# EVALUATING AND ESTIMATING DEMOGRAPHIC INDICATORS FROM THE 2016 SOUTH AFRICAN COMMUNITY SURVEY

Eric O. Udjo (PhD)
Independent Consultant
Eric Udjo Consulting
www.ericudjoconsulting.co.za
Email: bororue@yahoo.com

**Abstract**

**Background:** New data are a potential source for updating inputs in assumptions underlying demographic projections. However, it is imperative to evaluate the quality of such new data. It is often the case that the quality of the data cannot be taken at face value owing to reporting errors in the responses to the questions or inappropriate implementation of the questions during a census or survey.

**Objective:** This study evaluates the quality of the demographic aspects of the 2016 South African Community Survey conducted by Statistics South Africa with a view to obtaining plausible demographic indicators from the data.

**Data and Methods.** The study primarily utilised the 2016 South African Community Survey. The evaluation used at least one or combination of internal consistency check, use of demographic models, and comparison with other external sources.

**Results:** The results suggest that from the standpoint of the specific objectives of the 2016 Community Survey, the survey was unnecessary. Although there is no gold standard in judging the accuracy of a survey (or census) the calibration of the survey data to Statistics South Africa's 2016 mid-year estimates is highly questionable. The fertility aspect of the 2016 Community Survey is good and the most reliable of all the demographic aspects of the data. The demographic indicators of childhood mortality from based on reports of children dead of children ever born underestimate the level of childhood mortality in the South African population. Combining children dead of children ever born with orphanhood reports from the survey did not produce plausible estimates of childhood mortality either. The quality of reporting of deaths in households in the Survey was poor. The maternal mortality ratios obtained from the survey data are highly suspect.

**Conclusion:** The integrity of some of the demographic aspects of the 2016 Community Survey data are questionable. Users should therefore be cautious in the use of these aspects of the data.

# EVALUATING AND ESTIMATING DEMOGRAPHIC INDICATORS FROM THE 2016 SOUTH AFRICAN COMMUNITY SURVEY[1]

## INTRODUCTION

### Background

When new census or survey data become available, demographic indicators as well as the assumptions underlying population projections, should be revised in light of the new data. It is well established however that census and survey data usually contain errors hence the information from such data may not be taken at face value. The two forms of error often present in such data are coverage and content errors. The present study focuses on survey data, consequently, the two types of error are sampling and content errors. While sampling errors are easy to handle using appropriate statistical techniques, content errors are more difficult to detect and correct for in survey data. Despite this, it is critical that evaluation of the new data be carried out to enable estimation of plausible demographic indicators that feed into population projections and activities that require accurate demographic estimates. Statistics South Africa in 2016 carried out a national community survey (2016 CS). The survey data constitute a potential source for current demographic estimates that could be used among others, as inputs in updating or revising population projections for the country.

### Objectives

Thus the objective of this study was to evaluate the quality of the demographic aspects of the 2016 Community Survey (2016 CS) conducted by Statistics South Africa (see Statistics South Africa 2016a; 2016b) specifically focusing on age distribution, age at marriage, fertility, mortality and net migration with a view to obtaining plausible demographic indicators from the data.

---

[1] This paper is an extract from a broader study undertaken by this author for the Bureau of Market Research (see Udjo 2017).

## Purpose and Sample in the 2016 Community Survey

The stated purpose of the 2016 CS was to bridge the gap between censuses in South Africa in providing demographic and socio-economic data at lower geographical levels. Consequently, the stated specific objectives were:

1. To provide an estimate of the population count by local municipality.
2. To provide an estimate of the household count by local municipality.
3. To measure demographic factors such as fertility, mortality and migration.
4. To measure socio-economic factors such as employment, unemployment, and the extent of poverty in households.
5. To measure access to facilities and services such as piped water, sanitation and electricity for lighting (Statistics South Africa 2016b).

The target population for the survey was the non-institutional population residing in private dwellings (Statistics South Africa 2016b). Thus, homeless people were excluded from the survey. Very small enumeration areas (EAs) that were part of the target population were also excluded. The sampling frame was the geo-referenced dwelling frame based on the 2011 census EAs. The realised sample after editing of the data obtained by this author from the relevant data files consisted of 3,328,867 persons in 984,627 households although Statistics South Africa (2016b) in its report on the survey noted that the final sample realised for weighting after all the necessary checks and validation consisted of 1,422,928 households. Therefore, the realised sample size in the data file appears contradictory to what was stated in the Statistics South Africa's report.

## METHODS

## General Approaches in the Data Evaluation

At least one or a combination of the following approaches was used in evaluating and estimating demographic indicators from the 2016 CS. ***Internal Consistency Check:*** This entailed examining the holistic coherence of the data. For example, comparing indirect estimates of childhood and adult mortality with those from household deaths in the data. ***Use of Demographic Models:*** Demographic models are useful in identifying and separating error deviations from the real features of the data, for example the use of the relational Gompertz model to detect and adjust for errors in the fertility reports. ***Comparison with other External Sources:*** This entailed comparing one or several aspects of the data with an external source but bearing in mind that there is no gold standard in assessing the quality of particular data. The external source/sources may also contain errors.

**Technical Analysis and Estimation.**

***Age Composition***

The procedure employed in correcting and smoothing the age misreporting in the five-year age distributions from the 2016 CS entailed comparing the logit transformations of the reported cumulated age distribution with those of an appropriate stable population. The stable population was selected from Carrier's and Hobcraft's (1971) list of two-parameter stable populations. The selection was based on the closeness of the ratio of the number of persons aged 0-14 years to the number of persons aged 15-44 years as well as the gross reproduction rate in the observed 2016 CS distributions to the corresponding values in the stable population. Using the formulae given by Brass (1971) the age distributions were transformed as follows.

$Y_x = ½ \log_e(((1-P_x)/P_x))$ …………….(1)

$Y_{sx} = ½\log_e(((1-P_{sx})/P_{sx}))$ ………..…(2)

Where $Y_x$ and $Y_{sx}$ are respectively the logits of the reported and stable age distributions. $P_x$ and $P_{sx}$ are respectively the cumulated proportions under age x in the reported and stable populations. The $Y_x$ values were then plotted against the $Y_{sx}$ values and a straight line fitted to the "best" points using the least squares method. The smoothed age distribution was then computed as:

$Y'_x = \alpha + \beta Y_{sx}$ …………………(3)

Where $Y'_x$ is the logit smoothed age distribution, $\alpha$ and $\beta$ are respectively the intercept and slope of the fitted line. Finally, the expression:

$S_x = 1/(1 + e^{2Y'x})$ …………….(4)

produced the re-transformation of the fitted values.

Where $S_x$ is the re-transformation i.e. anti-logit of $Y'_{x.}$ De-cumulating $S_x$ produced the smoothed proportion of persons in in each five-year age group including the open age interval (Udjo, 2014).

*Age at First Marriage*

The mean age at first marriage was indirectly estimated from the observed proportions currently single. The value is called the singulate mean age at marriage,SMAM (Hajnal 1953). Assuming all first marriages took place by age 49, SMAM is computed as:

$$SMAM = \sum_{x=0}^{49} \{P_x - (50P_{45-54})\} / (1 - P_{45-54}) \text{ ............(5)}$$

Where $P_x$ is the proportion single at age *x* (Udjo, 2014). If persons living together are treated as married lowers the value of SMAM compared with the value if such persons were treated as never married. In this study, persons living together were treated as married. To minimise the effect of age misreporting on the value of SMAM, Van De Walle (1968) proposed that the entire population should be dichotomised into single and non-single. This is then translated into an age at first marriage by linear interpolation on the reported age distribution (i.e. the age at which the population cumulated since birth, equals the proportion single (Van De Walle, 1968 cited in Udjo, 2014).

*The P/F ratio and relational Gompertz models in evaluating and estimating fertility*

A starting point in evaluating data on fertility reports is usually the use of Brass (1971) P/F ratio model or its variant followed by fitting the relational Gompertz model (Brass, 1981) to the data. These models were fitted to the data on children ever born and births in the last 12 months tabulated by age of women in the reproductive age. Although the the P/F ratio model assumes that fertility has been constant in recent years which may not be the case in certain populations, it is useful in evaluating the quality of the reports. However, where this assumption is violated total fertility rate (TFR) may not be estimated from the data. Instead, the relational Gompertz model would be more appropriate in estimating TFR since the model relaxes the assumption of constant fertility. The P/F ratio model as given by Brass (1971) is expressed as

$$F_i = 5\sum_{j=0}^{i-1} f_j + kifi \text{ ...........(6)}$$

Where *fi* is reported ASFR, for a five-year age group, *ki* is a multiplier. Taking the ratios of the $P_i s$ and $F_i s$ constitute a measure of the consistency and accuracy of the reported children ever born and births in the last 12 months. The relational Gompertz model is expressed as

$z(x) - e(x) = \alpha + 0.48(\beta - 1)^2 + \beta(gx)$ …………(7)

where $z(x)$ is

$-ln(ln\ F_x / F_{(x+5)})$ ……………………….(8)

$F_x$ is cumulated ASFR up to age x, *e(x)* and *g(x)* are standard values, 0.48 is a constant, $\alpha$ and $\beta$ measure the location and spread of the fertility distribution (Brass, 1981). For mean parity (average children ever born),

*z(x)* in equation (7) is replaced by z(i) and defined as

$-ln(ln\ P_i\ /P_{(i+1)})$ ………………..(9)

Where $P_i$ is the mean parity in a five-year age group *i* (see Booth, 1979; Zaba, 1981).

***Indirect estimation and evaluation of childhood and adult mortality***

Childhood mortality was indirectly estimated using Brass (1971) model expressed as

$q_x = D_i * K_i$ …………………………(10)

Where $q_x$ is the probability of dying between birth and exact age x, $D_i$ is the proportion of deaths among children ever born to women in age group *i,* and $K_i$ is a multiplier. AIDS was incorporated using the INDEPTH (2004) standard life tables. For adult female and male mortality the estimation equation from the responses to the orphanhood questions (Is your biological mother still alive? Is your biological father still alive?) was

$l_{B+N}/l_B=W_N({}_5P_{N-5}) + {}_5P_N\ (1-W_N)$ ……………(11)

where $l_{B+N}/l_B$ is the probability of surviving from a base age B, to B+N.

Where *N* is the central age between two adjacent five year age groups; ${}_5P_{N-5}$ is the proportion in the age group *N-5* to *N* having a surviving parent; ${}_5P_N$ is the proportion in the age group *N* to *N+5* having a surviving parent; and $W_N$ is a weighting factor (see Brass 1971). The levels of adult mortality $\alpha$, in the logit system were then derived as

$\alpha = ½log_e\ [1 + ({}_NP_B - 1/l_{sB+N})/(1-{}_NP_B)]$ …………(12)

Where ${}_NP_B$ is the probability of surviving from base age *B* to age *B+N;* $l_{s(B+N)}$ is the probability of surviving to exact age *B+N* in a standard life table (Brass & Bamgboye, 1981). The reference dates of the $\alpha$ values were estimated using Brass' and Bamgboye's (1981) method. AIDS was incorporated into the estimates using the INDEPTH (2004) life tables. The separate childhood and adult mortality estimated estimates were spliced together to produce parameter estimates for a two-parameter life tables. The splicing procedure is described in Udjo (2008).

