

Estimating demographic parameters from the 2011 South Africa population census

Eric O Udjo

Bureau of Market Research, University of South Africa, PO Box 392 UNISA 0003
Pretoria, Email: udjoe@unisa.ac.za, Telephone: +27 12 429 3326, Fax: +27 12 429 6225

Abstract

Due to limited coverage of vital registration in most African countries, censuses and surveys constitute the primary sources for demographic information. Estimating demographic parameters from African censuses often poses a challenge. South Africa has a long history and high coverage of vital registration but censuses and surveys remain the principal sources of demographic information. The purpose of this study is to estimate demographic parameters from the 2011 South Africa census using various demographic models. The results indicate female mean age at first marriage of 26 years, infant mortality rate of 60 per thousand live births (both sexes), life expectancy at birth of 55 years (both sexes), total fertility rate of 2.7 in 2011 and Gauteng as the province with the highest volume of net migration in absolute terms. Despite some weaknesses in the 2011 census data, it is a rich data source for demographic estimation.

Keywords: Age distribution; marriage; fertility; mortality; net migration

Résumé

En raison de la couverture limitée des registres d'état civil dans la plupart des pays africains, les recensements et enquêtes constituent les principales sources d'informations démographiques. Estimation des paramètres démographiques des recensements africains souvent pose un défi. Afrique du Sud a une longue histoire et une couverture élevée de registres d'état civil mais les recensements et les enquêtes restent les principales sources d'informations démographiques. Le but de cette étude est d'estimer les paramètres démographiques du recensement 2011 Afrique du Sud à l'aide de divers modèles démographiques. Les résultats indiquent des femme âge moyen au premier mariage de 26 ans, les taux de mortalité infantile de 60 pour mille naissances vivantes (les deux sexes), l'espérance de vie à la naissance de 55 ans (des deux sexes), taux de fécondité de 2,7 en 2011 et de Gauteng, la province avec le plus grand volume de la migration nette en termes absolus. Malgré quelques faiblesses dans les données du recensement de 2011, c'est une source de données pour l'estimation démographique.

Mots-clés : Distribution d'âge ; mariage ; fertilité ; mortalité ; migration nette

Introduction

Problem Statement

Demographic information constitutes the bed-rock of all socioeconomic planning. Coverage of vital registration in most African countries is limited hence censuses and surveys are the primary sources of demographic information. For planning, demographic estimates need to be reasonably accurate to inform efficient allocation of scarce resources. Estimating demographic parameters from censuses (and surveys) however, often poses a challenge as the reported information may not be accepted at face value. Coverage and content errors are often present in the data. The magnitude of the errors differs from

one country to another depending on the level of statistical development in the country. Coverage errors in censuses are easier to deal with, through the use of a post enumeration survey (PES). Estimating demographic parameters from the data is more problematic because content errors are more difficult to detect and correct for. It is often the case that demographic parameters are extracted from censuses without adequate attention to errors that may be present in the data.

South Africa has a high coverage of vital registration. Groenwald et al (2005) estimated completeness of adult registration to be over 90%. Udjo (2006) estimated that registration of female

deaths increased from about 74% in 1997 to 84% in 2002. For males, it increased from 84% in 1997 to 92% in 2002. Despite the high coverage, censuses and surveys remain the principal sources of demographic information in South Africa. The focus of this study is on the census as a source for demographic estimation. The 1970 South Africa Census was the last apartheid census that covered the entire country. The apartheid censuses of 1980, 1985 and 1991 excluded the former homelands – Transkei, Bophuthatswana, Venda and Ciskei (TBVC). Since the demise of apartheid in 1994, three censuses have been undertaken, namely in October 1996, 2001 and 2011. The overall undercount in the 1970 South Africa Census was 2%. It increased to 11% in the 1996 census and 18% in the 2001 census but decreased to 14.6% in the 2011 census. There was an explosion in the number of questions in the post-apartheid census questionnaire (in comparison with the 1970 census questionnaire). Currently, South Africa probably has the longest census questionnaire in the world. Several aspects of each of South Africa's post-apartheid censuses are controversial. See for example, Dorrington (1999), Sadie (1999), Shell (1999), Phillips et al. (1999), (Udjo 1999, 2004a, 2004b). It is clear from these controversies that the demographic information from censuses and surveys in South Africa (as elsewhere) need careful evaluation.

Objectives of Study

The purpose of this study therefore is to estimate demographic parameters from the 2011 South Africa census data taking into account errors that may be present in the data. Such estimates are valuable inputs in developing assumptions for demographic projections. The demographic parameters cover broadly, age distribution, marriage, fertility, mortality and net migration. Where the quality of the data permit, the demographic parameters were estimated at national level as well as at provincial levels and by population group. The population group classification used in South Africa consist of Blacks, Coloureds (persons of mixed descent), Indians/Asians and Whites. According to the 2011 census results, Blacks constituted 79.2%, Coloureds 8.9%, Indians/Asians 2.5% and Whites 8.9% of South Africa's population of 51,770,050 persons in October 2011 (Statistics South Africa 2012).

Data

The primary data utilised for this study is the full 2011 Census data (Statistics South Africa 2012). There is a wealth of demographic information in the 2011 census data which includes broadly, age-sex, marital status, migration, parental survival, fertility, household mortality and pregnancy related deaths. Reference is made to other data sources notably, the 2007 Community Survey (Statistics South Africa, 2007), the 1996 and 2001 censuses (Statistics South Africa 1997, 2003) in evaluating the consistency of the estimates presented.

Methods

The principal methods utilised for the demographic estimation are outlined in what follows.

Estimating corrected age distribution

Reported age-sex distributions in censuses (and surveys) usually contain errors in the form of heaping (preference for certain ages typically ending in zeroes and fives) and shifting (transfer of one's age group to another which may not necessarily be the adjacent age group). Often, researchers compute indices such as Whipple's and Myers' indices or United Nations joint scores as a means of assessing the accuracy of reported age distributions. The summary measures produced by these indices do not reveal very much. Deeper insights may be obtained by producing population pyramids or plotting the age-sex distributions. This approach was utilised and the smoothed age-sex distributions were obtained by comparing the logit transformations of the reported cumulated age-sex distribution with an appropriate stable population (see Carrier and Hobcraft, 1971). The logit transformations given by Brass (1971) are denoted as:

$$Y_x = \frac{1}{2} \log_e \left(\frac{(1-P_x)}{P_x} \right) \text{ and}$$

$$Y_{sx} = \frac{1}{2} \log_e \left(\frac{(1-P_{sx})}{P_{sx}} \right)$$

Where Y_x and Y_{sx} are the logit transformations of the reported and stable age distributions respectively, P_x and P_{sx} are the cumulate proportions under age x in the observed and stable populations respectively. Plotting Y_x against Y_{sx} and fitting a straight line to the "best" points (in this case, the least squares method) enabled the estimation of the smoothed age distribution computed as:

$$Y'_x = \alpha + \beta Y_{sx}$$

Where Y'_x is the transformed smoothed age distribution, α and β are the intercept and slope of the fitted line respectively. Re-transformation of the fitted values were obtained using the expression:

$$S_x = 1/(1 + e^{2Y'_x})$$

Where S_x is the anti-logit of Y'_x .

