The age distribution of younger people in South Africa: What can it tell us about recent fertility trends?

Rob Dorrington and Tom A Moultrie

Centre for Actuarial Research (CARe)
University of Cape Town
rob.dorrington@uct.ac.za

Abstract
This paper assesses recent trends in South African fertility. We examine the long-term trajectory of fertility in the country from previous censuses and demonstrate that the age distribution of children enumerated in the 2011 census suggests a quantum break from that trend.

The data from the 10 per cent sample of the 2011 South African Census provides two mechanisms for assessing current and recent fertility in the country. The first approach is to make use of the answers to questions on the date of birth of women's last-born child. This allows a direct estimate of fertility levels to be calculated. However, direct estimates of fertility from census data are frequently underestimated as a consequence of generalised failure to report accurately all recent births. The relational Gompertz model was derived and refined to compensate for this error.

The second approach is to make use of the enumerated population under the age of 15 by single years of age and estimates of childhood and adolescent mortality to estimate the numbers of births that would have occurred in previous years to produce the enumerated population of children under age 15. These numbers in conjunction with estimates of the numbers of women aged 15-49 in each of the past 15 years (derived using reverse survival; or interpolation from data from previous censuses; or demographic models – the differences will not be material) permit the derivation of total fertility rates for each of the 15 years, allowing the time trend in fertility to be assessed.

Preliminary results suggest that, if the census count of young children is correct, the total fertility rate in the country has fluctuated markedly, especially in the period from 2002-2009, with fertility rates initially showing a continuation of the past downward trend in fertility, albeit at a slightly lower level, followed by a sudden upwards reversal (of the order of 10-15 per cent) before levelling off.

We seek to resolve whether this pattern is real or artificial using the data from this census, from previous South African censuses, as well as from publicly available data on recorded births (adjusted for incompleteness). In doing so, we produce estimates of the past trend and current level of fertility implied by the census data for each of the four population groups.

Introduction

As Coale (1964) observed half a century ago, the long-term evolution of a population is most strongly affected by levels of and trends in fertility. Accurate information on these aspects is therefore essential for understanding future population dynamics, as well as to allow better planning and resource allocation. There are three principal sources of data from which fertility dynamics can be evaluated: vital registration systems; nationally representative surveys; and the data collected in national censuses.

The data that are available in South Africa to investigate fertility trends and patterns are limited. The notification of births is incomplete, and despite a marked improvement in registration in the mid-1990s and again in more recent years, still many births are registered several years after their occurrence (Statistics South Africa 2013b). And, although policies to end late registration of births in the country by 2016 are being implemented, the unknown, and changing, level of completeness hinders the use of vital registration data to assess fertility dynamics and patterns.

The second source of data on fertility in the country, namely the results from representative surveys, has also been problematic. Under apartheid, South Africa was excluded from the international programme of Demographic and Health Surveys (DHSs) conducted with assistance from USAID (although a survey employing a very similar instrument was conducted in 1987-9 by the HSRC (Mostert and Lötter 1990)). Official DHSs were conducted in the country in 1998 and 2003 (with the next planned for 2015), and although the quality of the 1998 survey was reasonably good, that of 2003 was so poor that the data were never released officially through the DHS programme's data portal and are only available upon request to the Department of Health. Other routine surveys (such as the General Household Survey) do not ask questions in sufficient detail to allow the monitoring of fertility trends and patterns, although some attempts have been made in this regard in the past (Udjo 1997; 2005). Similarly, a Community Survey conducted in 2007 in place of a census turns out not to have been a particularly reliable source of data on fertility.

Data from national censuses in South Africa are also not without problems. Prior to the end of apartheid in 1994, only two censuses, those conducted in 1936 and 1970, were believed to have enumerated the African population of the country with any degree of accuracy. Thus, the three post-apartheid censuses conducted in South Africa (in 1996, 2001 and 2011) are relied upon to provide crucial information to help demographers, planners and others to understand and assess current fertility dynamics in the country. However, the data from the three censuses each manifested problems: in 1996, because of the structuring of the questions and/or poor training, respondents and/or enumerators frequently did not appreciate the distinction between the questions on lifetime and recent fertility, providing the same answers to both questions (Moultrie and Timæus 2002). In 2001 the process of trying to remove errors from the data on fertility collected was such that only approximately half of the information provided was deemed to be plausible in the estimation of fertility rates (Moultrie and Dorrington 2004). The recently-released 2011 South African census provides demographers with a new dataset from which to estimate fertility trends and patterns.

This study draws on data from all of these data sources (vital registration, surveys and census data) to understand better the recent trend in fertility in South Africa. While methodological developments, notably the process of reviewing and preparing material for the updated UN *Manual X*, *Tools for Demographic Estimation* (Moultrie, Dorrington, Hill *et al.* 2013), has led us to revisit some of our previous estimates of fertility trends and patterns in the country derived from earlier census data, the focus of this paper is explicitly on what the 2011 census data can tell us about recent fertility dynamics in the country.