***Evaluating and estimating mortality from reported deaths in households***

A major concern in estimating mortality from survey/census data on reported deaths in households is completeness of the reported number of deaths. This is also a concern in estimating mortality from registered deaths in a vital registration system. Completeness may be under- or over-reporting of deaths. In this study, the evaluation tool for the assessing the completeness of reporting of deaths in in the 2016 CS was the original Brass Growth Balance method. The method is based on the linear relationship of deaths and age distributions observed by Brass (1981) and expressed as

$$N(x)/N(x+) = r + D(x+)/N(x+) \quad \ldots\ldots\ldots\ldots\ldots(13)$$

Where $N(x)$ is the number of persons at exact age $x$, $N(x+)$ is the total number of persons above age $x$, $D(x+)$ is the total number of deaths occurring to persons aged $x$ and over and $r$ is the growth rate. Given that there would be an error pattern in the distribution of deaths by age, the above equation may be re-written as

$$N(x)/N(x+) = r + k[D(x+)/N(x+)] \quad \ldots\ldots\ldots\ldots(14)$$

Where $r$ (an estimate of the growth rate) in the above equation is the intercept of a straight line fitted to the plot by age, $k$ is the slope of the fitted line, a coefficient of the estimated ratio of true to reported deaths (a factor representing the completeness of reporting of deaths, see Hill 1987).

***Evaluating and estimating mortality from reported deaths in households***

A maternal death is defined as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes (World Health Organisation 2017). If the data were perfect, maternal mortality ratio, MMR (the commonly used measure of maternal mortality) may be computed as:

$$MMR = (D/B) * k \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(15)$$

Where MMR is the period maternal mortality ratio, $D$ is the number of maternal deaths in the period, $B$ is the number of live births in the period and $k$, a constant, usually 100,000 (Udjo & Lalthapersad-Pillay, 2014). However like other demographic data, maternal mortality data should be treated as suspect until proven otherwise. Hence the parameters in the above equation need to be adjusted for errors. The approach used in the adjustment adjustments in this study was based on the relationship between maternal mortality rate, and the maternal mortality ratio, MMR as defined by Stanton, Abderrahim and Hill (1997) as:

$$Mmrate/GFR = MMR \quad \ldots\ldots\ldots\ldots\ldots\ldots (16)$$

Where $Mmrate$ is the maternal mortality rate for a specified period and $GFR$ is the General Fertility Rate for the period. Maternal mortality rate is defined as:

$$Mmrate = (D/Women_{15\text{-}49}) * 1000 \quad \ldots\ldots\ldots\ldots (17)$$

Where $D$ is the number of maternal deaths and $Women_{15\text{-}49}$ is the mid-year population of women aged 15-49 years (see Stanton, et al., 1997). The adjusted maternal mortality ratio was then estimated as:

$$MMR' = Mmrate'/GFR' \quad \ldots\ldots\ldots\ldots\ldots\ldots (18)$$

Where *MMR'* is the adjusted maternal mortality ratio per 100,000 live births, *Mmrate'* is the adjusted maternal mortality rate and *GFR'* is the adjusted general fertility rate. The adjusted general fertility rate was estimated as:

$GFR' = B' / Women_{15\text{-}49}$ ………….. (19)

Where *B'* is the adjusted number of live births in the last 12 months (before the 2016 Community Survey ) and $Women_{15\text{-}49}$ is the number of women aged 15-49. The adjusted number of live births in the last 12 months, *B'* was estimated from a relational Gompertz model as:

$$B' = \sum_{x=15}^{49} (Women(x) * ASFR(x)') \quad \text{…………(20)}$$

Where $Women(x)$ is the number of women in a five-year reproductive age group, *x*, and $ASFR(x)'$ is adjusted age specific fertility in a five-year reproductive age group, *x*. The adjusted age specific fertility rates were estimated as:

$ASFR(x)' = TFR' * F(x)'$ ………………(21)

Where $ASFR(x)'$ is the decumulated age specific fertility rate for women in the reproductive age group aged x and becomes adjusted age specific fertility rate $ASFR(x)'$ in the relational Gompertz model, *TFR'* is the adjusted total fertility rate for the population in the relational Gompertz model and *F(x)'* is the model cumulative fertility rate up to age x in the relational Gompertz model. The relational Gompertz model was described in a previous section. The maternal mortality rate was then adjusted as:

$Mmrate' = ((D*c)/Women_{15\text{-}49}) * 1000$ ………………(22)

Where *Mmrate'* is the adjusted maternal mortality rate, *D* is as defined earlier and *c* is a coefficient for incomplete reporting of deaths in households assumed to be the same for all causes of death. Though not a realistic assumption, but was most practicable as not all causes of death were collected in the 2016 CS. It is not possible to assess completeness of registration for each cause of death even in a vital registration system given present knowledge on assessment of completeness of death registration.

***Evaluating and estimating net migration***

The 2016 CS included a battery of questions on migration. Evaluating and estimating net migration from the migration questions firstly entailed producing migration matrix tables from the responses to the questions. The matrix tables consisted of the following cross-tabulations: (1) Province of usual residence by year moved to current residence by sex; (2) Province of previous residence by year moved to current residence by sex; (3) Province by year in which emigrant left South Africa by sex of emigrant. Net provincial migration was then computed as

$$NM^{t}_{pjs} = IM^{t}_{pjs} - OM^{t}_{pjs} \quad \text{................(23)}$$

Where *NM* is the volume of net migration for each of the years *t*, 2011 – 2015 in a province, *pj* for a particular sex, s. The year 2016 was excluded from the analysis because the migration reported was not for a complete calendar year for that year. *IM* is the volume of in-migration for each of the years *t*, 2011 – 2015 into a province, *pj* for a particular sex, *s* including foreign-born while *OM* is the volume of out-migration for each of the years *t*, 2011 – 2015 from a province, *pj* for a particular sex, *s* including foreign-born. Next, total net migration was computed as

$$TNM^{t}_{pjs} = NM^{t}_{pjs} + E^{t}_{pjs} \quad \text{..............(24)}$$

Where $TNM^{t}_{pjs}$ is the total volume of net migration for each of the years *t*, 2011 – 2015 in a province, *pj* for a particular sex, *s* and $E^{t}_{ps}$ is the volume of emigration for each of the years *t*, 2011 – 2015 from a province, *pj* for a particular sex, *s*. The average annual percent growth rate in total net migration during the period 2011 - 2015 was then estimated as

$$RTNM^{t+n}_{pjs} = \{LN[(TNM^{t,2}_{pjs}/ TNM^{t,1}_{pjs})/(t_{,2} - t_{,1})]\} * 100 \text{......(25)}$$

Where $RTNM^{t+n}_{pjs}$ is the percent average growth rate in total volume of net migration during specified time periods in a province for a particular sex, *s*. $TNM^{t,2}$ and $TNM^{t,1}$ are respectively the total volume of net migration in the later and initial periods, $t_{,2} - t_{,1}$ is the number of years between the later and initial periods and *LN* is natural logarithm. The growth rate estimation was derived from the exponential growth formula of population growth.

Since the 2016 CS was a sample survey, total net migration at provincial population level was estimated as follows. First the ratio of estimated total net migrants for a particular period to the realised sample in a province was computed as follows:

$$r^{t}_{pjs} = TNM^{t}_{pjs}/n' \quad \text{...........(26)}$$

Where $r^{t}_{pjs}$ is the ratio of the estimated total net migrants for a particular year in a province for a particular sex and *n′* is the realised sample in that province. It was assumed that if a similar survey had been carried out many times in 2016, this ratio would be normally distributed with averages similar to those observed for the periods 2011 to 2015. On the assumption that the ratio of total net migrants to the realised sample in 2016 is similar to the ratio of net migrants to independent estimates of the population for different years, absolute number of total migrants at provincial population level was then estimated as:

$$PNM^{t}_{pjs} = r^{t}_{pjs} * P^{t}_{pjs} \quad \text{......... (27)}$$

Where $PNM^t_{pjs}$ is the provincial level total number of net migrants in a particular year for a particular sex, $P^t_{pjs}$ is an independent estimate of the provincial population for a particular year for a particular sex. In the present study, the independent estimates of the population were those provided by the Bureau of Market Research (See Udjo, 2015). Note that the 2016 Community Survey realised sample included migrants. The independent estimates of the provincial population also included assumptions about net migration in the net migration estimates. Theoretically, errors in the reported number of migrants in the 2016 CS realised sample as well as errors in the accuracy of the migration estimates in the independent population estimates should have biases in the accuracy of the estimates of $PNM^t_{pjs}$ but the magnitude of the biases cannot be ascertained.

## RESULTS

### Weighting of the Data

The 2016 CS data were weighted according to the design that were derived during the sampling process adjusting for additional dwelling units that were identified and enumerated but which were not in the sampling frame. The weighting process also adjusted for small EAs that were excluded from the sample and for non-response. Finally, the adjusted design weights at persons and household levels were calibrated to a demographic model – Statistics South Africa's 2016 mid-year estimates – using age, sex, and population group as controls at national, provincial and municipal levels (Statistics South Africa, 2016b). A number of issues arise from the weighting procedure.

***Implications of excluding institutions and small EAs from the sample:*** Arguably, the size of the excluded populations was probably small. However, technically, the results from the enumerated 2016 CS are not nationally representative of the total population of South Africa despite the calibration undertaken on the data. In other words, the variance in a specific indicator between the excluded populations and surveyed populations may be large. The calibration process (explained in a later section) makes the results even more questionable.

***Inconsistency in definition of In-scope and out-of-scope:*** The technical report by Statistics South Africa on the 2016 CS states: The institutional and transient population are out of scope for 2016 Community Survey. At the same time in the same report, Statistics South Africa (2016b) states:

*"The final result codes for each record were mapped to one of three final response status categories … where 1 = Respondent (i.e. having a completed or partly completed questionnaire for the household), 2 = Non-respondent (i.e. where the household did not respond and/or there was no questionnaire completed), and 3 = Out-of-scope (i.e. where no eligible household was found to be enumerated)" (Statistics South Africa, 2016b: 34). Category 3 above is inconsistent to the earlier statement where out-of-scope was defined as the institutional and transient population. The technical report further states that the "out of scope rate is defined as the proportion of DUs in which no eligible household was found to the total number of sampled DUs (including any additional DUs identified during data collection."*

This contradictory definition of out-of-scope has implication for the response rates. Re-classifying certain households that were not institutional or transient population as out-of-scope technically implies that the response rates provided in the report are biased upwards (i.e. exaggerated) since the re-classification means they were excluded from the numerator in the computations of the response rates.

***Potential bias in levels of indicators arising from calibration:*** The demographic model that the 2016 CS was calibrated to is the entire population of South Africa as estimated by Statistics South Africa for the year 2016. However, since institutional or transient populations were out of scope in the sampling for the 2016 CS, it means these groups were smuggled into the data through the back door (i.e. through the calibration process). By so doing and in computing indicators using the weighted and calibrated data, values of the computed indicators would be biased since they would be included in the denominators but not in the numerators used in computing indicators. Unbiased estimates would only be possible all other things equal if the analyst makes conscious efforts to identify and exclude the institutional and transient population during the tabulations of variables used in computing indicators.