Estimating the mean age at first marriage

Enumerated persons in the 2011 census were asked their current marital status. The singulate mean age at first marriage (SMAM) – an estimate of the mean number of years lived by a cohort before their first marriage (Hajnal 1953) – is an indirect estimate of the mean age at first marriage and was estimated from the responses to the current marital status question. Assuming all first marriages took place by age 49, the singulate mean age at first marriage (SMAM) is expressed as:

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

where P_x is the proportion single at age x . To minimise the effect of age mis-reporting on the SMAM, an overall dichotomy of the population into single and non-single was made and translated into an age at first marriage by linear interpolation on the smoothed age distribution (i.e. the age at which the population cumulated since birth, equals the proportion single (see Van De Walle, 1968).

Estimating fertility

Questions on fertility in the 2011 census included children ever born and date of last child born. It is well established (see Brass 1971) that there are often errors in the responses to these questions. The sources of error for various reasons, are reference period (births that were not born in the reference period reported as born in the reference period) and omission or over reporting of children ever born. Depending on the direction of the error, reported fertility will under- or over- estimate the “true” level of fertility. Fertility rates were therefore estimated from the data as follows. Although Brass (1971) P/F ratio method was designed to detect and correct for errors in the estimation of fertility from children ever born and births in the previous 12 months, the method assumes fertility has been constant. It is therefore inappropriate to use this method to estimate fertility if fertility has been subject to trend as is the case in South Africa as one of the key assumptions of the method would be violated. To disentangle errors from trends in the estimation of fertility from the fertility reports, the relational Gompertz model was therefore used.

The model is expressed as:

$$F(x) = F \cdot e^{-e^{[\alpha + \beta Y_s(x)]}}$$

Where $F(x)$ is the cumulated age-specific fertility rate up to age x and F represents the total fertility rate, $Y_s(x)$ is defined as $-\ln [-\ln F_s(x)/F]$ where $F_s(x)$ is a standard cumulative fertility rate up to age x (Brass 1981). The α and β parameters measure the location and spread of the fertility distribution for the particular population (Brass 1981). Further details about the development and use of the method are given in Brass (1974), Booth (1979), Zaba (1981).

Estimating mortality

Mortality had several components in the 2011 census namely childhood mortality, orphanhood, deaths in households and pregnancy related deaths.

Childhood mortality

From the responses to the questions in the 2011 census on the number of children ever born and children surviving the probability of surviving from birth to exact age x was estimated using the equation provided by Brass (1971):

$$q_x = D_i k_i$$

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 that adjusts for non-mortality factors determining the value of D_i . AIDS was incorporated into the estimates using the INDEPTH (2004) standard life tables for sub-Saharan Africa. The estimated q_x were converted into α values (level of childhood) in the logit system from the following:

$$\alpha = Y_x - Y_{xs} \text{ where}$$

$$Y_x = 1/2 \log_e \{(1 - I_x)/I_x\} \text{ and}$$

$$Y_{xs} = 1/2 \log_e \{(1 - I_{xs})/I_{xs}\}.$$

I_x and I_{xs} are, respectively, observed and standard life table (INDEPTH) probabilities of survival from birth to exact age x . The reference dates of the α values were estimated using a model for estimating the time location of childhood mortality developed by Brass (1985). The sources of error from the report on children ever born and children surviving usually lie in underreporting of children dead -some mothers tend to be reluctant to report the number of children dead in censuses and surveys because of the sensitivity of the deaths. This might lead to under estimating of childhood mortality from such reports.

Orphanhood

The orphanhood questions in the 2011 census were: Is this (person’s) own biological mother alive?

Is this (person's) own biological father alive? From the responses to these questions, the probability of dying from a base age in an adult age, B to age B+N can be estimated from the proportions of persons with a surviving parent from the equation developed by Brass (1971):

$$I_{B+N}/I_B = W_N ({}_5P_{N-5}) + {}_5P_N (1 - W_N)$$

Where I_{B+N}/I_B is the probability of surviving from a base age B, to B+N.

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 which depends on N and the location of child bearing, represented by the mean age of childbearing, M . The survival probabilities were converted into α values (levels of adult mortality) in the logit system using the following expression:

$$\alpha = \frac{1}{2} \log_e (1 + ({}_N P_B - I/I_B) / (1 - {}_N P_B))$$

Where ${}_N P_B$ is the probability of surviving from base age B to age B+N; $I_{s(B+N)}$ is the probability of surviving to exact age B+N in a standard life table (Brass and Bamgboye 1981). AIDS was incorporated into the estimates using the INDEPTH (2004) standard life tables for sub-Saharan Africa. The reference date of the α values was estimated using a model for estimating the time location of adult mortality developed by Brass and Bamgboye (1981). The sources of errors in responses to orphanhood questions is underreporting of dead mothers and fathers especially of the younger respondents due to the "adoption effect" (see Brass 1971, Blacker 1977, Hill 1977, Timaeus 1991, Mandela 1994, Udjo 2005).

Mortality Evaluation from Distribution of Deaths in Households

The main issue about reports on deaths in households during a census or survey (or vital registration) is completeness of reporting. The main source of error in this information is that the number of deaths reported in households over a reference period tends to be underreported or over reported relative to the census or survey coverage. The Growth Balance method (Brass, 1971) is a useful model for detecting and adjusting for such errors and was used in this study. The model is based on the linear relationship of deaths and age distributions that may be expressed as:

$$N(x)/N(x+) = r + k(D(x+)/N(x+))$$

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 which implies that there should be a straight line relationship of intercept r , slope k , a coefficient of the estimated ratio of true to reported deaths (a factor representing the completeness of reporting of deaths). Further details about the method are given in United Nations (1983), Hill (1987).

There was a large number of unspecified age and sex in the data that amounted to about 23.5% of the total number of reported household deaths. These cases were not excluded from the analysis because doing so, would have exaggerated the magnitude of incomplete reporting of household deaths. The solution was to prorate these cases on the assumption that the distribution of the unknown cases is similar to the distribution of the known cases. The assumption may not necessarily be correct but is difficult to establish otherwise from the data.

Maternal mortality from the distribution of pregnancy related deaths in households.