In addition to the general reasons provided above as to why the estimation of fertility is important, there is a further significant reason why the investigation of fertility patterns and trends described by the 2011 South African census is of particular interest to demographers and planners: one of the most unexpected results to come out of the 2011 Census was the reported age distribution of the population under age 20, presented in Figure 1. The figure indicates a pronounced trough in the population aged between ages 7 and 16 and implies within it a history of births/fertility over the last two decades that is distinctly at odds with what demographers had been estimating and projecting on the basis of the numbers of births reported by women in censuses and surveys prior to the 2011 Census.

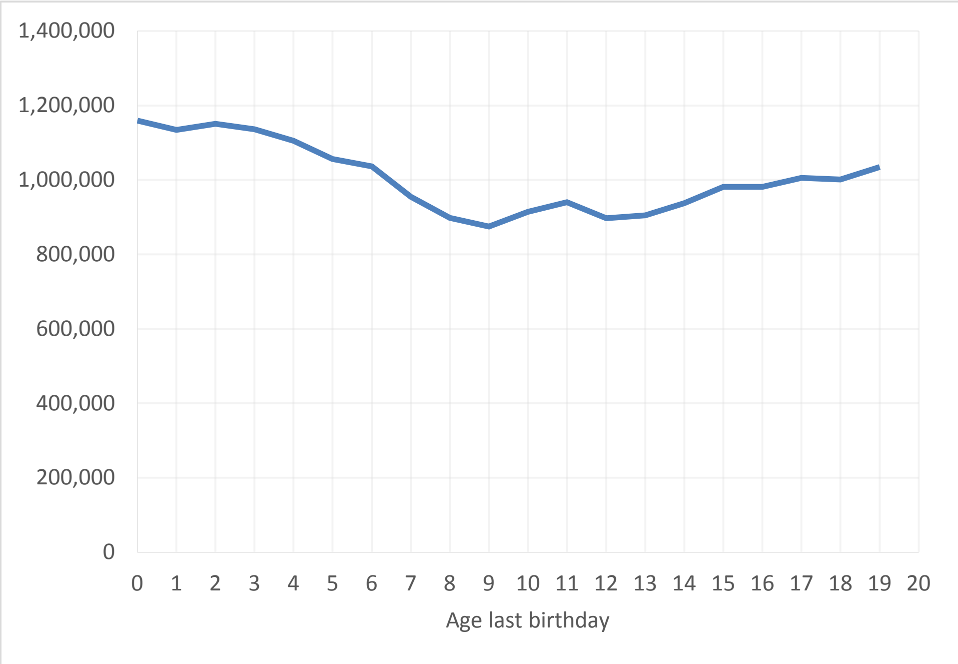

Figure 1: Age distribution of the population under the age of 20, 2011 South Africa Census

The number of people enumerated at any age, *x*, represents the number of survivors of births that occurred *x* years previously plus the survivors of immigrants less emigrants who migrated before age *x*. While the number of immigrants can be identified with some degree of accuracy using data collected on place of birth and removed from the analysis, the number of emigrants is more difficult to establish and, since the numbers are expected to be trivial relative to the stock of children and young adults enumerated in the census, have for the purposes of this exercise, been ignored. Hence, Figure 1 leads to the observation that the numbers of births in the country in the last two decades must have shown essentially the same pattern (but in reverse) to produce the enumerated population under the age of 20. However, prior to the release of the 2011 Census most demographic projections for the country (for example, those produced by the UN Population Division, the 2011 Mid-Year Estimates produced by Statistics SA, as well as the ASSA Aids and Demographic Model) assumed the numbers of births per annum to be roughly constant over this period (reflecting the underlying momentum of the population of women of reproductive age counterbalanced by the observed slow pace of fertility decline in the country since the 1960s (Caldwell and Caldwell 1993; Moultrie and Timæus 2003) with the number of births expected to show a gradual decline since the turn of the millennium. The resulting age distributions are compared to the Census in Figure 2.