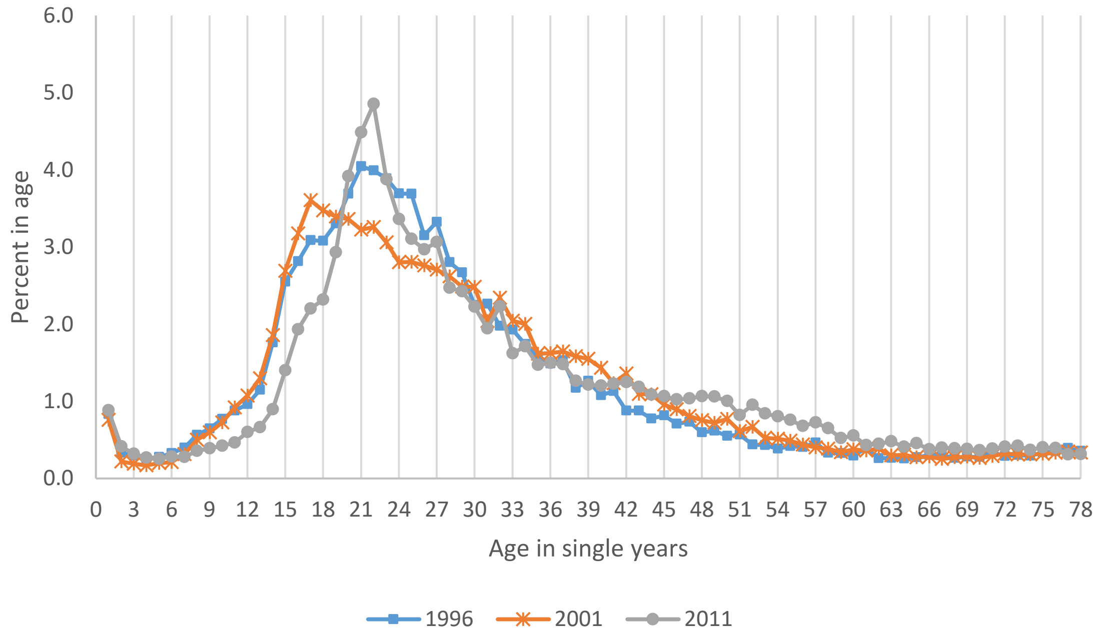
Through personal communication, this author was informed that Statistics South Africa accounted for the institutional and transient population in the 2016 CS (as in the 2007 CS) through the calibration process using the proportions of these populations based from the last census. The implicit assumption underlying this is that the distributions (demographic and socio-economic characteristics) of these populations are constant in all census years. Figures 1-2 show the outcome of two approaches in checking the veracity of this implicit assumption. In the first approach, the single year age distributions of the institutional population in 1996, 2001 and 2011 based on the unweighted census data (to eliminate post enumeration survey (PES)) biases are shown in Figure 1. In the second approach, the ratio of the institutional to non-institutional population in each single year based on the unweighted census data are shown in Figure 2. If the implicit assumption of constant distribution in the census years were true, then one should have a single curve for all the three census years. Figures 1-2 clearly show that the distributions are not the same as depicted by the three separate curves in each graph except for the youngest ages where there is some convergence. Therefore, accounting for the institutional and transient population in the 2016 CS using the proportions of this population in the previous census in the calibration process is questionable.

***Calibrating the design weights to a demographic model illogical*:** As earlier noted, survey (and census) data provide among others, demographic information for updating demographic indicators that serve among others, inputs in improving or updating demographic models. This was the spirit behind Statistics South Africa's decision in 1998 not to adjust the post-apartheid 1996 census to a demographic model contrary to the practice during the apartheid era. In that era census results were adjusted to a demographic model as seen in the following statements. Arguing why the 1996 census results were the best ever, Statistics South Africa (1998) noted that:

> "*During the 1991 census, roughly half of the population were incompletely counted, or not counted at all. Instead, a demography model reaching back 20 years was used to estimate the size of the population*". Statistics South Africa (1998) further noted that: *"We have recalled how the 1991 population estimates, in total and by population group, reflected a demographic model. The actual counts from enumeration from aerial photography and sample surveys were adjusted to fit the model; not the other way round… By contrast, the estimates of census '96 represent a countrywide count adjusted on the basis of a country wide PES. They take the empirical data, rather than the match to a demographic model, as primary".*

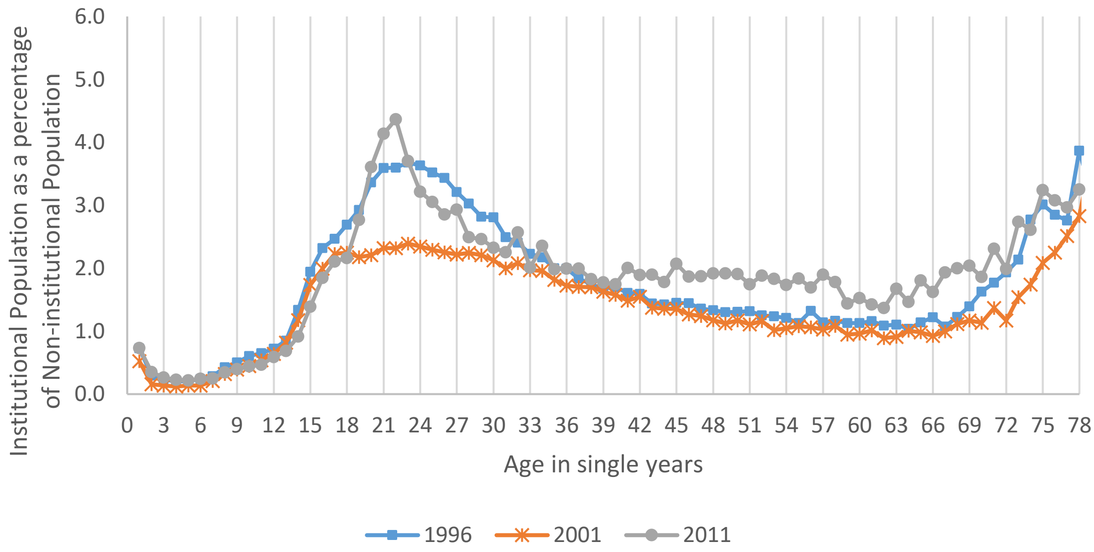
Statistics South Africa's mid-year estimates is a demographic model that is based on assumptions about the future course of fertility, mortality and net migration. These assumptions are not empirical observations whereas survey data are empirical observations. The calibration of the 2016 CS to a demographic model is a contradiction of Statistics South Africa's position in 1998 and illogical.

**FIGURE 1: PERCENT SINGLE YEAR AGE DISTRIBUTION OF THE INSTITUTIONAL POPULATION, BOTH SEXES (UNWEIGHTED DATA)**




Sources: Author's computation from the 1996, 2001 and 2011 Census data.

**FIGURE 2: RATIO OF INSTITUTIONAL TO NON-INSTITUTIONAL POPULATION, BOTH SEXES (UNWEIGHTED DATA)**




Sources: Author's computation from the 1996, 2001 and 2011 Census data

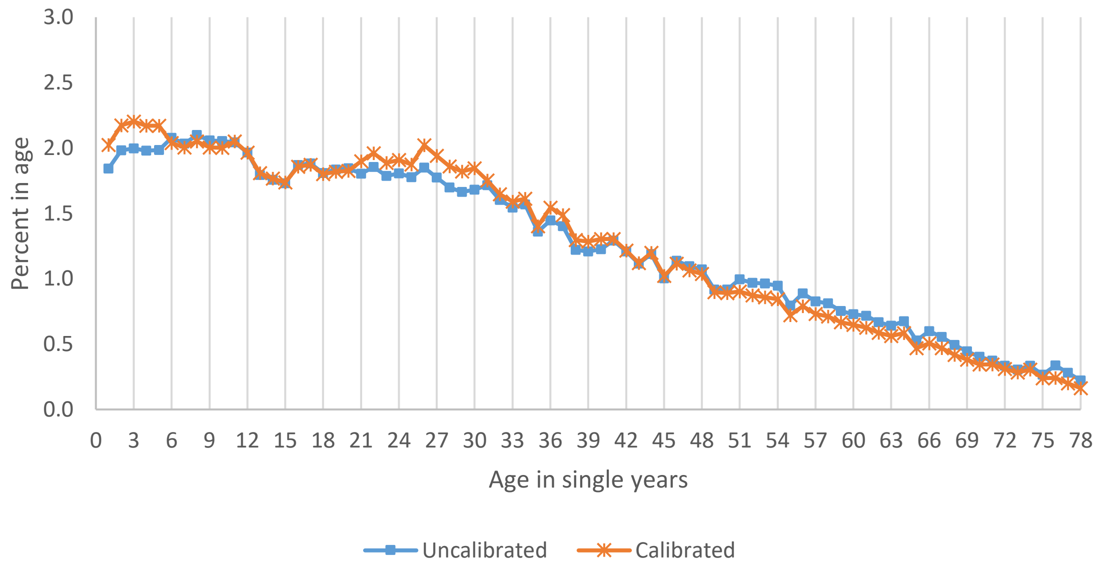
***Distortions arising from the calibration:*** Calibration should reproduce the 2016 CS but this was not the case. The calibration distorts the data as seen from the following. Summary measure of age from the 2001 and 2011 census data suggests that the median age of the population of South Africa is increasing (Table 1). This makes sense in view of the declining fertility and gradual ageing of the population. However, comparing the median age of the population in the uncalibrated and calibrated 2016 CS suggests otherwise. The calibrated 2016 CS data exhibit a younger median age compared with uncalibrated 2016 Community Survey data. This is largely because in the process of calibrating the data, proportionately more persons were assigned to ages 0 – 6 and 19 – 32 (Figure 3). Therefore, the calibration produced two different curves instead of one curve in the age distribution as seen in figure 3.

**TABLE 1: IMPLIED MEDIAN AGE (YEARS) OF THE POPULATION FROM 2001, 2011 CENSUS AND 2016 COMMUNITY SURVEY, BOTH SEXES**

| 2001 Census | 2011 Census | 2016 CS uncalibrated | 2016 CS calibrated |
|---|---|---|---|
| 23 | 25 | 26 | 25 |

Source: Author's computation from 2001, 2011 Censuses and 2016 CS data

**FIGURE 3: PERCENT SINGLE YEAR AGE DISTRIBUTION, BOTH SEXES 2016, COMMUNITY SURVEY CALIBRATED AND UNCALIBRATED: NATIONAL**

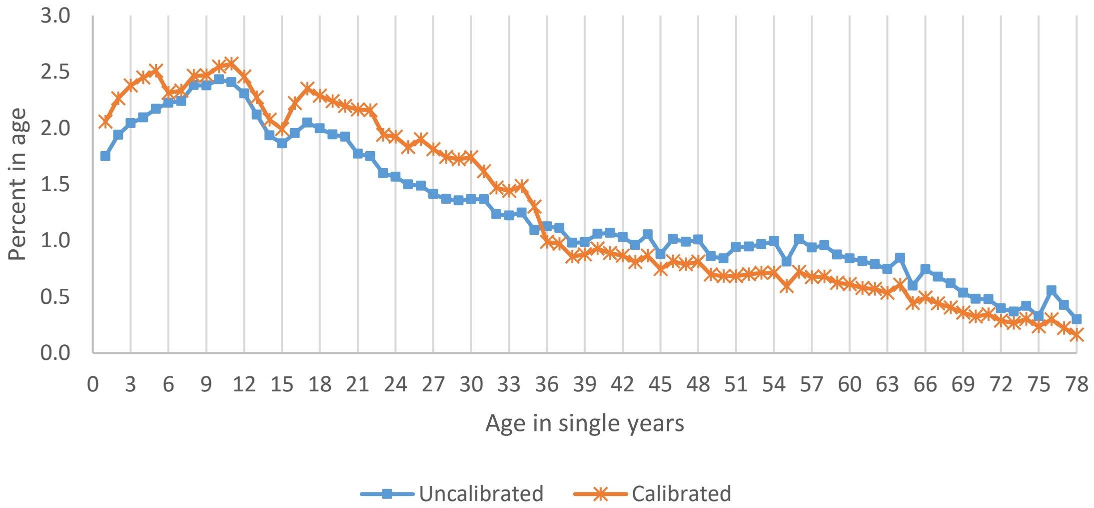



Source: Author's computation from 2016 CS data

Regarding the distribution at provincial level, for brevity of presentation, only the calibrated and uncalibrated single year age distribution for one province is shown – that of the Eastern Cape (Figure 4). It can be seen from the graph that proportionately more people were assigned to the age group 0-35 years, and less people assigned to the age group 36 years and over in the calibrated figures compared with the corresponding uncalibrated figures. All the provinces exhibited distortions comparing the calibrated and uncalibrated single year age distributions.