The WHO defines a maternal death as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and the site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes. The 2011 census included questions about how many members of the household passed away in the last 12 months (before the census) as well as the age and sex of the deceased and whether the cause of death was natural or unnatural. If the deceased was a female aged 12-50 years, it was then asked whether she was pregnant at the time of death or whether death occurred within six weeks after childbirth. Maternal mortality ratio (the number of maternal deaths per 100,000 live births) – the most commonly used measure of maternal mortality – was estimated from the reported distribution of pregnancy related deaths as follows.

If the data are perfect, unadjusted maternal mortality ratio, MMR is computed as:

$$MMR = \frac{D}{B} k$$

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 while k is a constant, usually 100,000 (United Nations

2003: 36). But because the data are not perfect, adjusted maternal mortality ratio MMR' was estimated as

$$MMR' = \frac{D'}{B'} k$$

Where D' is adjusted number of maternal deaths in the period, B' is adjusted number of live births in the period. D' was estimated by fitting the Growth Balance method to the reported distribution of pregnancy related deaths under the assumption that errors in the reported distribution was constant for all causes of deaths (an assumption that is not necessarily correct). B' was estimated by fitting the relational Gompertz model to the fertility reports and estimating model age specific fertility rates from the model (see above) as

$$f(x)' = F \cdot F(x)'$$

Where $f(x)'$ is the decumulated model age specific fertility rate for women aged x and $F \cdot F(x)'$ is the model cumulative fertility up to age x . The adjusted number of live births was then estimated as

$$B' = \sum_{x=15}^{49} (w(x) \cdot f(x)')$$

Where B' is the adjusted number of live births for a specific period and $w(x)$ is the number of women aged x in a five-year reproductive age group.

Maternal mortality using Life table method based on information on household deaths

An alternative approach was to estimate maternal mortality ratio indirectly using life table methods and is summarised as follows.

$$MMR' = \sum_{x=15}^{49} \{ [(w(x) \cdot m(x)')] \cdot p / B' \} k$$

Where MMR' is the estimated mortality ratio for the time period (2011), $w(x)$ is the number of women aged x in a five-year reproductive age group, $m(x)'$ is the adjusted (using the Growth Balance method) life table age specific death rate for women aged x in a five-year reproductive age group, B' is the adjusted (using the relational Gompertz model) number of live births for the given period, p is the proportion dying of maternal causes of all causes of death among women of reproductive age. The proportion may come from an external data source within the same population. One of the challenges in this approach is a reliable estimate of p .

Estimating net migration

Estimating net migration is problematic worldwide with the exception of the Scandinavian countries that operate efficient population registers.

In the context of South Africa the challenge arises from the following: (1) out-dated data on immigration and emigration: there has been no new processed information on immigration and emigration from Stats SA (due to lack of data from the Department of Home Affairs) since 2003. (2) Although information on in- and out-migration as well as immigration can be obtained from the censuses, censuses are not appropriate and usually do not collect information on emigration. (3) The complication of undocumented migration. Although the manner the census questions on migration were phrased in South Africa's censuses, theoretically captured both documented and undocumented migrants, it is possible that undocumented migrants may not have provided accurate information about their country of birth.

Bearing the above in mind, estimating current trends in inter-provincial net migration in this study was based on the 2011 Census questions on province of usual residence and province of previous residence by year of move since 2001. If the answer to the question on: "Has this person been living in this place since October 2001" was "yes", then the person did not move between October 2001 and October 2011. The opposite was the case if the answer to the question was "no". Migration matrix tables were obtained by tabulating province of usual residence by year of move to obtain in-migration and province of previous residence by year of move to obtain out-migration. The tabulations included foreign born (as this category could be identified). Thus, the results presented are for the volume of net internal migration and immigration but excludes emigration for reasons noted above. Annual growth rates in interprovincial net migration plus immigration were compute using the geometric formula of the exponential form:

$$NM_{t+n} = NM_0 \cdot e^{rt}$$

and

$$r = [LN(NM_t / NM_0)] / t$$

where NM_{t+n} is the number of net migrants at time $t+n$, NM_0 is the number of net migrants at an initial period, r is the annual growth rate, NM_t is the number of net migrants at a later date after the initial period and t is the interval between the initial period and the later date.

Results

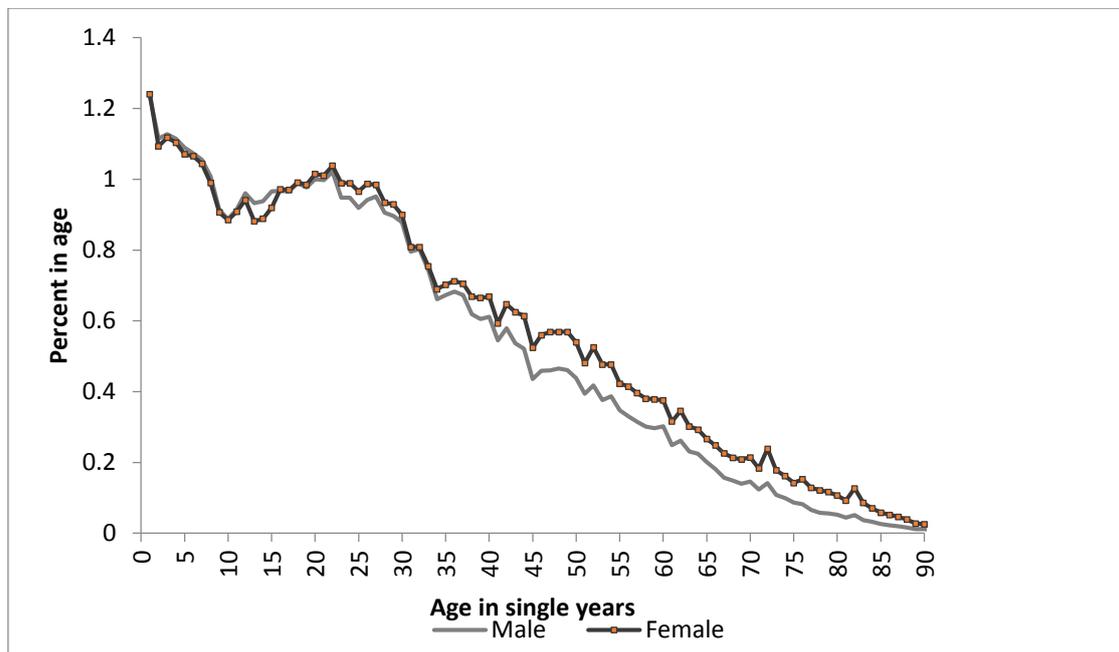
The following section presents the results and in doing so, the internal and external consistency of the estimates are highlighted. Demographic phenomena

are not independent of each other and operate jointly to produce a particular outcome. Thus, checks on internal consistency of demographic estimates entails examination of the holistic internal coherence of the data. External consistency in this study entails comparing one or several demographic estimates from the 2011 census data with one or several other independent sources of data. It should however be borne in mind that there is no gold standard that can be used as a benchmark for assessing the 'correctness' of demographic estimates as the external sources themselves may contain biases.

