116,269	3,661,890	- 545,622	-14.9	3,661,890	-0	-0.00
GP	10,921,385	12,803,500	- 1,882,115	-14.7	12,803,499	1	0.00
MP	3,561,521	4,214,818	- 653,297	-15.5	4,214,818	0	0.00
LP	5,074,949	5,638,832	- 563,883	-10.0	5,638,832	0	0.00
RSA	46,121,323	54,011,585	- 7,890,262	-14.6	54,006,233	5,352	0.01

Conclusion – the conclusions the authors have come to are erroneous because they have been drawn from an inappropriate analysis of the data. A quick re-analysis would reveal that the NT BSG allocations would have been different from their own in the Cape provinces, where Western Cape seems to have been short-changed, while the Eastern Cape has received more than its due.

A similar argument can be made with respect to the allocation of parliamentary seats in the National Assembly.

General remarks – If my own calculations are correct, then there might be a case for the Western Government to challenge the allocation of the BSG, but this would be to the disadvantage of the Eastern Cape. Due to the relevance of the analysis of the quality of census data with a view to informing governments accurately, I would suggest the authors consider a thorough rework of their paper.

Appendix I: Sample of published work as first author

Gender & Behaviour Vol. 12 No.4, 2014

Power Dynamics and men's vulnerability to HIV infection in Sub Sahara Africa

Jeremy Gumbo[1], Clifford Odimegwu, Stella Makoni, Nyasha Chadoka and Garikai Chemhaka

Demography & Population Studies Programme, School of Social Sciences, University of Witwatersrand

Abstract

Background: *Majority of researchers in Sub Sahara Africa have largely concentrated on investigating how power dynamics exposed women to HIV infection. However, in this study, we investigated the association between power dynamics and risk of HIV infection among men in this region.*

Methods: *Data was drawn from Demographic and Health Surveys of Cameroon, Kenya, Nigeria, and Zimbabwe. Our population of interest were sexually active men aged between 15-54 years. For univariate analysis we did tabulations to produce frequency distribution of respondents, among key variable of the study. For multivariate analysis, Poisson Regression model was applied to test the association between power dynamics and number of life time sexual partners.*

Results: *Estimates of men reporting having had more than one life time partners were above 70% for each country. Testing for HIV infection ranged between 13% in Nigeria and 40% in Kenya and Cameroon. Multiple sexual partnerships practice was highest in latter country, 48%. Adjusted coefficients suggested that expected number of life time partners was likely to be higher among men wielding more power in all countries except Zimbabwe.*

Conclusion: *Our findings suggest that power dynamics in Sub Sahara Africa are associated with men's vulnerability to HIV infection. This is largely because, certain risk behaviours among men in SSA are nurtured and promoted out of excess power they have.*

Keywords: Power dynamics, Risk of HIV infection, Socio-cultural, Sub Sahara Africa, vulnerability.

---

[1] Corresponding author: jeremy.d.gumbo@gmail.com

such subordination is reduced control over issues that affect their sexuality and consequently their reproductive health (Gupta, 2000). For instance lack of power to make independent decisions often means women are not able to negotiate for safe sex, e.g. through condom use, or reduced coital frequency. Condom use is a key behavioural practice that greatly reduces risk of any form of sexually transmitted disease. Yet on the other hand high coital frequency is believed to increase risk of infections from various diseases that are sexually transmitted, particularly when there is inconsistent or no condom use.

However, in this study we noted that such power dynamics may not only expose women to the risk of HIV infection, but men as well. As UNFPA (2011) has noted, addressing power dynamics between men and women especially on how they affect sexual relations is critical in eventually strategizing on how to reduce further HIV transmissions. Such an observation further highlights the fact that power dynamics in SSA equally expose men to STI and/or HIV infection. Societal norms, beliefs and practices in SSA often built the sense of masculinity among men. Examples include; men taking prestige in having multiple sexual partnerships, disregarding some health practices like HIV tasting, and none condom use during sexual intercourse. These practices increase an individual's risk of infection from any form of STI. UNFPA has noted that due to power dynamics in SSA, boys and men are usually expected to show knowledge and experience in sexual matters, behave in ways that prove they are "really" men, which at times might mean not seeking medication when they are sick.