In view of the fore going it is inappropriate (and probably misleading) to use the calibrated 2016 CS data in estimating demographic and socio-economic indicators. The results presented below were therefore based on the uncalibrated data.

**FIGURE 4: PERCENT SINGLE YEAR AGE DISTRIBUTION BOTH, SEXES 2016, COMMUNITY SURVEY CALIBRATED AND UNCALIBRATED: EASTERN CAPE**



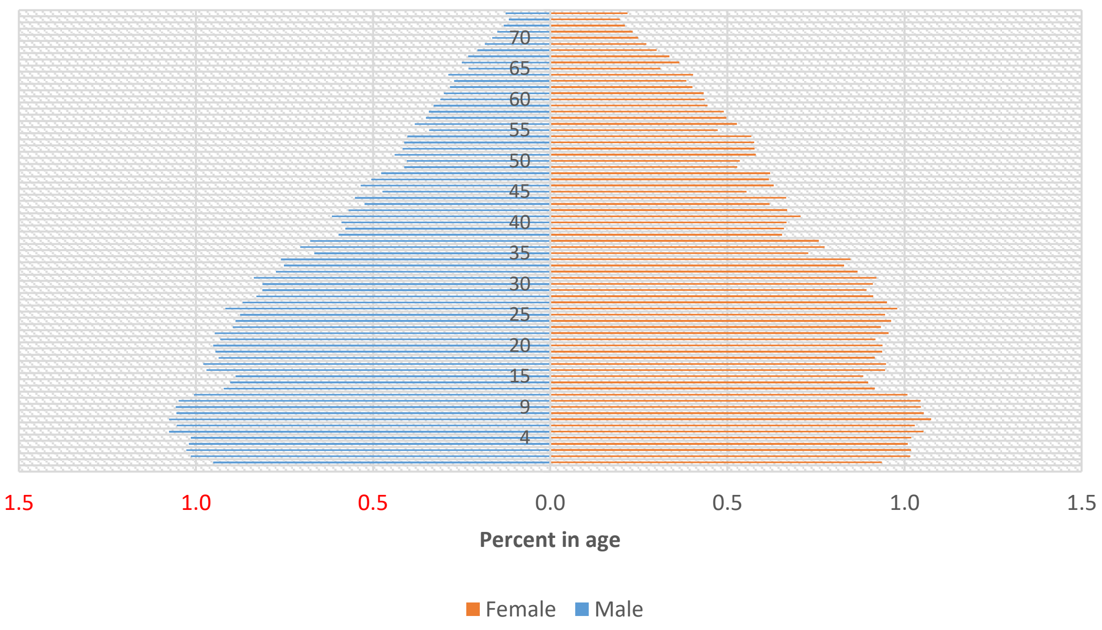

Source: Author's computation from 2016 CS data

**Age Composition**

The single year population pyramid at national level from the 2016 CS is shown in Figure 5 and as seen in the graph, the age heaping in the 2016 CS does not conform to the usual pattern of systematic heaping in ages ending in zeros, fives and even numbers. Instead, the spikes exhibit a mixed pattern of heaping in ages ending in odd numbers (for example 17 and 41) as well as ages ending in fives and even numbers among adult males. Among females the pattern was ages ending in odd numbers (17, 31, 41, 51) as well as in fives and even numbers. The deep indentations in Figure 5 between ages 10 and 15 suggest avoidance in the reporting of ages 10, 11, 12, 13, 14, 15 by many of the respondents though heavier mortality or emigration among children born 10-15 years before the survey is a theoretical possibility.

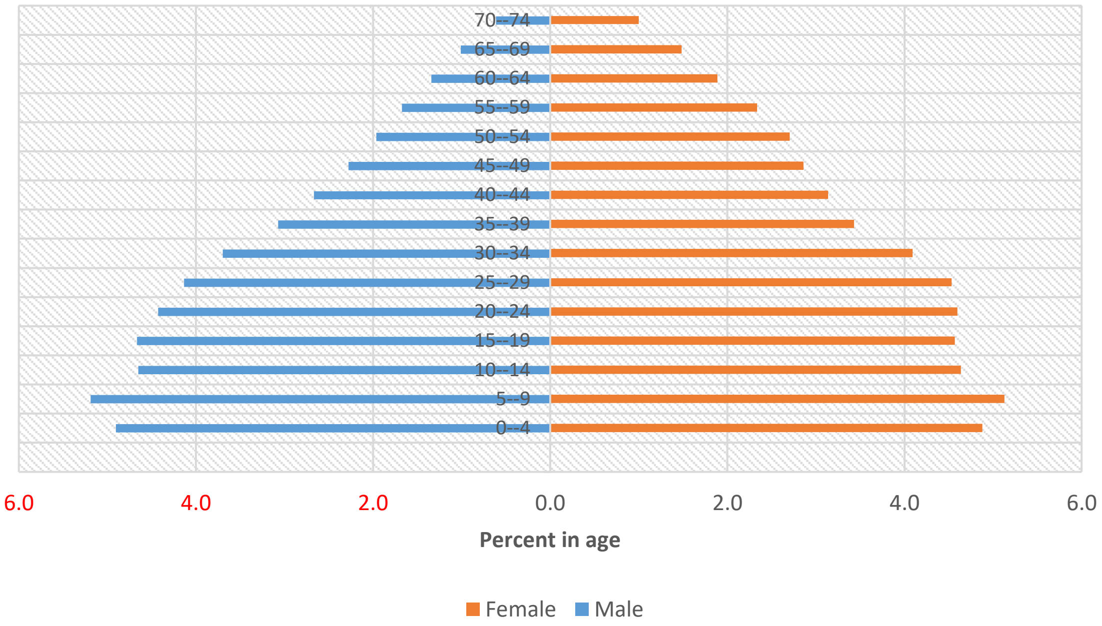
Figure 6 shows the reported five-year age distribution plotted as a pyramid at national level. The following are evident. (1) Indentation at the base of the pyramid corresponding to persons aged 0-4 years; (2) indentation in the pyramid corresponding to persons aged 10-14 years; (3) a bulge in the pyramid corresponding to females aged 15-29 years. One may invoke several theoretical explanations for these patterns. As in many African countries, indentation at the base of the population pyramid is a common feature in Statistics South Africa's data and is most likely due to underreporting of persons aged 0-4 years during the 2016 CS. The indentation corresponding to the age group 10-14 years in the population pyramid may also be due to underreporting of persons in this age group during the 2016 C S while the bulge corresponding to females aged 15-29 years may be due to age shifting – some females older than 15-29 years reported as being in this age group.

**FIGURE 5: PERCENT REPORTED SINGLE-YEAR POPULATION PYRAMID 2016 CS, NATIONAL**




Source: Author's computation from 2016 CS data

**FIGURE 6: PERCENT REPORTED FIVE-YEAR POPULATION PYRAMID 2016 CS, NATIONAL**




Source: Author’s computation from 2016 CS data

The reported five year age distribution was compared with a demographic model as described in the methodology section. The result at national level is summarised in Figure 7. A negative percentage in Figure 7 denotes that the percentage of persons reported in an age group during the 2016 Community Survey was lower than the corresponding percentage in the demographic model. Thus, Figure 7 suggests that the number of persons aged 0-4, 10-14, 15-19, and 35-54 were under-reported during the 2016 CS. It also suggests that the under-reporting was higher for females in the age groups 0-4, 15-19 and 35-49 than for males. The demographic model furthermore suggests there was substantial over-reporting of persons aged 5-9, 25-29 and 65-69 during the 2016 CS.

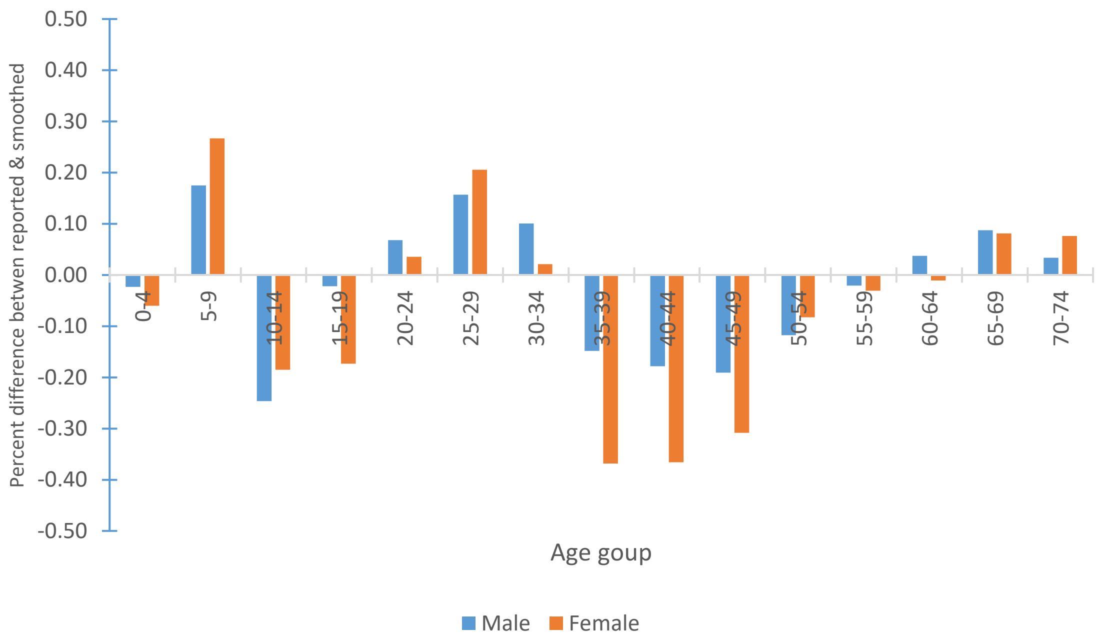
A similar approach to the foregoing may be used to evaluate the quality of the reported provincial single and five-year age distributions. However, owing to the complicating impact of internal and international migration, this was not attempted since drawing firm conclusions from such evaluation is more difficult at provincial levels.