Age Distribution

The age-sex distribution of a population is a crosscutting variable in planning since all sectors need to take this into consideration directly or indirectly, hence the need for reliable information on age-sex composition. As can be seen from Figure 1 the quality of the age report in the 2011 census was reasonably good. However, heaping or digit preference are present indicating preference for ages ending in even numbers and zeroes.

Figure 1. Single year age-sex distribution from the 2011 Census



Source: Own computation from 2011 South Africa Census data

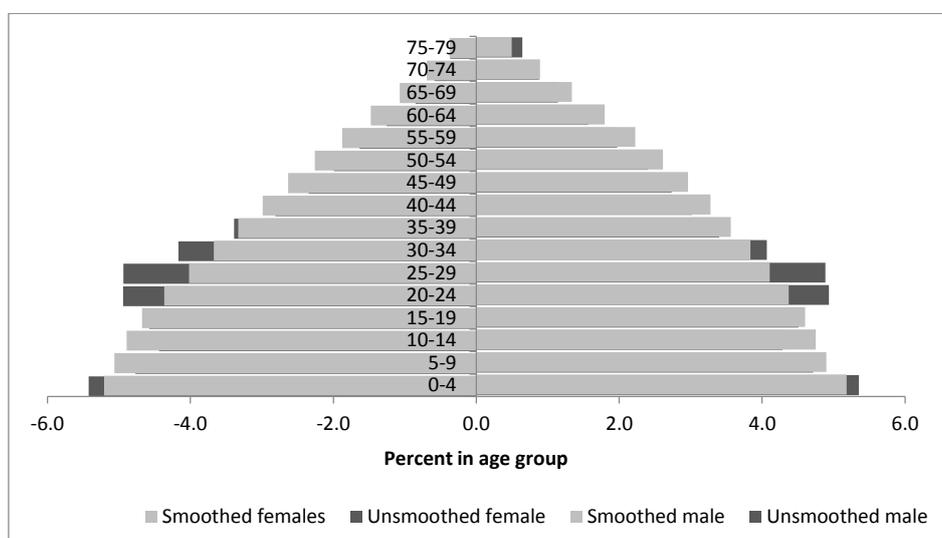
The five-year population pyramid (see Figure 2 unsmoothed age distribution) shows three striking features. Firstly, there is a bulge in the 0-4 age group. This was not the case in the 1996 and 2001 censuses, instead, there was an undercutting in that age group that were inconsistent with the country's fertility and mortality levels trends (Udjo, 1998). Secondly, there are indentations in the 5-9 and 10-14 age groups. The theoretical explanations are: (a) excess childhood mortality or (b) sharp decline in fertility in the 10-14 years ago. Although fertility has declined in South Africa in the last few decades, the decline is not erratic, so a sharp decline in fertility 10-14 years ago cannot be accepted as the explanation for the indentation in the 5-9 and 10-14 age groups in the 2011 census. Although the quality of the child

mortality data from previous censuses and surveys had limitations there are no indications that the country experienced unprecedented high mortality among children 10-14 years ago. Such a situation usually arises from natural disasters like famine, war etc. Thus, high mortality among children cannot be accepted as the reason for the indentation in the 5-9 and 10-14 age groups in the 2011 census. (c) Massive emigration of persons aged 5-9 and 10-14 in the last 5-10 years: This is unlikely because if that were the case, one should have observed indentation in the higher ages corresponding to the ages of the parents in the population pyramids since children are often taken along when their parents emigrate. (d) Upward age shifting of persons aged 5-9, and 10-14: This is unlikely as the usual pattern of age shifting is

downwards except among elderly people who tend to exaggerate their age. (e) Under- enumeration of persons aged 5-9 and 10-14: By a process of elimination therefore, it appears the most likely explanation of the indentation in the 5-9 and 10-14 age groups is under-enumeration of persons in that age group in the 2011 census. Thirdly, there is a bulge in the 15-29 age group possibly due to downward age shifting of older persons. For planning purposes, these seemingly anomalies were “corrected” through smoothing. The results (see Figure 2 smoothed age distribution) indicate that the percentages aged 0-4, 20-24, 25-29, 30-34 and 35-39 and among females

aged 75-79 in the unsmoothed distribution are slightly higher than those in the smoothed age distribution. It would appear from this that the seemingly age errors in the 2011 census age-sex distributions were largely in the age groups 20-34 and to a lesser extent in the 0-4 and 75-79 age groups. Smoothing of the provincial age-sex distributions was not attempted because at provincial levels, internal migration compounds the usual pattern of age mis-reporting making it difficult to disentangle migration impact from age mis-reporting.

Figure 2: Percent in smoothed and unsmoothed five-year age distribution



Source: Own computation from 2011 South Africa Census data

Age at first marriage

The singulate mean age at marriage derived from the 2011 census among females (29.6 years) is high by African and European standards and this was primarily due to the value derived from the Black population group. The value for males is 32.2 years. If accepted as correct, it would mean that South African females on average have one of the highest mean age at first marriage in the world: Tunisia (20.7), Senegal (21.6), Uganda (20.0), Botswana (26.4), Sweden (31.4), Switzerland (29.8), India (20.2). (United Nations 2012). The high singulate mean age at marriage observed from the 2011 census was probably due to African/Black females and to a lesser extent Coloured females not reporting their true marital status category - some people reporting they

have never been married when in fact they were divorced, separated or widowed. The observed singulate mean age at marriage were 30.3 years, 28.4 years, 25.8 years and 25.3 years among Blacks, Coloureds, Indians and White females respectively and for males, it was 33.0 years, 30.3 years, 28.8 years and 27.8 years respectively. The phenomenon of high singulate mean age marriage was also observed in the 1996 and 2001 censuses as well as in the 2007 community survey. The application of van De Walle’s method described earlier yielded lower age at first marriage (26.1 years for female and 27.4 years for males nationally) than the observed singulate mean age at marriage. This suggests that part of the high singulate mean age marriage observed was due to age mis-reporting.