This may further be exacerbated by the fact that such power dynamics consequently shape societies in such a way that certain risky behaviours become condoned, acceptable, and even encouraged among men. For example research identified that having many life time sexual partners is prevalent particularly among adult men in Cameroon (Oluoch et al, 2011). In this country, multiple sexual partnerships (MSPs) practice among men is estimated to be above 27% (Bingenheimer, 2010). In Kenya the same source further noted that MSPs practice among men is above 10.4% and under reporting of the practice is possible. Other researchers have also noted that in Nigeria having many life time partners and MSPs practice are common phenomenon among men (Mitsunaga et al, 2005), and the latter practice is estimated at 44.9%. In Zimbabwe MSP practice among men is also very common, and the name "small house" has been coined to describe the mistresses that married men go out with (Gumbo, 2010). Having many life time partners is equally common among men in Zimbabwe. For instance, statistics indicate that about 33% of adult males had 2 to 3 life time partners, and 30% had 4 or more (Ministry of health & Child welfare and National AIDS council of Zimbabwe, 2004).

Our realisation of high prevalence of MSPs practice among men in SSA and their dominance over women (Leclerc-Madlala, 2004), in a region where majority of transmission are through heterosexual intercourse motivated this study. Our aim was to investigate how excess power among men which is largely culturally driven exposed men to HIV infection. Studies that have taken the same dimension were mainly conducted among developed communities, particularly the United States of America (Pettifor et al, 2004). In such societies power imbalance might not be as wide as it is in SSA. Hence the finding may not be easily generalized to this region.

The study is nested within the context of two theories which are; Gender and Power theory, and Social Dominance Theory. The former states that power is disproportionately distributed according to gender, and men tend to be more powerful than women. Power imbalance according to the theory, can be perpetrated through; force, monopoly of access to resources, societal obligations, and ideological consensus (Sanchez, et al, 2013). The relevance of the theory to the study is that it acknowledges that the patriarchal nature of Sub Sahara Africa allows men to be understood as the owners of resources. For instance household resources like land, livestock, children, and women themselves including what they work for, are often perceived to be properties of the household head who is often a man. Such power which is linked to access of resources by men partly influences them to engage in risky sexual behaviours like MSPs because they are likely to afford doing so.

The latter theory posits that societies have hierarchical classes that work to the advantage of certain groups of people at the expense of others. Dominant groups often seek to override those who are less powerful to them, and in many cases this is partly achieved through discrimination. When bringing gender into perspective, this theory argues that men will often be more social-dominant compared to women. Meaning, the societal classes created by a patriarchal society classify men and women into distinct classes. The former occupy the upper class, and can override the lower classes of women and children on issues that concern everyone. Such excess power among men may reduce for example the effect of spousal monitory by their partners. Yet research has noted that effective spousal monitoring reduces individuals' chances of engaging in risky sexual behaviours (Gumbo, 2010).

Methods

Design

This is a cross sectional study that uses Demographic and Health Survey (DHS) data drawn from Cameroon, Kenya, Nigeria, and Zimbabwe. The four countries represent sub regions of SSA; i.e. Central, Eastern, Western, and Southern regions respectively. Despite time-variant in dates when the data was collected among these countries, the content of the information portrayed is largely the same. The latter characteristic makes the data more ideal for this study which investigates the same association across the respective countries. Population of interest to the study were sexually active men aged 15-54 years, i.e. Cameroon, 7188; Kenya, 3465; Nigeria, 15478 and Zimbabwe, 7480.

Variable description

The study identified 4 outcomes variables that measure vulnerability to HIV infection. The first one was number of life time sexual partners, a count variable. Secondly was, multiple sexual partners that respondents have had in the last 12 months, including spouse. Though this can also be treated as a count variable, in this study it was managed as binary. Respondents with one sexual partner were classified as sexual exclusive, and those with more than one, as having had multiple sexual partners. Third outcome variable was Ever had an STI in the last 12 months, and it also had binary outcome, yes or no. The fourth was; Ever been tested for HIV, and also had binary outcomes of either yes or no.

Power dynamics was measured using four decision making variables of: Who decides; over fertility; how to spend wife's salary; daily purchases; and on large household purchases? The question asked e.g. for the latter was; who usually makes decisions about making major household purchase? Questions that collected information on the other three variables were also phrased in a similar way. Men who made decisions alone were termed sole deciders, those who decided together with partner were termed joint deciders, and those whose partners were sole deciders were termed; partner is the decider. Sole decider suggested men were autonomous, joint decider suggested men who were fairly powerful, and partner decider suggested non autonomous men. These variables were separately used in the analyses to estimate power dynamics. In the case of Zimbabwe whose DHS data did not have the above variables, power dynamics was measured by; land ownership, and house ownership.

Our control variables were; education level of the respondents, age of the respondent, and religion.

Procedure

The first stage involved some data quality checks, where all the male records drawn from the four DHSs of the selected countries were interrogated. We did this by testing for digit preference. The quality of data has a large bearing on the validity of findings arrived at by the study.