**Sex Ratios**

The reported overall sex ratio of 89 in the 2016 CS (see last row, column 2 of Table 2) is implausibly low given fertility, mortality and net migration levels and trends. This suggests that males were under-reported during the 2016 CS as in most Statistics South Africa’s surveys and censuses. The reported overall sex ratio in the 2011 census was 93 although the final weighted dataset that was put out by

Statistics South Africa gave an overall sex ratio of 95. Though more plausible, it is not clear how Statistics South Africa came to an overall sex ratio of 95 in comparison with overall sex ratios of 93 and 92 respectively from the final weighted datasets of the 1996 and 2001 censuses (see table 2). Did Statistics South Africa reclassify women as men in the weighting procedure? In justifying the low sex ratio of 93 from the 1996 census, Statistics South Africa pointed out that the adjusted high sex ratio from the 1991 census was because "… a large proportion of women were reclassified as men" by Loubser and Van Wyk (Statistics South Africa 1996). Did Statistics South Africa do the same thing in the published data with regard to the 2011 Census in its weighting procedure as well as in the 2016 CS? Overall sex ratio in the absence of social political upheavals such as wars, changes very slowly in a population.

**FIGURE 7: DIFFERENCE BETWEEN PERCENT REPORTED AND SMOOTHED FIVE-YEAR AGE DISTRIBUTION 2016 CS, NATIONAL**




Source: Author's computation from the 2016 CS data

**TABLE 2: OVERALL SEX RATIOS IN SOUTH AFRICA'S CENSUSES AND 2016 CS**

| | Reported | Weighted* |
|---|---|---|
| 1996 Census | 92.2 | 92.7 |
| 2001 Census | 90.1 | 91.7 |
| 2007 Community Survey | 93.0 | 93.3 |
| 2011 Census | 93.2 | 94.8 |
| 2016 Community Survey | 88.8 | 95.9 |

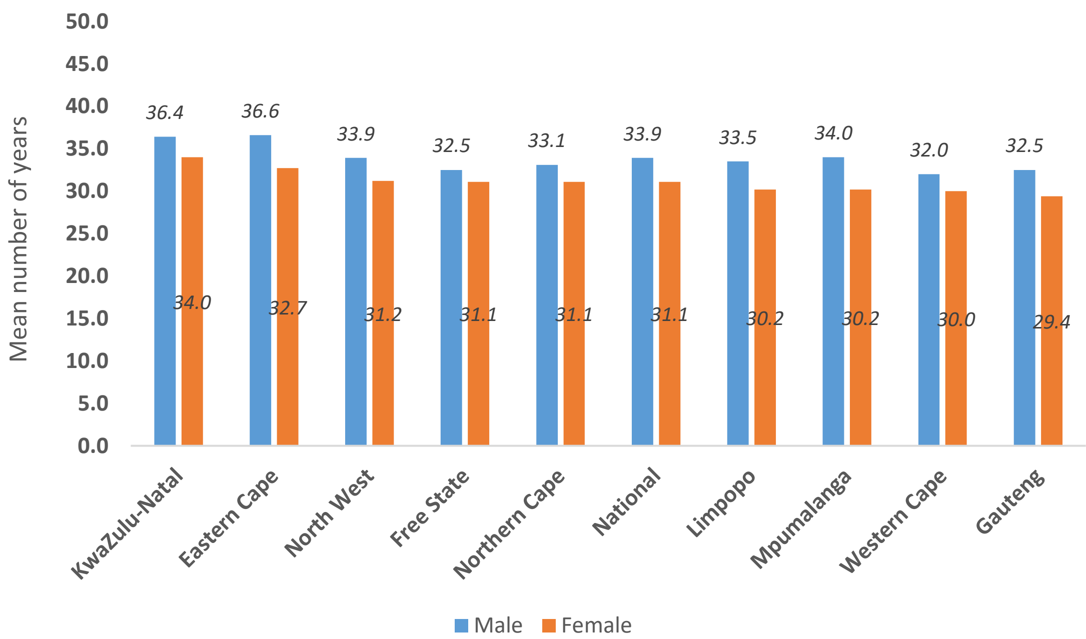
*Weighted by Statistics South Africa
Source: Author's computation from censuses and 2016 CS data

**Age at First Marriage**

There was no direct question on age at first marriage in the 2016 CS hence was estimated from the information on current marital status resulting in unadjusted SMAM as described in the methods section. Figure 8 summarises the results. The unadjusted age at first marriage nationally (31.1 years for females, 33.9 years for males) is very high by European standards - Sweden 33.0 years (Statistics Sweden, 2015); Switzerland 29.6 years (UNECE, undated) - let alone by African standards - Ethiopia 16.5 years (UNICEF, 2016); Kenya 20.2 years (Kenya Bureau of Statistics, 2015). This casts doubt on the accuracy of the reports. As seen from figure 6 the provincial levels of unadjusted SMAM are also high. The values of SMAM in the 2016 CS as in other surveys and censuses in South Africa may have been exaggerated by a number of factors including (1) incorrect declaration of current marital status during the survey i.e. some persons who were probably divorced, separated or widowed reporting they had never married. (2) Incorrect statement of age.

It is difficult to assess the magnitude of incorrect marital status declaration and its impact on the SMAM. However, the impact of age errors on SMAM may be assessed by computing the age at which the reported cumulated percentage age distribution equals the percentage of the population single as described in the methods section. The results denoted as adjusted SMAM are summarised in Figure 9.

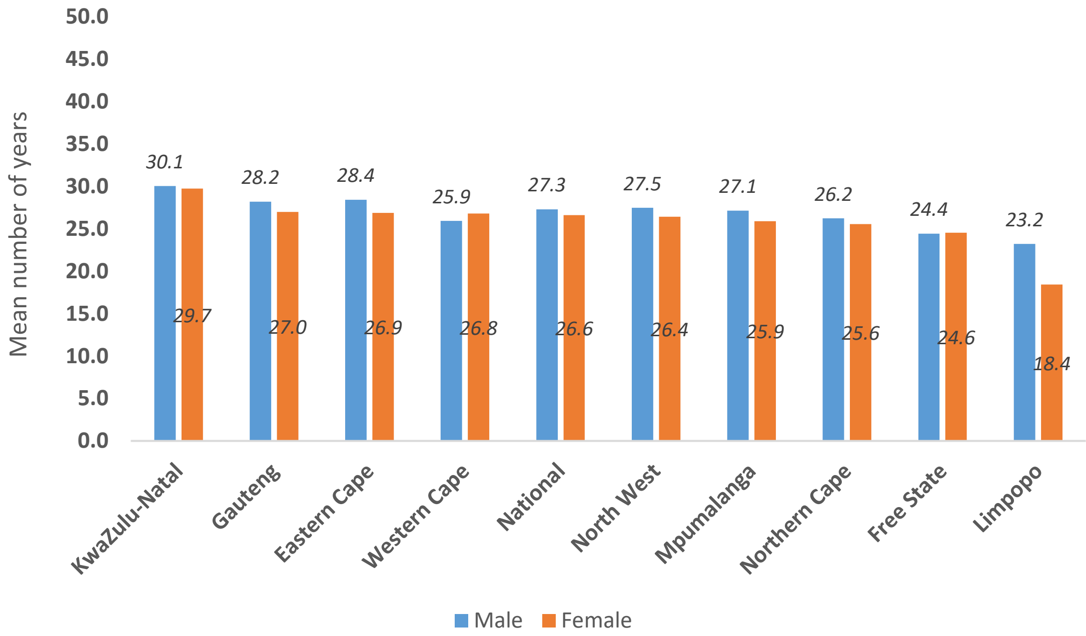
**FIGURE 8: UNADJUSTED SMAM (YRS), 2016 CS**



Source: Author's computation from the 2016 CS data

As seen from the graph, the adjusted SMAMs are lower than the corresponding values of the unadjusted SMAMs for both sexes. This implies that age errors exaggerated the unadjusted mean ages at marriage. At national level, for example, while female unadjusted SMAM is 31.1 years, the adjusted value is 26.6 years. Note that the adjusted value of SMAM for females in Limpopo (18.4 years) may to be too low.

**FIGURE 9: ADJUSTED SMAM (YRS), 2016 CS**




Source: Author's computation from 2016 CS data.

**Fertility**

It can be seen from the results of the application of the P/F ratio method to the 2016 CS depicted in Table 3 that, aside the age groups 40-44 and 45-49, the P/F values are roughly 1.0. This implies consistency and reasonable accuracy in the report on children ever born and births in the last 12 months by women aged 15-39 during the 2016 CS. The high values of P/F corresponding to women in the age groups 40-44 and 45-49 appear to suggest that women in these age groups exaggerated the number of children ever born. One possible reason for this is the "adoption effect" i.e. some women in these age groups reporting non-biological children as their children.

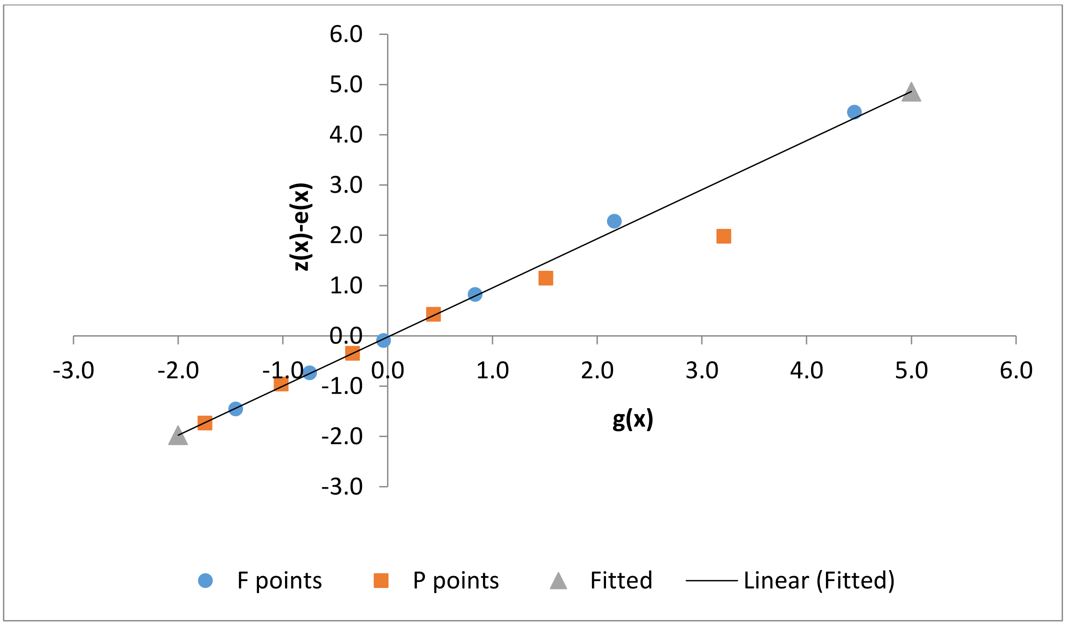
**TABLE 3: P/F RATIOS FROM 2016 CS USING HAMAD'S MULTIPLIERS, NATIONAL**

| Age group | Reported Age specific Fertility rates | Cumulated Age specific Fertility rates | Multipliers | Current Birth F | Mean Parity P | P/F |
|---|---|---|---|---|---|---|
| 15-19 | 0.050 | 0.000 | 2.568 | 0.129 | 0.131 | 1.018 |
| 20-24 | 0.104 | 0.251 | 3.044 | 0.568 | 0.623 | 1.097 |
| 25-29 | 0.101 | 0.772 | 3.077 | 1.082 | 1.161 | 1.073 |
| 30-34 | 0.089 | 1.276 | 3.160 | 1.558 | 1.661 | 1.066 |
| 35-39 | 0.062 | 1.722 | 3.314 | 1.928 | 2.044 | 1.060 |
| 40-44 | 0.027 | 2.033 | 3.435 | 2.127 | 2.389 | 1.123 |
| 45-49 | 0.005 | 2.170 | 4.227 | 2.190 | 2.652 | 1.211 |

$f_1/f_2 = 0.481$; $m = 28.261$
Source: Author's computation from 2016 CS data.