Table 1: Estimated Gompertz parameters and total fertility rates from the 2011 census data

	α	β	TFR
Population Group			
Blacks	0.086	0.965	2.8
Coloureds	0.133	1.036	2.6
Asians	-0.163	1.183	2.1
Whites	-0.292	1.131	1.9
Provinces			
Eastern Cape	0.092	0.981	2.8
Free State	0.095	1.018	2.5
Gauteng	0.041	0.977	2.4
KwaZulu-Natal	0.162	0.995	2.7
Limpopo	-0.002	0.987	3.1
Mpumalanga	0.109	0.943	2.9
Northern Cape	0.129	0.989	2.9
North West	0.072	1.001	2.9
Western Cape	0.061	0.975	2.4
National	0.085	0.978	2.7

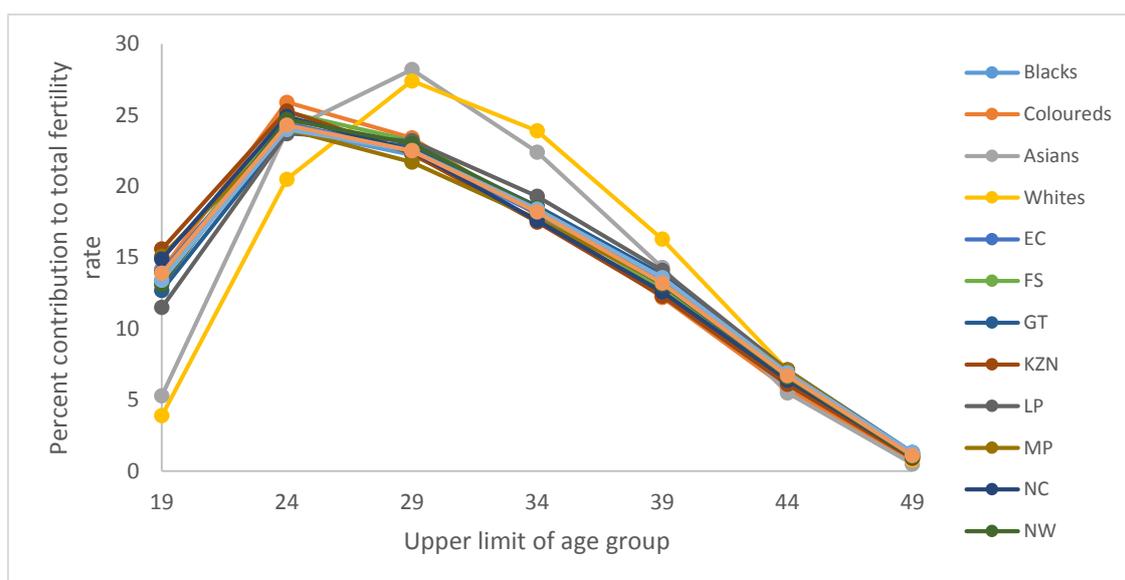
Source: Own computation from 2011 census data

Fertility

Figure 3 shows the results of the fitting of the relational Gompertz model to the reports on children ever born (P points) and births in the last 12 months (F points) in the 2011 census. As seen from the graph, all the points corresponding to the younger women lie very close to a straight line indicating the quality of the fertility reports from the younger women was reasonably good. The curving downward of the P points is an indication that older women tended to overstate the number of children ever born during the census. These children may be "adopted" children (not biological children). Table 1 shows the estimated Gompertz parameters (α and β) and the total fertility rates (TFR). The α values are not consistent across population groups and provinces - while a few are negative, most are positive - and

therefore difficult to draw firm conclusions regarding the timing of fertility implied by the α values. The β values appear to suggest that child bearing is concentrated in a wider age band among Blacks and in most provinces (β lower than 1) than among the other population groups (β greater than 1), the Free State and North West. The estimated Gompertz parameters at national level imply a total fertility rate of 2.7 which is slightly higher than the observed total fertility rate (2.5). The estimates indicate that Whites had the lowest (and below replacement) level of fertility (1.9) while Blacks had the highest (2.8) as of 2011. The estimated total fertility rates in the provinces in 2011 ranged from 2.4 in the Western Cape and Gauteng to 3.1 in Limpopo.

Figure 3: Relative Age specific fertility rates (%) from the 2011 census data



Source: Own computation from 2011 census data

The estimated relative age specific fertility rates (with total fertility rate = 1) are shown in Figure 4. The estimates indicate the percentage contribution of each reproductive age group to total fertility and also depicts the age pattern of fertility controlling for differences in total fertility rates. The graph shows that at national level, the 15-19 age group contributed about 14% to the country's total fertility rate in 2011 which suggests a substantial level of teenage fertility. The graph further suggest that teenage fertility is substantial among Blacks, Coloureds and all the provinces but low among Indians/Asians and Whites.

While among Blacks, Coloureds and in the provinces, relative age specific fertility rate was above 10% in the age group 15-19 in 2011, it was between 4-5% among Indians/Asians and Whites in the same period. Furthermore, the estimates suggest that among Indians/Asians and Whites, while the peak age of childbearing was in the 25-29 age group in 2011, it was in the 20-24 age group among Blacks and Coloureds in the same period. In the provinces, the peak age of child bearing was in the 20-24 age group in 2011.

Table 2: Estimated probability of dying between birth and age x from children ever born and Children dead, 2011 census

Age of children, x	Probability of dying between birth and age x, (q_x)	Years before 2011 census
1	0.022	1.7
2	0.030	3.4
3	0.035	5.4
5	0.038	7.7
10	0.039	10.2
15	0.043	13.1
20	0.058	16.6