Second stage was data analysis at univariate level, involving tabulations of all variables of interest to this study. Then, followed by multivariate analysis where only one outcome variable, i.e. number of life time sexual partners was used. This was so because information on the other three variables was either restricted to only two countries or the frequency distribution of responses across categories were heavily skewed towards one category, which could not allow the intended binary logistic regression analysis.

As for key independent variables only two were used for multivariate analysis, and these were who decide on usage of wife's earning, and decide on daily purchases. The selection of the two was based on the fact that the distribution of their respondents was not heavily skewed in one category. The regression model used was the Poisson. The model applies when the outcome variable is count i.e. all integers are positive (i.e. 0, 1, 2, 3, and 4 etc), and when the data is not over dispersed. The Poisson regression model applies well when the data in the count variable is skewed around the count 0. If the data becomes evenly distributed across all counts, implying wide dispersion, or when the data is symmetrically distributed it is not advisable to use the model. In most cases data from count variables are often skewed around 0. For this reason most studies that have used this model often assume that the data in their count outcome variable is skewed around 0. We made the same assumption in this paper.

Results

Data Quality

Suggestions from data plotted on graphs below (Figs 1-4) were that, in all countries there was age misreporting though there were country variations in the extent of age misreporting. Most preferred digits were multiples of 5. Suggestions are that quality of age reporting in Southern Africa was better than in other regions. For example age reporting for Zimbabwe in most cases suggested less digit preference, except in few instances e.g. where ages 20, 25, 40 and 50 seem to have had some clear preference. The overall trend of age distribution for Zimbabwe therefore looked normal and expected for a developing country. For example, the age distribution clearly indicated higher population in younger ages and constantly dwindling population with increase in age.

With regards to Cameroon, Kenya and Nigeria age reporting looks somehow bad. In all the three countries there were clear indications of high digit preference for multiples of 5; from age 15 years to the last age reported in the respective data sets. To some extent Cameroon's age distribution suggested a trend again consistent with that of developing countries. As for Kenya age preference was more pronounce, and become worse for Nigeria, which might suggest poor age reporting in Eastern and Western Africa.