Further evaluation of the fertility reports was done by fitting the relational Gompertz model to the data. The results at national level are summarised in Figure 10 and Table 4.

**FIGURE 10: FITTING THE RELATIONAL GOMPERTZ MODEL TO CHILDREN EVER BORN (P) AND BIRTHS IN THE LAST 12 MONTHS (F), 2016 CS, NATIONAL**




Source: Author's computation from 2016 CS data.

**TABLE 4: OBSERVED AND ADJUSTED TFR FROM BIRTHS IN THE LAST 12 MONTHS (F POINTS) AND CHILDREN EVER BORN (P POINTS), 2016 CS, NATIONAL**

| Age group of women | TFR Based on F points |
|---|---|
| 15-19 | 2.45 |
| 20-24 | 2.49 |
| 25-29 | 2.39 |
| 30-34 | 2.39 |
| 35-39 | 2.38 |
| 40-44 | 2.48 |
| 45-49 | 2.66 |
| Observed TFR | 2.19 |
| α | -0.022 |
| β | 0.976 |
| Adjusted TFR | 2.4 |

Source: Author's computations from the 2016 CS data.

The slopping downward of the P points in Figure 10 appears to confirm exaggeration of children ever born by the older women noted earlier from the results of the P/F ratio method. The TFRs derived from each reproductive age group show remarkable consistency (Table 4) indicating that the quality of the fertility reports in the 2016 CS was generally good. Averaging the TFRs derived from the 15-19 … 35-39 age groups produced adjusted TFRs of 2.4 compared with an observed TFR of 2.2.

The relational Gompertz model was also fitted to the provincial fertility reports. After correcting for errors, the results indicate that Limpopo Province had the highest level of fertility in 2016 with adjusted total fertility rate of 3.0 followed by the North-West. In contrast, Gauteng, KwaZulu-Natal and the Western Cape had the lowest with adjusted total fertility rate of 2.2 in each of the three provinces (Table 5).

**TABLE 5: OBSERVED AND ADJUSTED TOTAL FERTILITY RATES BY PROVINCE, 2016 CS**

| Province | Observed TFR | α | β | Adjusted TFR |
|---|---|---|---|---|
| Eastern Cape | 2.4 | 0.006 | 0.963 | 2.5 |
| Free State | 2.2 | -0.042 | 0.990 | 2.5 |
| Gauteng | 2.0 | -0.081 | 1.009 | 2.2 |
| KwaZulu-Natal | 2.0 | 0.082 | 0.971 | 2.2 |
| Limpopo | 2.7 | -0.119 | 0.953 | 3.0 |
| Mpumalanga | 2.1 | 0.005 | 0.965 | 2.5 |
| North West | 2.8 | 0.072 | 1.001 | 2.9 |
| Northern Cape | 2.7 | 0.032 | 1.024 | 2.7 |
| Western Cape | 2.0 | -0.052 | 1.025 | 2.2 |

Source: Author's computations from the 2016 CS data.

**Mortality**

***Childhood mortality***

The results of the indirect estimation of childhood mortality from the reports on children dead of children ever born are summarised in Table 6. Since column 4 of the table are cumulative probabilities of dying from birth, the probabilities should increase with the age of children (and age of mother) in the absence of errors in the data. As seen from column 4, the increase is very gradual and appears to suggest incomplete reporting of children dead by some women. Better insight was sought by constructing a one-parameter logit life table (i.e. a life table based on mortality of children alone) using the level of childhood mortality in the logit system shown in column 7 of Table 6 for a reference date of 2012 (column 6) i.e. 3.1 years before the 2016 CS (column 5).

The childhood mortality indicators derived from the one-parameter logit life table (Table 7) seemingly suggest that of every 1000 children born in 2014, 40 were not likely to survive to their fifth birthday whereas applying a similar approach by this author to the 2016 South Africa Demographic and Health Survey (SADHS) data suggest that 75 children were not likely to survive to their fifth birthday of every 1000 children born in 2014.

**TABLE 6: INDIRECT ESTIMATION OF CHILDHOOD MORTALITY, 2016 CS, BOTH SEXES, NATIONAL**

| Age of Women | Age of Children (Age x) | Proportion dead of children ever born | Probability of dying from birth to age x ($q_x$) | No of years before survey | Reference date | α |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 15-19 | 1 | 0.025 | 0.024 | 1.4 | 2014.6 | -0.624 |
| 20-24 | 2 | 0.029 | 0.029 | 3.1 | 2012.9 | -0.594 |
| 23-29 | 3 | 0.034 | 0.033 | 5.0 | 2011.0 | -0.590 |
| 30-34 | 5 | 0.039 | 0.039 | 7.3 | 2008.7 | -0.628 |
| 35-39 | 10 | 0.044 | 0.045 | 9.7 | 2006.3 | -0.612 |
| 40-44 | 15 | 0.050 | 0.050 | 12.6 | 2003.4 | -0.591 |
| 45-49 | 20 | 0.064 | 0.063 | 16.0 | 2000.0 | -0.507 |

P2/p3 = 0.536, m (from births last 12 months) = 28.3
Source: Author's estimates from 2016 CS data

The estimates of childhood mortality derived from the reports on household deaths (next section) also indicate higher childhood mortality than suggested by the indirect estimates derived from children dead of children ever born. In view of this seemingly large-scale omission of children dead of the reported number of children ever born, the estimates were not disaggregated further by province.

**TABLE 7: CHILDHOOD MORTALITY INDICATORS INDIRECTLY ESTIMATED FOR YEAR 2014 (BOTH SEXES), 2016 CS**

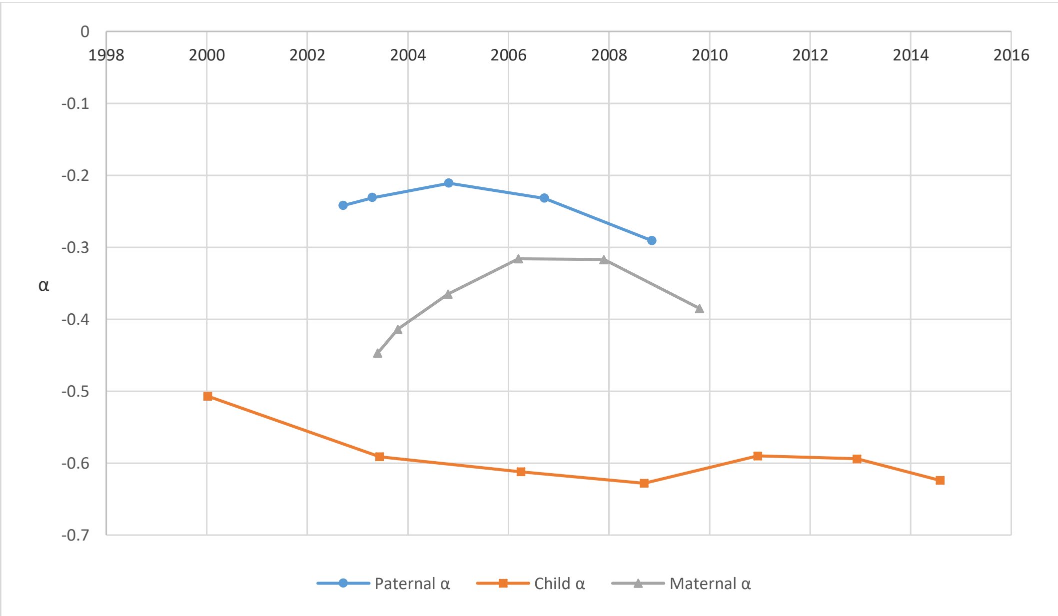
| Measure | Childhood Mortality Indicator (per thousand) |
|---|---|
| α | -0.624 |
| β | 1.000 |
| Infant mortality rate (1q0) | 24 |
| Child mortality rate (4q1) | 17 |
| Under five mortality rate (5q0) | 40 |

Source: Author's estimates from 2016 CS data.

***Adult Mortality***

The results of the application of the orphanhood method to responses to the orphanhood questions in the 2016 CS are summarised in Figure 11. It appears from the graph that that adult male mortality increased during the period 2002 and 2005 and declined during the period 2005 and 2009 while adult female mortality increased during the period 2003 and 2006, declined during the period 2007 and 2009. Despite the decline in the more recent periods, the low negative α values for both adult male and female mortality appear to suggest high adult mortality in the population (as α slides from negative towards positive on the y-axis in the graph, mortality increases).

**FIGURE 11: TRENDS IN MORTALITY, 2016 COMMUNITY SURVEY, NATIONAL**




Source: Author's computations from the 2016 Community Survey

***Combining Adult Mortality Estimates from Orphanhood with Indirect Estimates of Childhood Mortality***

The adult mortality estimates from the orphanhood reports were combined with the indirect estimates of childhood mortality in the previous section and as described in the methods section to produce a two-parameter logit life table for a reference date of 2014. The results suggest infant mortality rate of about 23 per thousand live births (both sexes), life expectancy at birth of about 62 years and 63 years for males and females respectively during the period 2014 nationally (Table 8). The overall childhood mortality rate as represented by the total mortality rate in the first five years of life, $_5q_0$ is lower than that estimated from the reports on household deaths (see next section). Therefore, as noted earlier, the indirect approach appears to underestimate childhood mortality.

**TABLE 8: ESTIMATED MORTALITY INDICATORS FROM CHILDREN DEAD OF CHILDREN EVER BORN AND ORPHANHOOD REPORTS 2016 CS, NATIONAL**

| Measure | Mortality Indicator 2014 |
|---|---|
| α (both sexes) | -0.400 |
| β Males | 1.258 |
| β Females | 1.208 |
| Infant mortality rate ($_1q_0$) per thousand (both sexes) | 23 |
| Child mortality rate ($_4q_1$) per thousand (both sexes) | 28 |
| Under five mortality rate ($_5q_0$) per thousand (both sexes) | 49 |
| Male life expectancy at birth | 61.8 |
| Female life expectancy at birth | 63.1 |

Source: Author’s estimates from 2016 CS data.

***Mortality from Household Deaths***

The life expectancies at birth computed directly from the age specific deaths rates derived from the reports on household deaths using life table method are shown in Table 9. The values are improbably high. This prompted the fitting of the Growth Balance method to the data to assess the completeness of reporting of deaths in households during the survey. For brevity of presentation only the outcome at national level of this process is shown and for males only (Figure 12).