Source: Own computation from 2011 census data

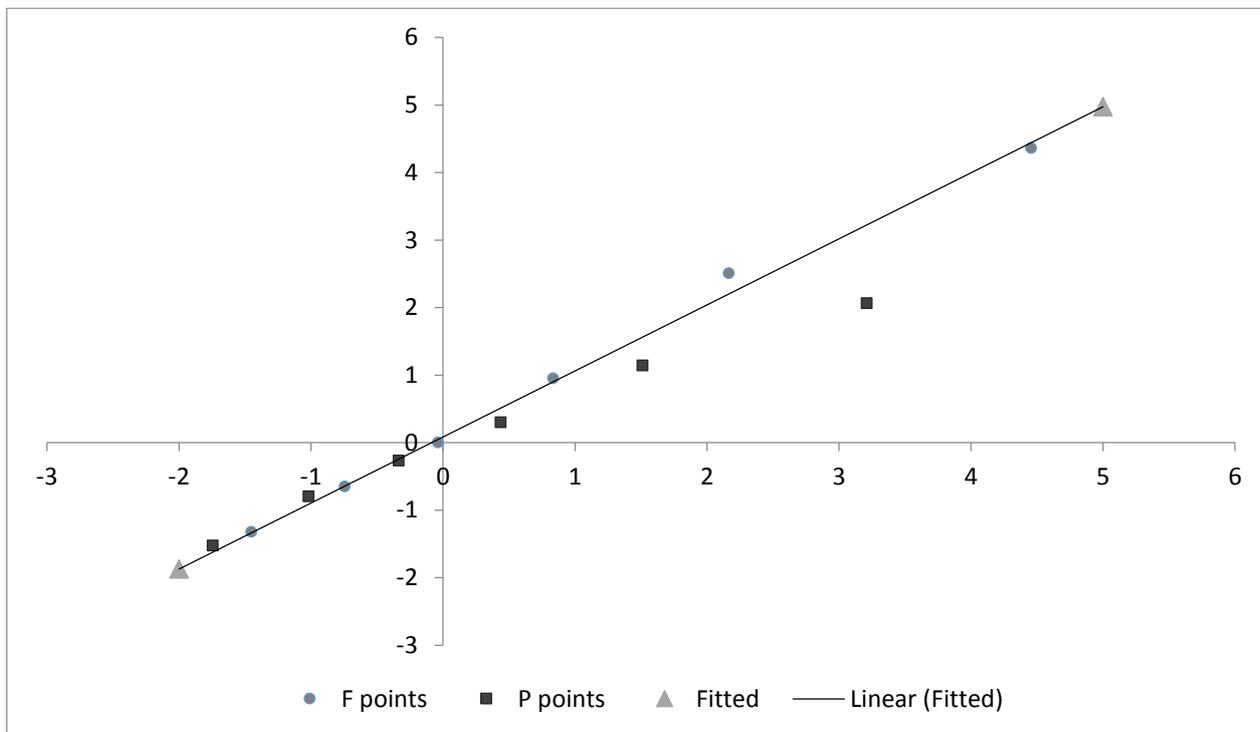
Mortality

Childhood mortality

Table 2 shows the estimated probability of dying between birth and age x q_x from the 2011 census both sexes combined. Both sexes were combined in the

estimation due to some odd pattern observed in the data – higher female than male childhood mortality that could not be explained. In the absence of errors in the reporting of the number of children

Figure 4: Fitting the relational Gompertz model to current (F) and life time fertility (P), 2011 Census, National



Source: Own computation from 2011 South Africa Census data

surviving the probability of dying between birth and higher ages, q_x should increase with mother's age due to longer exposure to the risk of dying of children of older mothers. As can be seen from Table 2, the probabilities increase monotonically and are more or less constant between ages 2 and 10. This is an indication of substantial underreporting of dead

children by mothers in the 2011 census since children do not cease to die after age 2. Therefore estimates of childhood mortality based on these probabilities at any age would underestimate childhood mortality. In view of this, estimates of childhood mortality was therefore derived from information on household deaths. The results are shown in Table 3.

Table 3: Estimated mortality indicators for 2011 from household deaths (both sexes)

	Infant mortality rate (q_x)	Uder five mortality rate ($5q_0$)	Life expectancy at birth (yrs, e_0)
Provinces			
Eastern Cape	57.3	75.6	51.0
Free State	80.6	104.1	53.6
Gauteng	56.3	75.1	57.0
KwaZulu-Natal	79.2	102.9	50.3
Limpopo	49.7	67.0	57.2
Mpumalanga	72.3	96.0	56.1
Northern Cape	56.2	76.5	60.0
North West	68.5	89.4	59.0
Western Cape	27.2	36.1	66.0
National	60	80	55

Source: Own computation from 2011 census data

Table 4: Annual percent growth rates in net internal migration plus immigration 2002-210 by sex

	Males	Females
Provinces		
Eastern Cape	30.7	25.8
Free State	19.4	16.4
Gauteng	17.7	21.1
KwaZulu-Natal	23.1	23.5
Limpopo	24.9	26.3
Mpumalanga	20.9	21.2
Northern Cape	29.1	22.0
North West	18.7	22.8
Western Cape	23.8	23.4

Source: Own computation from 2011 census data

In contrast with infant mortality rate of 29 per thousand live births derived from the proportions dead of children ever born the estimates derived from information on household deaths indicate infant mortality rate of 60 per thousand live births in 2011. The results also indicate that provincial disparities in infant mortality rates are almost similar to that of

under-five mortality rates in South Africa. While the Western Cape had the lowest infant mortality rate (27 per thousand live births), the Free State had the highest (81 per thousand live births) in 2011. Life expectancy at birth is largely driven by the level of infant mortality. The values in Table 3 appear to suggest that while the Western Cape had the highest

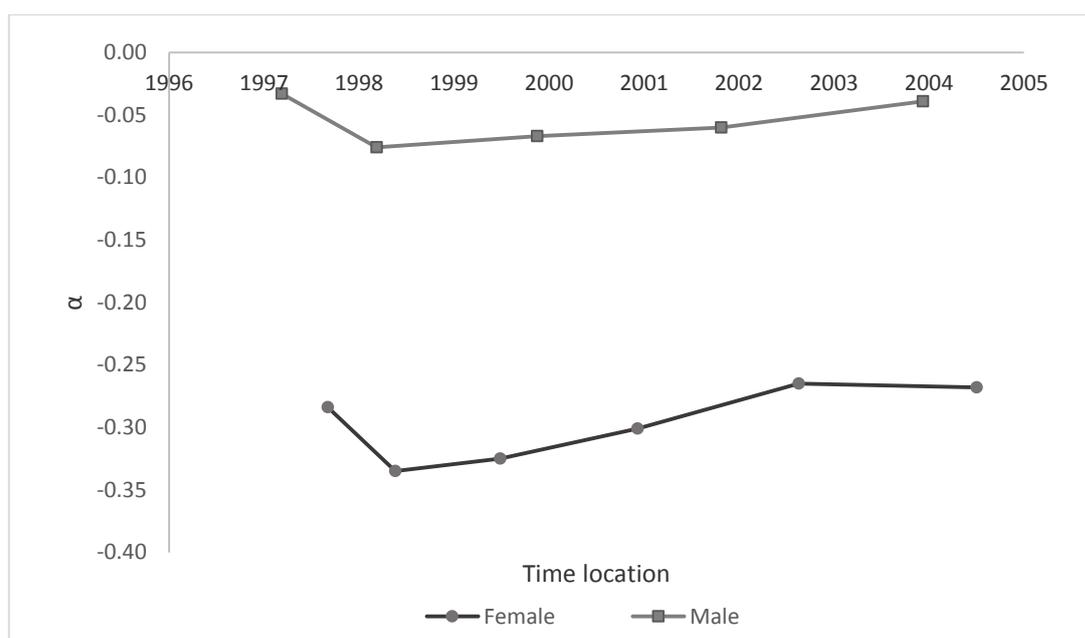
life expectancy at birth, KwaZulu-Natal had the lowest in 2011. Estimates of mortality by population group from the information on household deaths in the 2011 census (and in previous censuses) is not possible because information on the population group of the deceased was not asked.