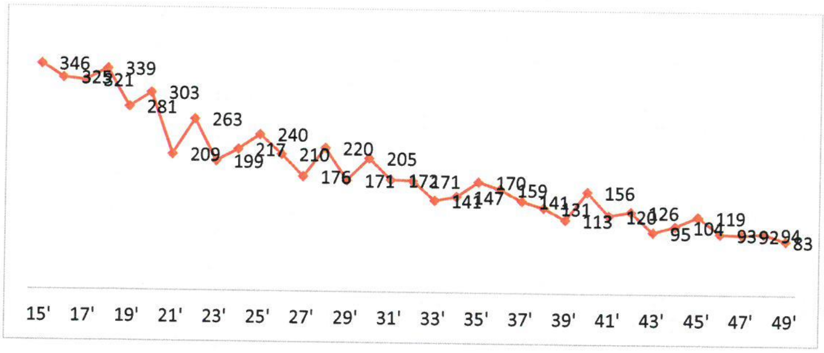


Fig 1: Test for age preference, Cameroon

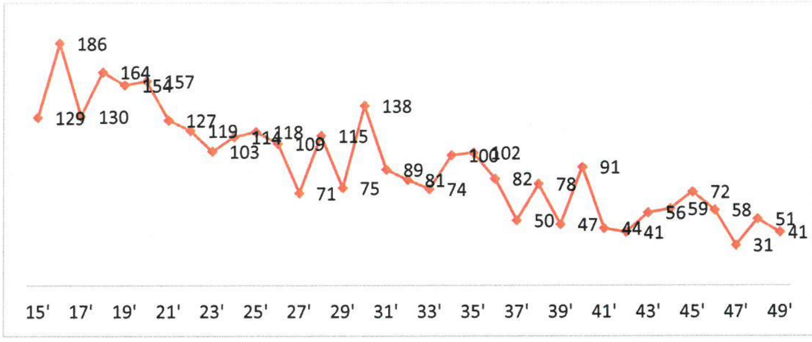


Fig 2: Test for age preference, Kenya

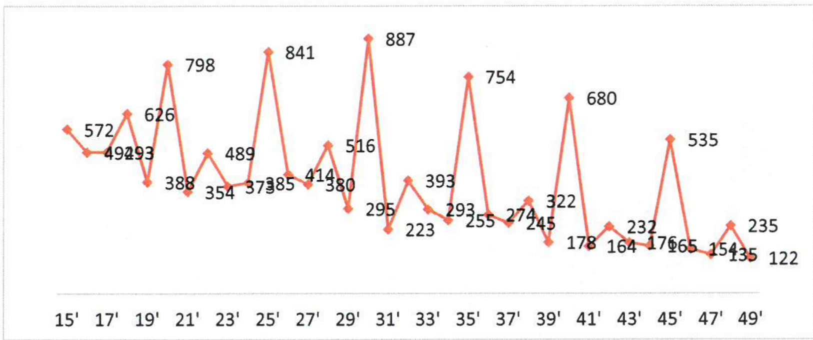


Fig 3: Test for age preference Nigeria

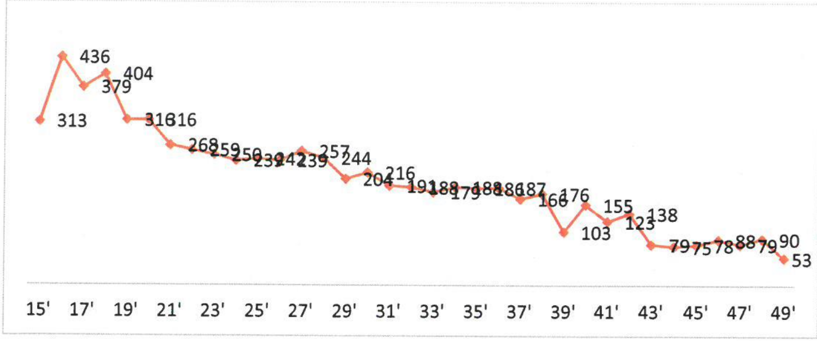


Fig 4: Test for age preference, Zimbabwe

Univariate

Number of life time partners among males by country

In Fig 5, all the countries were plotted in one graph that describes and compares the levels of life time partners by age. Indications were that majority of men from all the sub regions of Sub Sahara Africa (SSA) have had many life time partners. In Nigeria 70% of men have had more than one partner in their life. Statistics from Kenya and Zimbabwe suggest that above 80% may have had more than one sexual partner in their life time. In Central Africa, about 90% of adult males report having had more than one sexual partner in their lives. A common trend observed from all countries analysed was that majority of men were found within the range of those who have had between 1 and 5 life time partners.

The frequency distribution of males reporting having had many life time partners seem to fall significantly after 7 lifetime partners. However, the frequency surges up for those reporting having had 10 life time partners, in all countries.

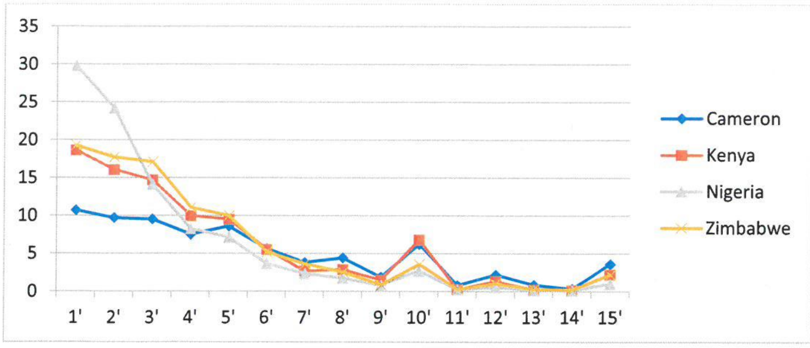


Fig 5: % distribution of life time partners for men in selected SSA countries

Multiple concurrent partnerships (MCPs) practice

The latest DHS data that have information on MSPs practice were for Zimbabwe and Cameroon only. The proportion of men reporting having MSPs in Cameroon was higher compared to those who were sexually exclusive. The pattern was contradictory in the case of Zimbabwe where proportion of men who were sexual exclusive was more than that for men who had MSPs. Yet the difference between the two comparative groups was minimal for Cameroon, it was evidently very wide for Zimbabwe

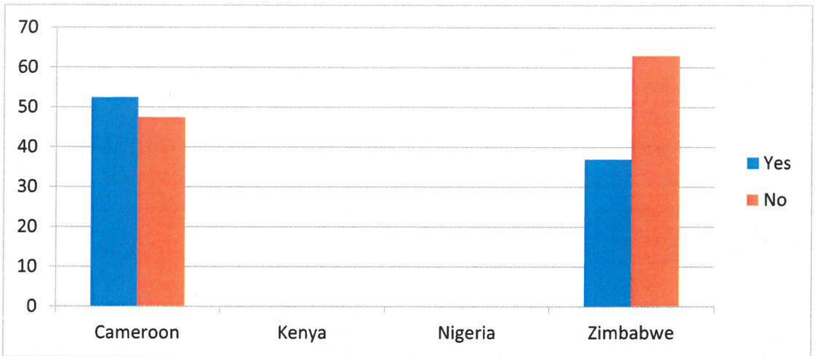


Fig 6: % of men who had MSPs relative to those who were exclusive

Prevalence of sexually transmitted infections (STI) among selected countries was generally low across the entire region. For countries like Nigeria and Kenya STI prevalence was close to only about 1%, being slightly higher in Zimbabwe at about 2%, and in Cameroon around 3.5%.