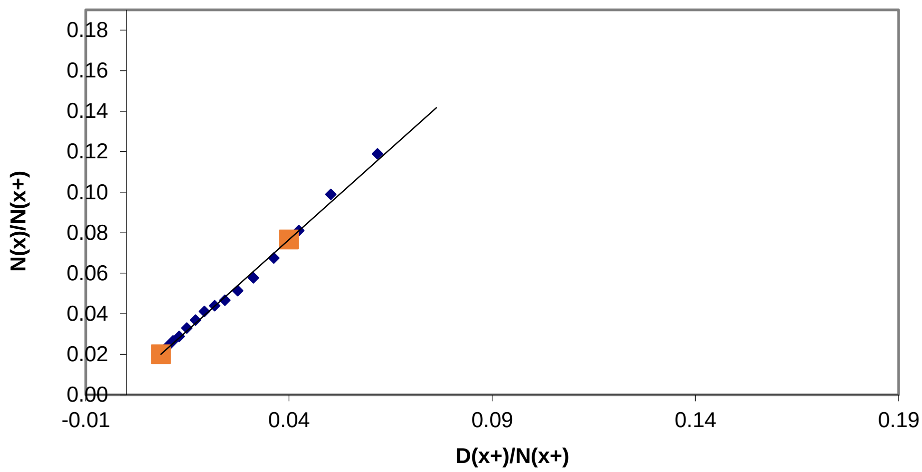
**TABLE 9: UNADJUSTED LIFE EXPECTANCIES AT BIRTH FROM REPORTS ON DEATHS IN HOUSEHOLDS, 2016 CS**

| | Life expectancy at birth (years) | |
|---|---|---|
| **Province** | **Male** | **Female** |
| Eastern Cape | 57.7 | 65.0 |
| Free State | 57.9 | 81.2 |
| Gauteng | 77.5 | 79.0 |
| KwaZulu-Natal | 58.9 | 72.3 |
| Limpopo | 72.4 | 75.0 |
| Mpumalanga | 62.3 | 70.9 |
| North West | 67.3 | 61.7 |
| Northern Cape | 58.3 | 64.2 |
| Western Cape | 83.4 | 87.4 |
| **Total*** | **70.0** | **73.4** |

Source: Author's computations from the 2016 Community Survey

*Estimates are not weighted averages of the provincial estimates but estimated from the national data.

**FIGURE 12: FITTING THE GROWTH BALANCE METHOD TO REPORTS ON DEATHS IN HOUSEHOLDS: MALES 2016 CS, NATIONAL**




Source: Author's computations from the 2016 CS data.

The following were evident and from the summary shown in Table 10. The quality of reporting of deaths in households during the 2016 Community Survey was poor at national and provincial levels. At national level completeness of reporting of deaths that occurred during the 12 months period preceding the survey was 55.7% for males and 49.3% for females suggesting that only about one half of the deaths that occurred in the reference period were reported.

**TABLE 10: ESTIMATED PERCENT COMPLETENESS OF REPORTING OF DEATHS IN HOUSEHOLDS BY SEX, 2016 COMMUNITY SURVEY**

| Province | Male | Female |
|---|---|---|
| Eastern Cape | 60.0 | 47.9 |
| Free State | 52.8 | 39.8 |
| Gauteng | 50.8 | 44.0 |
| KwaZulu-Natal | 41.7 | 41.9 |
| Limpopo | 59.7 | 53.7 |
| Mpumalanga | 61.4 | 44.8 |
| North West | 61.0 | 59.8 |
| Northern Cape | 66.9 | 57.4 |
| Western Cape | 40.4 | 32.3 |
| **Total*** | **55.7** | **49.3** |

Source: Author's computations from the 2016 CS data.
*Estimates are not weighted averages of the provincial estimates but Estimated from the national data.

Estimated completeness of reporting of deaths in households at provincial level ranged between 40.4% (Western Cape) and 66.9% (Northern Cape) for males and for females, it ranged between 39.8% (Free State) and 59.8% (North West). It was difficult to fit a straight line in the application of the Growth Balance method for some of the provinces owing to the poor quality of the data especially in the Eastern Cape, Free State, KwaZulu-Natal. The adjusted mortality indicators from the reports on deaths in households are summarised in Table 11. The estimates suggest that life expectancy at birth in 2015/2016 nationally was about 59.4 years for males and 64.5 years for females.

**TABLE 11: ADJUSTED MORTALITY INDICATORS FROM REPORTS ON DEATHS IN HOUSEHOLDS, 2016 CS**

| | | | Life expectancy at birth (years) | |
|---|---|---|---|---|
| **Province** | **Infant mortality ($_1q_0$ both sexes) Per thousand 2015/2016** | **Under five mortality ($_5q_0$ both sexes) Per thousand 2015/2016** | **Male 2015/2016** | **Female 2015/2016** |
| Eastern Cape | 45 | 65.4 | 49.7 | 55.9 |
| Free State | 51 | 64.9 | 51.2 | 66.0 |
| Gauteng | 35 | 49.6 | 67.1 | 69.4 |
| KwaZulu-Natal | 52 | 74.8 | 50.7 | 61.9 |
| Limpopo | 43 | 56.4 | 62.5 | 69.0 |
| Mpumalanga | 45 | 61.2 | 55.6 | 61.7 |
| North West | 63 | 79.4 | 56.7 | 56.4 |
| Northern Cape | 45 | 60.7 | 53.7 | 58.5 |
| Western Cape | 25 | 33.9 | 68.9 | 74.6 |
| **Total*** | **44** | **60.3** | **59.4** | **64.5** |

Source: Author's computations from the 2016 Community Survey
*Estimates are not weighted averages of the provinces but estimated from the national data.

*Maternal Mortality*

Estimates of maternal mortality from the 2016 CS based on the methods described above are summarised in Table 12. At national level both the unadjusted (298/100,000) and adjusted (274/100,000) maternal mortality ratios are almost three times lower than the estimates obtained by this author from the 2016 SADHS using the same methods described above. The 2016 SADHS was carried out about the same time as the 2016 CS and interestingly, also implemented by Statistics South Africa as a client of the National Department of Health for the survey. The level of maternal mortality ratio obtained from the 2016 CS is also about three times lower than the levels estimated by other researchers for South Africa (see for example, Garenne et al. (2009), Moszynski (2011), Udjo and Lalthapersad-Pillay, (2014). In view of these, the data on maternal deaths from the 2016 Community Survey should be treated as highly suspect and for this reason, it was not worthwhile disaggregating the estimates by province.

**TABLE 12: UNADJUSTED AND ADJUSTED MATERNAL MORTALITY RATIOS 2015/2016: DIRECT ESTIMATES**

| Variable/Measure | 2016 Community Survey | | 2016 SADHS | |
|---|---|---|---|---|
| | Unadjusted | Adjusted | Unadjusted | Adjusted |
| Years before Survey | 0-1 | 0-1 | 0-1 | 0-1 |
| No. of maternal deaths | 182 | | 5 | |
| No. of women aged 15-49 | 903,052 | 903,052 | 8,514 | 8,514 |
| No. of births last 12 months | 61,176 | 100,165 | 645 | 716 |
| General Fertility Rate (per 1000) | 67.7 | 110.9 | 75.8 | 84.1 |
| **Maternal mortality ratio (per 100,000)** | **298** | **274** | **775** | **698** |

Source: Author’s estimates from 2016 CS and 2016 SADHS

*Migration*

Migration has two components – internal and international migration. On examining the 2016 CS data file, this author observed that the data corresponding to the question: In which province does this person usually live? – were given two different variable names, namely, POUR and Usual_Province. However, the frequency distributions obtained from these two variable were different. Similarly, the data corresponding to the question: In which province did this person live before moving to this province? – were given two different variable names, namely, PopRes and Previous_Province. The frequency distributions from the two variables were also different. The values from POUR and Usual_Province should be the same. To illustrate the discrepancy, it can be seen from Table 13 that in all the provinces, the numbers in the second column are lower than the numbers in the third column.

Whereas in the second column of table 13, the 'Not applicable' is 3,303,559, the corresponding number in column 3 of table 13 is 0 i.e. in both columns, the total is 3,328,867.

**TABLE 13: NUMBER OF PERSONS REPORTED AS USUAL RESIDENTS OF PROVINCE, 2016 CS**

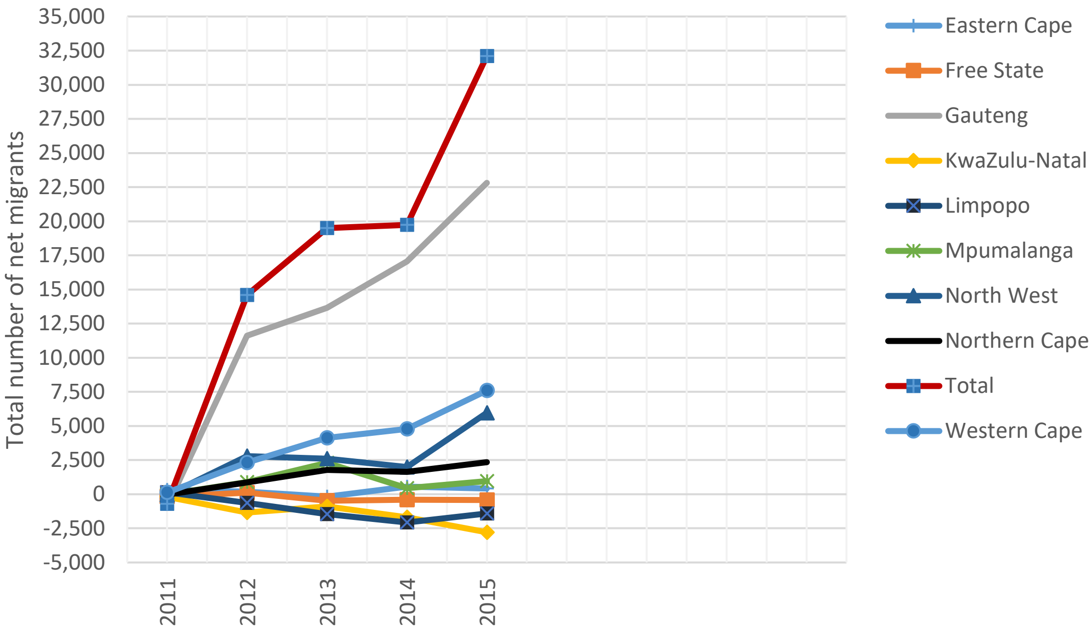
| Province | In which province does this person usually live? | |
|---|---|---|
| | Variable POUR | Variable Usual_Province |
| Eastern Cape | 3,580 | 463,332 |
| Free State | 968 | 195,319 |
| Gauteng | 5,438 | 725,005 |
| KwaZulu-Natal | 2,911 | 659,518 |
| Limpopo | 3,656 | 402,890 |
| Mpumalanga | 1,797 | 270,149 |
| North West | 1,623 | 247,598 |
| Northern Cape | 516 | 82,327 |
| Western Cape | 1,548 | 279,455 |
| Outside South Africa | 1,227 | 1,227 |
| Do not know | 215 | 215 |
| Not applicable | 3,303,559 | 0 |
| Unspecified | 1,832 | 1,832 |
| **Total** | **3,328,867** | **3,328,867** |

Source: Tabulated from 2016 CS data

In email communication by this author, these anomalies were brought to the attention of Statistics South Africa but did not receive satisfactory explanation of these anomalies from the organisation. Given that the reported numbers of usual province of residence and previous province of residence by province are implausibly low probably perhaps owing to inappropriate implementation of the questions, the derived variables in the data file were used in the estimates provided below. The total net migration estimates provided below incorporated immigration into province as well as emigration from the province. The results should however be treated with skeptism in light of the anomalies noted above.