Adult mortality from orphanhood

Estimates of adult male and female mortality can be derived from orphanhood questions. The summary of the results at national level with regard

to the 2011 census is illustrated in Figure 5. The trends indicate increasing adult mortality since the late 1990s. Usually, estimates of adult mortality orphanhood are combined with estimates of childhood mortality from children dead of children ever born to produce estimates of mortality for the general population. In view of the poor quality of the information on children surviving, as indicated above, further disaggregation in the orphanhood estimates was deemed fruitless in this study.

Figure 5: Trends in adult male and female mortality, 2011 census



Source: Own computation from 2011 South Africa Census data

Maternal mortality

Maternal mortality ratio was estimated from the distribution of pregnancy related deaths reported in households. The results indicate a maternal mortality ratio of 271 per 100,000 live births in 2011. This value is inconsistent with estimated values derived from the 2001 census (473 per 100,000 live births) and the 2007 Community Survey (764 per 100,000 live births) using the same methods. For such apparent sharp decline in maternal mortality ratio in a short period, there would have had to be an intensive and wide scale intervention specifically targeting reduction of maternal mortality. There is no evidence that there was such an intervention during the period.

The estimated proportion, p dying of pregnancy related deaths of all causes of death among women of reproductive age after prorating the non-response

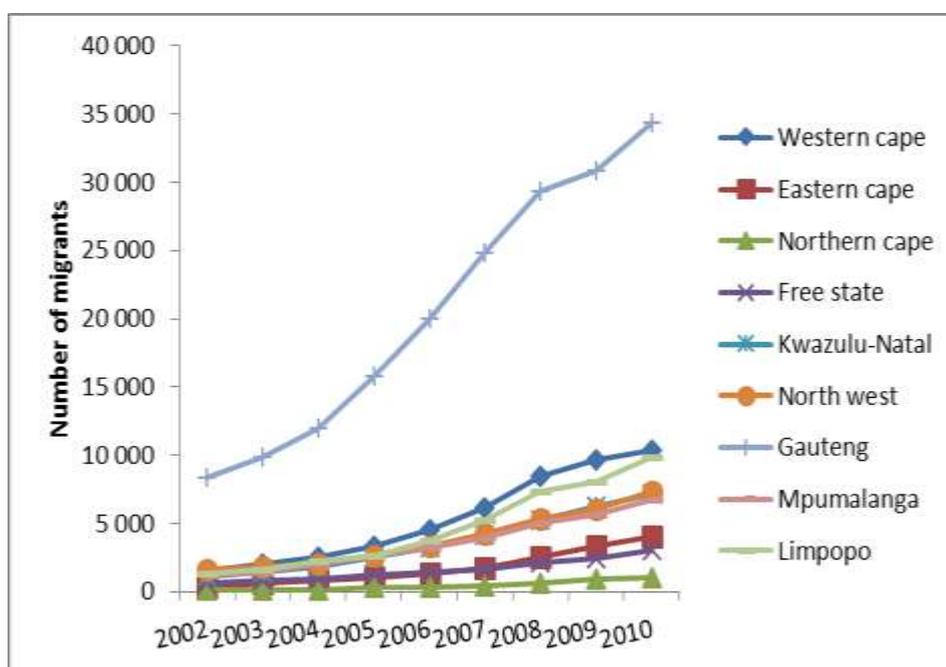
and unspecified cases in the 2011 census data on household deaths was 3.2%. Substituting this proportion in the life table approach of estimating maternal mortality ratio, yielded a value of 397 per 100,000 live births. This value is also inconsistent with the estimates derived from the 2001 census and 2007 Community survey. The life table approach is very sensitive to errors in the value of p . Boerma (1987) observed that the minimum level of the proportion of deaths due to maternal causes by women of reproductive age in sub-Saharan Africa (i.e. p) was 10%. Although this proportion may have declined, it would appear that the proportion derived from the 2011 census was underreported. (A 5% value of p and using the life table approach yields a maternal mortality ratio of 613 per 100,000 live births). In view of these, the estimated maternal mortality ratio of 271 or 397 per 100,000 live births from the census data should be regarded as unreliable.

Net migration

As already indicated, it is often not possible to obtain information on emigration from censuses hence was excluded in the estimates. The periods 2001 and 2011 were excluded from the estimates of migration since the information pertaining to these years was not for a full year in each case, thus the estimates presented pertain to the period 2002-2010. The results indicate that net migration was positive in all the provinces during the period (Figure 6) due to immigration. Without immigration (i.e. if the estimates only focused on in-migration and out-migration of South African born), some of the provinces would have experienced negative net

migration but this would not have been a good reflection of net inter-provincial migration as foreign born also move into and between provinces. The results further indicate that among males, the growth in net migration in absolute terms was largest in Gauteng province during the period (it increased from about 8,333 in 2002 to about 34,351 in 2010). The Northern Cape experienced the least volume of net internal migrants and immigrants during the period (the number of net migrants increased from about 96 in 2002 to about 982 in 2010). A similar pattern was observed among females (graph not shown).

Figure 6: Trend in Interprovincial net migration: Males



Source: Own computation from 2011 South Africa Census data

The provincial pattern in growth rate in net internal migration plus immigration is shown in Table 4. As seen in the table, even though Gauteng province experienced the largest volume in net interprovincial migration during the period 2002-2010 (Figure 6), it was not the province with the highest annual growth rate in net migration. Instead, the Eastern Cape experienced the highest annual growth rate in net interprovincial migration plus immigration (30.7% among males, 25.8% among females). Gauteng experienced the least annual growth rate in net interprovincial migration plus immigration among males (17.7%) during the period.

Discussion and conclusion

The results of this study show that estimating demographic parameters from censuses poses a challenge. As with other censuses elsewhere South Africa's 2011 census data have certain strengths and weaknesses as seen from the results of this study. Censuses worldwide have quality issues though in varying degrees. Most of the quality issues relate to content errors as coverage errors are usually easier to deal with, especially if a post-enumeration survey was carried out. There is no known standard in judging content errors in demographic data from censuses, but present demographic knowledge and

use of appropriate demographic models do provide useful insights.

Although, it was not possible to estimate childhood mortality from the questions on children ever born and children dead from the 2011 South African census, the estimates derived from the report on household deaths provide indication of childhood mortality. Aside this and using various models, demographic parameters estimated from the 2011 South African census nationally in 2011 include mean age at first marriage (26.1 years for females and 27.4 years for males), total fertility rate (2.7), infant mortality rate, (60 per thousand live births both sexes), under-five mortality rate (80 per thousand live births both sexes), life expectancy at birth (55 years both sexes), growth in provincial net migration (Gauteng province with the highest net positive migration in absolute terms).