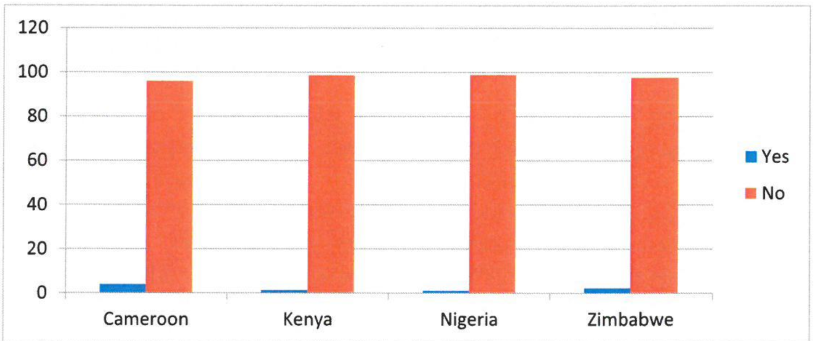


Fig 7: % of men who ever had an STI compared to those who have never had

HIV testing by country

For all the countries, proportion of men who have never tested for HIV was always higher than for men who have been tested. In Zimbabwe less than 40% of men were tested, and if this is the trend across the entire Southern African region this could be a cause for concern in a region believed to be the epicentre of the pandemic. Nigeria had lowest proportion of men who have had an HIV test i.e. about 13% compared to 87% who have never been tested. For other countries, at least men who have had an HIV test where above 40%, though still this is quite very low.

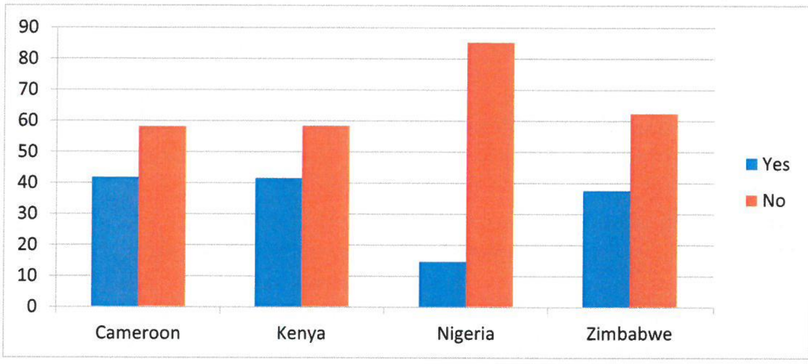


Fig 8: % of men who have tested for HIV compared to those who have never tested

Decision making

There was very little to suggest that women were sole decision makers over their fertility, with all indications pointing to men as key decision makers. Except in the East, both West and Central Africa majority of men had more powers in fertility decision, as indicated by high proportion of men who were sole decision makers in Nigeria and Cameroon. Men, who made joint decisions with partners in East Africa, were the majority. In West and Central Africa there were also substantial proportions of men who made joint decision with partners on fertility. However, the proportion of respondents who were sole decision makers on fertility was overall low. In West Africa, where Nigeria had about 7% of women who had sole decision making powers over their fertility was the highest.

Decision on how to use wife's salary was also an indicator of men's dominance over women. For all the countries examined none suggested that at least half of the women population were autonomous in deciding how to use the money they earn. Rather it was apparent that a substantial number of men had influence on how their partners should use their money. For example, in Nigeria up to 30% of men were sole deciders on how their partners' earning had to be used, compared 34% of women who were sole deciders. Men's power over their partners was also evident on who decides on purchase of large house hold goods. Except in Southern Africa where such decisions were mainly jointly done, elsewhere across SSA, majority of men were sole deciders. Majority of women only appear to have power to decide over daily purchases. However, this is not to say there aren't substantial proportions of men who are also sole deciders over daily purchases. For example in West Africa indications were that slightly over 50% of men were sole deciders on daily purchase, where as in central Africa it was about 28% of men.