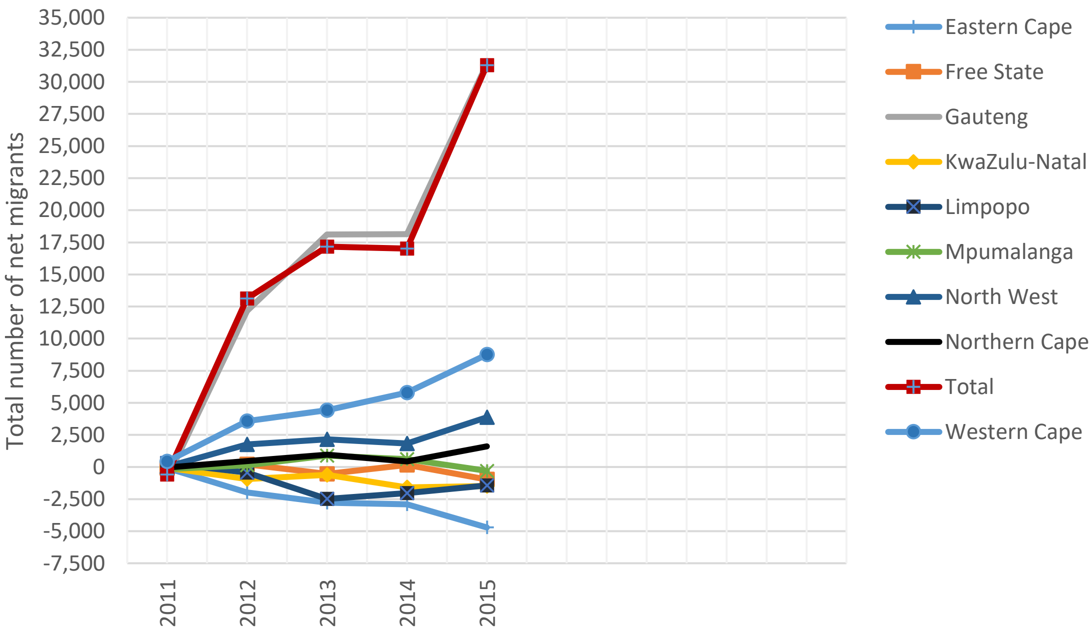
Figures 13-14 suggest that total that total net migration was positive during the period 2012-2015 among males and females in Gauteng, Western Cape, North-West and Northern Cape while it was negative among males and females in the other provinces during the period. Of the provinces that suffered increasing loss in total net migration, KwaZulu-Natal and Limpopo had the steepest trend among males while among females, Limpopo and the Eastern Cape had the steepest trend.

**FIGURE 11: TRENDS IN PROVINCIAL TOTAL NET MIGRANTS, MALES**




Source: Author’s estimates from 2016 CS data.

**FIGURE 13: TRENDS IN PROVINCIAL TOTAL NET MIGRANTS, FEMALES**




Source: Author’s estimates from 2016 CS data.

**DISCUSSION AND CONCLUSION**

The rationale and justification for the 2016 CS are embedded in the survey objectives. The survey objectives raise the question: Was the 2016 CS really necessary? Regarding the first two objectives – provide an estimate of the population count as well as household count by local municipality - a sample survey is not an appropriate medium for providing population estimates at local municipality level. Rao (2008), Tanton et al. (2011) have noted that survey approach in providing population estimates entails various methods of re-weighting survey data to a number of known totals for small area. However, because of operational cost, it is usually not possible to have an overall sample that is large enough to provide direct estimates for all the small areas (Rao 2008). Consequently, because the standard errors for the small area estimates are generally too high, they render the estimates unreliable (Datta & Ghosh, 2012).

The other three objectives of the 2016 CS were: measurement of fertility, mortality and migration, measurement of socio-economic factors; measurement of access to facilities and services. Fertility, mortality and migration measurements are necessary for updating demographic indicators. However, some other annual surveys by Statistics South Africa such as the General Household Survey and Quarterly Labour Surveys are already doing this to some extent. Since demographic phenomena change very slowly what would be more cost effective is to include additional fertility, mortality and migration questions every three or five years in the General Household Surveys. From the standpoint of the specific objectives of the 2016 CS therefore, the CS (programme that began in 2007) in the view of this author was unnecessary.

Regarding the quality of the data, there is no gold standard in judging the accuracy of survey (or Census) however, the following to mention a few were evident from the results. Based on the evaluation provided in this study, the fertility aspect of the 2016 CS is good and the most reliable of all the demographic aspects of the 2016 CS data. The indirect estimation of childhood mortality suggested that there was large-scale omission of dead children in the women's reports on child survival during the survey. Demographic indicators of childhood mortality relying on the 2016 CS would therefore underestimate the level of childhood mortality in the South African population. The quality of the reporting of deaths in households in the 2016 CS was poor. The levels of maternal mortality ratio obtained from the 2016 CS even after adjustment are highly suspect.

Statistics South Africa appears to be obsessed with calibrating survey data to mid-year estimates to produce population numbers of socio-economic and demographic phenomena. Such practice is dangerous from a policy viewpoint because the numbers produced from such an exercise can potentially mislead policy makers and other users. Take for example the numbers of unemployed people provided in Statistics South Africa's Quarterly Labour Survey reports (or the numbers from the 2016 CS using the weighted data). In the case of the Eastern Cape, the seemingly exaggeration of the population in 2016 mid-year estimate for the province - 7,061,717 in 2016 compared with 6,498,683 in 2017 (Statistics South Africa 2016c, 2017) - would have overstated the number of unemployed persons in that province in 2016 since Statistics South Africa calibrates its survey data to the mid-year estimates. Therefore, policy makers and other users should be wary of survey figures derived from weighted and calibrated data to Statistics South Africa's mid-year estimates. The purpose of a survey is to derive and update demographic and socio-economic indicators (i.e. rates and ratios) and not to produce population estimates.

**ACKNOWLEDGEMENTS**

This author wishes to thank Statistics South Africa for access to their data. The views expressed in this study are, however those of the author.

**REFERENCES**

Booth, H. 1979. The estimation of fertility from incomplete cohort data by means of the transformed Gompertz model. PhD thesis, London School of Hygiene Medicine, University of London.

Brass, W. 1981. The use of the relational Gompertz model to estimate fertility. *Proceedings of the International Union for the Scientific Study of Population Conference*, Manilla, Liege, vol. 3: 345-361.

Brass W., & Bamgboye, E.A. (1981). *The time location of reports of survivorship: estimates for maternal and paternal orphanhood and the ever-widowed*. Centre for Population Studies, London School of Hygiene and Tropical Medicine, Working Paper no. 81-1.

Carrier, N., & Hobcraft, J. 1971. *Demographic estimation for developing societies*. London: Population Investigation Committee.

Datta, G., Ghosh, M. 2012. Small area shrinkage estimation. *Statistical Science* 27 (1): 95-115.

Garenne, M., McCaa, R., & Nacro, K. 2009. Maternal mortality in South Africa: An update from the 2007 community survey. *Paper prepared for the 26th IUSSP International Conference*, Marrakech, 27 September – 2 October.

Hill, K. 1987. Estimating census and death registration completeness. *Asian and Pacific Population Forum*, 1 (3):8-13.

Kenya Bureau of Statistics. 2015. *Kenya Demographic and Health Survey 2014*. Nairobi: Kenya Bureau of Statistics.

INDEPTH Network 2004. *INDEPTH model life tables for sub-Saharan Africa*. Aldershot: Ashgate Publishing Limited.

Moszynski, P. 2011. South Africa's rising maternal mortality is due to health system failures. *British Medical Journal* 343, d5089.

Rao, J.N.K. 2008. Some methods of small area estimation. *Rivista Internazionale di Sociali*, 116 (4): 387-406.

Statistics Sweden. 2015. Marriages and births in Sweden. Eurostat. Retrieved from: www.ec.europa.eu/eurostat/statistic-explained/index.php/Marriages_and_births_in_Sweden. (Date accessed 14 August 2017).

Stanton, C., Abderrahim, N., & Hill, K. 1997. *DHS maternal mortality indicators: An assessment of data quality and implications for data use.* DHS Analytical Reports No.4. Calverton, Maryland: Macro International Inc., pp.1-2.

Statistics South Africa 1996. *The people of South Africa population Census, 1996: The count and how it was done*. Pretoria: Statistics South Africa.

Statistics South Africa 1998. *Population Census, 1996: Census '96 overview*. Pretoria: Statistics South Africa.

Statistics South Africa. 2016a. *Community Survey 2016 Statistical Release*. Pretoria: Statistics South Africa.

Statistics South Africa. 2016b. *Community Survey 2016 Technical Report*. Pretoria: Statistics South Africa.

Statistics South Africa. 2016c. Mid-year population estimates 201. Pretoria: Statistics South Africa.

Statistics South Africa. 2017. Mid-year population estimates 201. Pretoria: Statistics South Africa.

Tanton, R., Vidyattama, Y., Nepal, B., & McMara, J. 2011. Small area estimation using a re-weighting algorithm. *Journal of the Royal Statistical Society Series A (Statistics in Society)* 174 (4): 931-951.

Udjo, E.O. 2008. A re-look at recent statistics on mortality in the context of HIV/AIDS with particular reference to South Africa. *Current HIV Research*, 6: 143-151.

Udjo, E.O. 2014. Estimating demographic parameters from the 2011 South Africa population census. *African Population Studies*, 28, (1): 564-578.

Udjo, E.O. 2015. *Population estimates for South Africa by district municipality and province, 2015*. Bureau of Market Research, Research Report No 461, University of South Africa, Pretoria.

Udjo, E.O. 2017. Evaluating the demographic aspects and estimating demographic indicators from the 2016 South Africa Community Survey. Pretoria: Bureau of Market Research Report No. 486.

Udjo, E.O. & Lalthapersad-Pillay, P. 2014. Estimating maternal mortality and causes in South Africa: national and provincial levels. *Midwifery*, 30 (5): 512-518.

UNICE undated. Switzerland, mean age at first marriage. Retrieved from: www.w3.unece.org/PXWeb/en/TimesSeries?IndicatorCode=303&CountryCode=756. Accessed 14 August 2017

UNICEF. 2016. *Child marriage in Ethiopia: a review of the evidence and an analysis of the prevalence of child marriages in hotspot districts*. Ethiopia: Overseas Development Institute and UNICEF, Ethiopia.

Van De Walle, E. 1968. Marriage in African censuses and inquiries. In, *The Demography of Tropical Africa*. Princeton: Princeton Press, p 203.

World Health Organisation. 2017. Health statistics and information systems. Retrieved from: www.who.int/healthinfo/statistics/indmaternalmortality/en/ (Date accessed15 June 2017).

Zaba, B. 1981. *Use of the relational Gompertz model in analysing fertility data collected in retrospective surveys*. Centre for Population Studies Working Paper No 8, London School of Hygiene & Tropical Medicine, University of London.