It is not uncommon for demographers and other researchers to estimate demographic and other population parameters from census and survey data without taking into account errors that may be present in the data. For example, regression analyses are quite common among demographers but it is often not clear in some of the analyses whether the estimated differences among sub-groups are due to real differences or spurious differences arising partly from differentials in the quality of the reports among the sub-groups. Coale and Trussel (1996) have cautioned about the use of demographic data without comparison with demographic models to assess their validity. They noted that the techniques demographers use to analyse these data (including Demographic and Health Surveys) are increasingly the techniques of statistics particularly event-history analysis. They further noted that while the trend toward the use of sophisticated statistical models appropriate for the problem to be analysed and the data that are available are healthy, the trend toward accepting demographic data at face value is not. In providing demographic estimates to aid socioeconomic planning it is incumbent on demographers and other researchers providing such estimates to be wary of the quality of the data by providing adequate assessment of the quality of the data. Statistically less developed countries are not yet at the stage where their census and survey data can be accepted at face value for statistical and demographic estimation. Doing so may very well miss lead policy makers and planners. A critical modelling of the data is essential. This has been the

main task in this study. Overall, the results indicate that despite some weaknesses in some parts of the 2011 South African census data, it is a rich data source for demographic and other population estimation for informing socioeconomic planning for South Africa.

Acknowledgement

I wish to thank Statistics South Africa for providing access to its data. The views expressed in this paper are however, those of the author.

References

- Blacker, J.G.C. 1977. The estimation of adult mortality in Africa from data on orphanhood. *Population Studies*, 31: 107-128.
- Boerma T (1987). The magnitude of the maternal mortality problem in sub-Saharan Africa. *Social Science and Medicine*, 24(6): 551-558.
- Booth, H. 1979. The estimation of fertility from incomplete cohort data by means of the transformed Gompertz model. PhD Thesis, London School of Hygiene and Tropical Medicine, University of London.
- Brass W (1971). Methods for estimating fertility and mortality from limited and defective data. Chapel Hill, NC: University of North Carolina.
- Brass, W. 1974. "Perspectives in population prediction: illustrated by the statistics of England and Wales". *Journal of the Royal Statistical Society, Series A* 137 (4): 532-583.
- Brass W (1981). The use of the Gompertz relational model to estimate fertility. Paper presented at the International Union for the Scientific Study of Population Conference, Manila, 3: 345-361.
- Brass W (1985). A simple approximation of the time location of estimates of child mortality from proportions dead by age of mother, in *Advances in methods for estimating fertility and mortality from limited and defective data*. Centre for Population Studies, London School of Hygiene and Tropical Medicine, Working Paper no 1-16.
- Brass W, Bamgboye EA (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.

- Coale A, Trussel J (1996). The development and use of demographic models. *Population Studies*, 50(30):469-484.
- Dorrington R. (1999). To count or to model that is not the question: some possible deficiencies with the 1996 census results. Paper presented at the Arminel Roundtable workshop on the 1996 South Africa Census, Hogsback, 9-11 April.
- Groenwald P, Nanan N, Bourne D, Laubscher R, Bradshaw D (2005). Identifying deaths from AIDS in South Africa. *AIDS* 19(2): 193-201.
- Hajnal J (1953). Age at marriage and proportion marrying. *Population Studies* 7(2):129-131.
- Hill, K. 1977. Estimating adult mortality levels from information on widowhood. *Population Studies*, 31: 75-84
- Hill, K. 1987. Estimating census and death registration completeness. *Asian and Pacific Forum*, 1, (3): 8-13.
- INDEPTH Network (2004). INDEPTH model life tables for sub-Saharan Africa. Aldershot: Ashgate Publishing Limited.
- Mandela, N. R. 1994. The long walk to Freedom. Transvaal, South Africa: Macdonald Purnell, 10.
- Phillips HE, Anderson BA, Tsebe P (1999). Sex ratios in South African census data 1970 – 1996. Paper presented at the Workshop on Phase 2 of Census 1996 Review on Behalf of the Statistical Council, Johannesburg 3-4 December.
- Sadie JL (1999). The missing millions. Paper presented to NEDLAC meeting. Johannesburg.
- Shell R (1999). An investigation into the reported sex composition of the South African population according to the census of 1996. Paper presented at the Workshop on Phase 2 of Census 1996 review, Wanderers Club, Johannesburg, 3-4 December.
- Statistics South Africa (1998). The people of South Africa, Population Census, 1996. Pretoria: Statistics South Africa.
- Statistics South Africa (2003). 2001 Census: Census in brief. Pretoria: Statistics South Africa.
- Statistics South Africa (2007). Community survey 2007: Methodology, processes and highlights of key results. Pretoria: Statistics South Africa.
- Statistics South Africa (2012). Census 2011: Statistical release. Pretoria: Statistics South Africa.
- Timaeus, I. M. 1991. Estimation of mortality from orphanhood in adulthood. *Demography*, 28: 213-227.
- Udjo EO (1998). Additional evidence regarding fertility and mortality trends in South Africa and implications for population projections. Pretoria: Statistics South Africa.
- Udjo EO (1999). Comment on R. Dorrington's "To count or to model that is not the question". Paper presented at the Arminel Roundtable Workshop on the 1996 South African Census, Hogsback, 9-11 April.
- Udjo EO (2004a). Comment on Moultrie and Dorrington: Estimation of fertility from the 2001 South Africa census data. Statistics South Africa workshop on the 2001 population census.
- Udjo EO (2004b). Comment on Dorrington et al: Estimation of mortality using the South Africa census 2001 data. Statistics South Africa Workshop on the 2001 population census.
- Udjo, E. O. 2005. An examination of recent census and survey data on mortality within the context of HIV/AIDS. In, Zuberi, Sibanda, and Udjo (eds). The demography of South Africa. New York: M. E. Sharp, 90-119.
- Udjo EO (2006). Estimation of mortality from vital registrations in South Africa. *Current HIV Research*, 4: 469-474.
- United Nations. 1983. Manual X: Indirect Techniques for demographic estimation. New York: United Nations, 129-146.
- United Nations. 2003. Indicators for monitoring the millennium development goals: definitions, rationale, concepts and sources. New York: United Nations, p 36.
- United Nations (2012). World marriage data 2012. Available at: http://www.un.org/en/development/desa/population/theme/marriage-unions/marriage_estimates.shtml Accessed 5 December 2013.
- Van De Walle E (1968). Marriage in African censuses and inquiries. In, The Demography of Tropical Africa, Princeton: Princeton Press, p 203.
- Zaba, Basia. 1981. Use of the relational Gompertz model in analysing fertility data collected in retrospective surveys. Centre for Population Studies Working Paper 8 1-2, London School of Hygiene & Tropical Medicine.