Table 1 Power dynamics estimation

Power Dynamics variable	Cameroon % distribution	Kenya % distribution	Nigeria % distribution	Zimbabwe % distribution
Decision on fertility				
.Respondent	43.41	18.98	44.79	-
.Joint	47.11	73.71	45.81	-
.Wife	5.50	3.00	2.11	-
.Depends	3.98	4.31	7.29	-
Decision on wife's earning usage				
.Respondent	25.03	17.68	30.82	-
.Joint	27.40	48.44	31.52	-
.Wife	46.17	30.36	34.85	-
.Depends	1.39	3.52	2.81	-
Decision on large house hold purchases				
.Respondent	57.76	44.59	82.26	10.46
.Joint	27.51	46.81	14.52	74.57
.Wife	13.51	7.45	2.24	14.42
.Depends	1.22	1.15	0.98	0.56

Decision on daily house hold purchases				
.Respondent	28.23	16.42	54.12	-
.Joint	13.36	28.46	14.84	-
.Wife	57.79	53.25	29.98	-
.Depends	0.63	1.88	1.05	-
Land ownership				
. Respondent doesn't own	-	-	-	65.37
. Respondent owns alone	-	-	-	15.88
. Jointly	-	-	-	13.32
. Both jointly & alone	-	-	-	5.43
House ownership				
.Respondent doesn't own	-	-	-	64.77
.Respondent owns alone	-	-	-	14.73
.Jointly	-	-	-	14.21
.Both jointly & alone	-	-	-	6.28

Multivariate

All interpreted coefficients in this section were adjusted and should be understood as significant unless specified as either way.

For Cameroon, expected number of sexual partners was 0.08 lower for men who made joint decision on wife's earnings compared to men who make sole decision. As for men whose wives were sole decision makers on usage of own earnings, the expected number of sexual partners was 0.05 lower compared to men who were sole deciders on usage of wife's salary. As for Kenya, like in Cameroon expected number of sexual partners was lower among men who made joint decision, as well as those whose wives were sole deciders on usage of own earnings compared to men who were sole deciders on wife's earnings. However, for Nigeria, the expected number of sexual partners did not vary for both men who were joint deciders and those whose wives were sole deciders on their own earnings, when each was compared to men who were sole deciders.

In Zimbabwe, the expected number of sexual partners among men who solely owned land was likely to be 0.21 lower compared to men who did not own land. Among men who were joint owners of land, the expected number of sexual partners was 0.23 lower relative to men who did not own land. Also the expected number of sexual partners was likely to be 0.04 and 0.31 lower for men who jointly own house with partners and those whose wives were solo house owners respectively, compared to men who did not own a house.

For other variables, firstly age; if a man was to have an age increase by a year, the difference in expected number of sexual partners would increase by 0.04 for Cameroon, Kenya, Zimbabwe, and 0.03 for Nigeria. The effect of Education on number of sexual partners had a similar direction of flow in all the selected countries. The expected number of sexual partners was likely to be higher for men with primary education, secondary education, and tertiary education, compared to men with no education. As for religion; number of sexual partners would be expected to be lower for men who were Protestants, and muslins relative to Catholics, in Cameroon, Kenya and Nigeria. In Zimbabwe, and men who were Pentecostal, Apostolic and None Religious were expected to have lower number of sexual partners compared to men who were Catholics.

Table 2 Coefficients from poison regression analysis

Variable	Cameroon		Kenya		Nigeria		Zimbabwe	
	Coeff	P<0.05	Coeff	P<0.05	Coeff	P<0.05	Coeff	P<0.05
Decider on wife earning usage								
.Respondent	RC		RC		RC			
.Joint	-0.08	0.001	-0.32	0.001	0.003	0.68		
.Wife	-0.50	0.001	-0.17	0.001	0.001	0.88		
Decider on daily purchases								
.Respondent	RC		RC		RC			
.Joint	-0.09	0.001	0.07	0.001	0.22	0.001		
.Wife	0.17	0.001	0.13	0.001	0.37	0.001		
Age of respondent	0.04	0.001	0.04	0.001	0.03	0.001	0.04	0.001
Education level of respondent								
.None	RC		RC		RC		RC	
.Primary	0.60	0.001	0.30	0.001	0.17	0.001	0.30	0.001
.Secondary	0.82	0.001	0.04	0.101	0.32	0.001	0.48	0.001
.Tertiary	0.70	0.001	0.10	0.001	0.43	0.001	0.43	0.001
Religion								
.1	RC		RC		RC		RC	
.2	-0.06	0.001	-0.05	0.001	-0.25	0.001	-0.23	0.001

.3	-0.38	0.001	-0.25	0.001	-0.76	0.001	-0.41	0.001
.4	-	-	-	-	-	-	-0.64	0.001
.5	-	-	-	-	-	-	-0.10	0.001
Land								
ownership							RC	
.Doesn't							-0.21	0.001
.owns alone							-0.23	0.001
.jointly with wife								
House								
ownership							RC	
.Doesn't							-0.04	0.001
.owns alone							-0.31	0.001
.jointly with wife								

Religion 1= Catholics; 2 = Protestants; 3 for Kenya, Cameroon & Nigeria = Muslims, but for Zimbabwe is Pentecostal; 4 = Apostolic; 5 = No religion

Discussion

Our study investigated how excess power which is largely culturally bestowed on men in SSA may increase their risk of HIV infection. However, like any other study, our investigation had some limitations. For instance, the variables intended to be used as proxy measurement for power dynamics became non uniform, as the latest Zimbabwean DHS data had different measurements for power dynamics. For countries from West, East and Central Africa, power dynamics was measured by; who made decisions on various issues that relates to fertility, purchases, and wife's earnings. For Zimbabwe we used variables that asked if the man owned property i.e. land and/or house. But such a limitation only went as far as compromising our comparisons for the effect of power dynamics on men's vulnerability to HIV infection among sub regions.

Our key findings were that; excess power among men is likely to increase their vulnerability to HIV infections. Such a finding was consistent with theories that guided this study. The Gender and power theory suggests that men have excess power over everything including resources. In such societies, men are often economically powerful than women. As a result men can take of advantage of their economic muscle to have many sexual partners, as well as changing them over time. Such practices expose men to STIs and HIV infection. Equally the same the social dominance Theory was also supported by our findings. By emphasising on variation in classes, where men are more dominant the implications are that men may override decisions, feelings and or views from their partners/spouses. For instance powerful men may get poor spousal monitoring from their partners since such women may lack confidence to trace their spouses' practices. This may result in such men engaging in risky sexual behaviours which increase their exposure to HIV and other STIs infections. Our findings are also consistent with prior research, which suggested that power dynamics in SSA associates with higher likelihood of HIV infection (Shisana, 2004).

At descriptive level it was also evident that few men ever had HIV test, a behaviour which should be associated with the 'mancho man' that society may have constructed about men. In this way health behaviours like going for HIV tests are likely to be viewed as a sign of weakness. Whereas, being infected by an STI may be understood as a sign of being a man. Like in the Shona culture in Zimbabwe, they have a saying that gives pride to men infected with STI that goes like: "bhuru rinogona kurwa rinoonekwa nemavanga.", literally transferred to English as: a bull that fights well can only be identified by scares it obtained during the fight.

In a way such key findings of our study indicate two way dimension of how power dynamics in SSA may escalate HIV infection among both males and females. As for women risk of HIV infection through power dynamics has been identified to a result of their limited power to negotiate safer sex in face of their more powerful male sexual partners. On the other side, power dynamics in Africa expose men to HIV infection as they become powerful much that there have less restrictions that may protect them against risk behaviour practices. Our findings have been consistent with those from few related studies that have recently carried out in developed countries. For example a recent study in Spain suggested that power dynamics among adolescent was a likely risk factor for young males to be infected with HIV (Sanchez et al, 2013).

In the case of Zimbabwe those assumed to have less power were significantly associated with many life time sexual partners. This completely contradicted our findings from the other countries. This may be due to the fact that a different proxy variable for measuring power dynamics was used for the country.

Minor findings suggested that increase in age was related to increase in numbers of sexual partners, in the entire region. This should also be expected since increase in age is likely to relate with increase in chances of coming across more sexual partners. Increase in education was also significantly related with having had more life time partners. This corresponds with what has largely been documented in literature that more time spend perusing education especially by those who may be single may lead to eventual accumulation of many life time partners. The understanding is that marriage often leads to sexual exclusivity (Hattori, 2006) than remaining single. Also across entire SSA, suggestions were that men religiously protestants, Muslim in East, Central and West Africa, as well as those who were Pentecostals, Apostolic, and none religious in Southern Africa were each less likely to have many life time sexual partners than those who were Catholics. This may be partly explained by the Catholics doctrines of prohibition of contraceptive use, encouragement of having more kids as fulfilment of biblical quotations like; like more children as many as sand of the seas.

With regards to current discourse, our study finds relevance in discussions around the importance of including men when dealing with issues that affect welfare of women. By indicating that power dynamics which men enjoy may in other instances increase their risk of HIV infection, this may influence men to reconsider their risk behaviours for the good of their health, and equally for the good of their partners as well.

Equally the same, our underpinning theories for this study i.e. Gender and Power Theory and Social Dominance Theory also found relevance from the study. There were clear indications that power was disproportionately distributed in SSA in favour of men. This finding confirmed the former theory. Also true to the assertions of the latter theory, indications were also that SSA societies are made up of heterogeneity classes, and men generally belong to are higher class than that assigned to women.

Conclusion

Power dynamics in SSA which are largely tilted in favour of men may not only expose women to HIV infection. Rather, as suggested by our findings, such power dynamics have equal potential of increasing exposure of HIV infection among men as well. Issues of masculinity among men which come with such power dynamics greatly encourage them into risky sexual behaviours thereby increasing their vulnerable to HIV infection. The study's findings therefore suggest that certain risky behaviours among men in SSA should be understood in the context of excessive power that they have.

Author contribution

Jeremy Gumbo: did the conceptualizing of the topic and discussion; while Clifford Odimegwu did conceptualisation, interpretation and editing; Stella Makoni did reporting of results, and Garikai Chemhaka and Nyasha Chadoka conducted the literature review.

Acknowledgement:

We thank DHS for permitting us to use their data from countries studied.

References

Bingenheimer, J, B. (2010) "Men's multiple sexual partnerships in 15 Sub Saharan African countries: Sociodemographic patterns and implications." Stud Fam Plann 41 (1): p 1-17

Bove, R. and Valegia. C. (2009). "Polygyny and women's health in Sub Saharan Africa." *Social science and medicine*, 68(2009): 21-29

Gregson, S., Garnett, G.P., Nyamukapa, C.A:., Hallet, T.B., Lewis, J.C.J., Manson, P.R., Chandiwana S.K and Anderson. R, M. (2007). "HIV Decline associated with behaviour change in Eastern Zimbabwe." Science Journal, 311 (5761): 664-666

Gumbo J,D (2010) "Is marriage a viable strategy for reducing HIV infection among women in Zimbabwe?" University of Witwatersrand Library

Gupta, G. R. (2000). "Gender, sexuality, and HIV/AIDS. The whats, why, and how." *Canadian HIV AIDS Policy Law Rev*, 5(4): 86-93.

Hattori, M. K. and F.N. Dodoo. (2006). "Cohabitation, Marriage, and 'sexual monogamy' in Nairobi's slums." *Social Science and Medicine*, 64 1067-1078.

Leclerc-Madlala, S. (2004). "Transactional Sex and the pursuit of Modernity." *Social Dynamics* 29(2): 1-21.

Mitsunaga, Tisha M MSc; Powell, Antonia M MSc; Heard, Nathan J MSc; Larsen, Ulla M PhD (2005) "Extramarital Sex Among Nigerian Men: Polygyny and Other Risk Factors" *Epidemiology and Social Science* Vol 39. Issue. 4 p 478- 488

Muthengi, E,. (2009). "Socioeconomic Status and HIV infection among women in Kenya." *African Population Studies Supplement*, 23: 161-176.

Pettifor, A. E., Measham, M., Rees, V, H and Padian, N, S (2004) "Sexual Power and HIV Risk, South Africa." *Emerging Infectious diseases.www.cdc.gov/eid*. Vol. 10, No. 11

Rosenberg, Melanie D. BA; Gurvey, Jill E. MPH; Adler, Nancy PhD; Dunlop, Miranda B. V. BA; Ellen, Jonathan M. MD (1999) "Concurrent Sex Partners and Risk for Sexually Transmitted Diseases Among Adolescents" *Sexually Transmitted Diseases* Vol 26. Issue. 4 p 208-212

Sanchez, M, T, R., Bermudez, M and Buela-Casal, G (2013) "Power dynamics in adolescent couple relationships and risk of sexually transmitted Infections and HIV." *Current HIV Research*, 11, p 536-542

Shisana, O. N. Zungu. N-Dirwayi, Y. Toefy, L. Samba, S. Malik, K. Zuma. (2004). "Marital status and risk of HIV infection in South Africa." *SAMJ*, 94(7): 537-543.

UNAIDS Fact Sheet (2011) Sub Sahara Africa http://www.who.int/hiv/mediacentre/200605-FS_SubSaharanAfrica_en.pdf (downloaded on 16-06-2014)

UNFPA (2011) "Preventing HIV infection: Addressing Gender perspectives in HIV prevention." http://www.unfpa.org/hiv/prevention/hivprev4b.htm (downloaded on 06-06-2014)

UNICEF (2012) Statistics Cameroon http://www.unicef.org/infobycountry/cameroon_statistics.html (downloaded on 05-06-2014)

UNICEF (2012) Kenya http://www.unicef.org/infobycountry/kenya.html (downloaded on 16-09-2014)

UNICEF (2012) Nigeria http://www.unicef.org/infobycountry/nigeria_statistics.html (downloaded on 16-09-2014)

UNICEF (2012) Zimbabwe http://www.unicef.org/infobycountry/zimbabwe.html (downloaded on 16-09-2014)