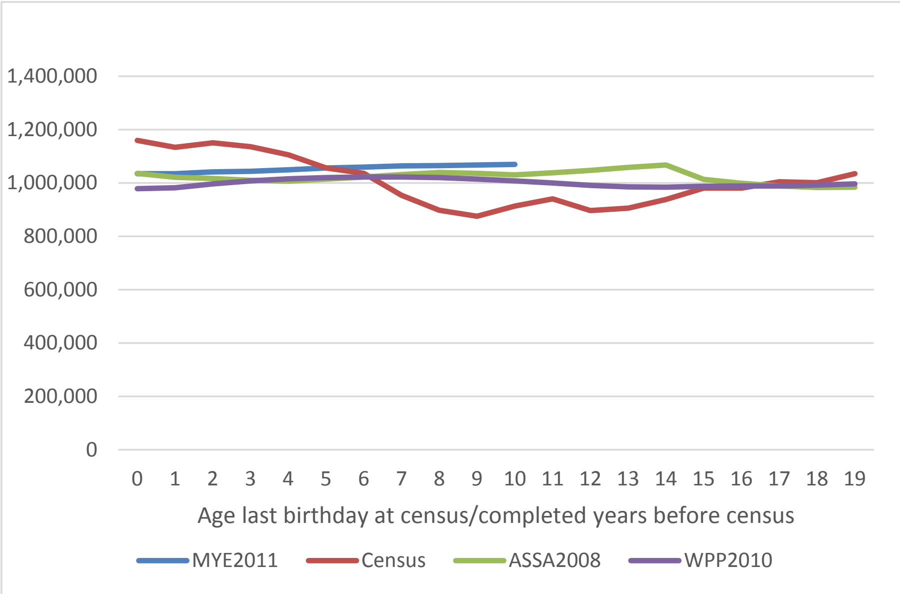

Figure 2 Comparison of the age distributions of the population under age 20: Pre-Census

The resolution of this discrepancy - the lack of close correspondence between the various models on the one hand, and the number of births implied by the 2011 census on the other – is important for understanding both the recent fertility history in the country, as well as its likely future demographic course.

Although it is now some two and a half years since the release of the initial results from the Census, this issue is yet to be resolved. Although one of the consultants contracted by Stats SA to help assess the Census, Griff Feeney, did show that it was possible to produce an age distribution similar to that of the Census using the numbers of registered births and deaths, by making suitable assumptions about the changing completeness of registration over time[1], this work, applied to a version of the Census prepared some months before the final version, is not in the public domain, and apparently, did little to convince Stats SA of the correctness of the Census count at these ages, since for the past three years they have produced mid-year estimates of the population with age distributions inconsistent with that of the Census (Bogue 1993; Dorrington 2013; Statistics South Africa 2013a; 2014; 2015)

As can be seen from Figure 3, Stats SA is not the only organisation unconvinced that the Census count reflects past fertility. This is a particularly surprising example since the UN Population Division (WPP2012) manage to reconcile their very different trend in births over time with an age distribution in 2011 that close to that of the 2011 Census, presumably through an implausible pattern of migration by age at the younger ages, which is not made public.

[1] Consultant report prepared for Stats SA and the Statistics Council, 2012.

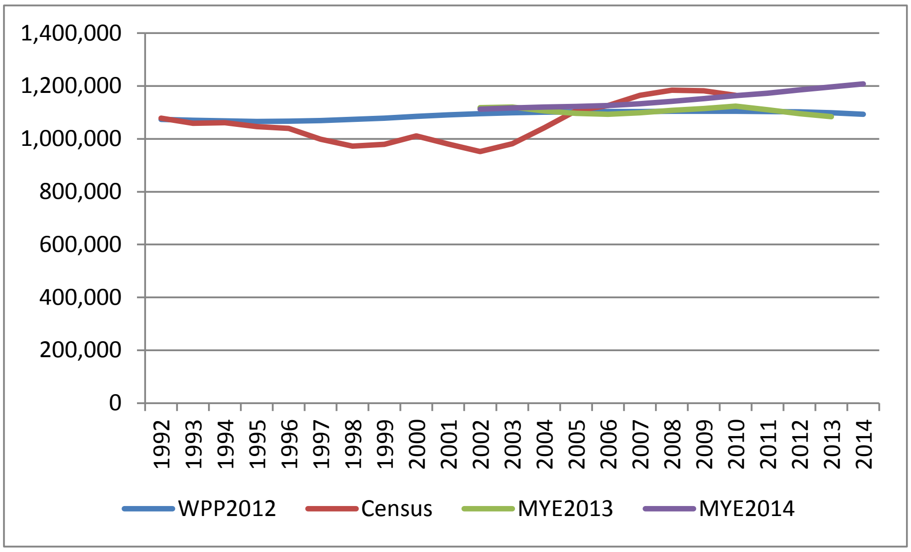

Figure 3 Comparison of numbers of births: Post-Census estimates

This paper is thus focussed on examining the consistency of the estimates of total fertility implied by the age distribution enumerated in the 2011 Census with previous estimates of fertility, in an effort to come to a view as to whether that age distribution could, despite its rather curious shape, be plausible.

Estimates of fertility based on reports on recent births

In order to frame the results from the investigations pursued in this paper, we present the estimates of total fertility derived from censuses and nationally-representative surveys conducted since the end of apartheid. The 1996 and 2001 censuses, as well as the 2007 Community Survey, collected fertility data allowing estimates of fertility to be derived from women's reports on births occurring in the year before the census date or interview date. These estimates apply, approximately, to a point 6 months before the census or survey. A fourth estimate of fertility in the post-apartheid era can be obtained from the data collected in the 1998 Demographic and Health Survey. Owing to the much smaller sample size, estimates of fertility derived from DHSs are usually based on the births occurring in the 3 years before the survey, and hence apply (approximately) to a point 18 months before the survey. Estimates of fertility from these inquiries (Moultrie and Timæus 2002; Moultrie and Dorrington 2004) are presented in Table 1.

Table 1 Estimates of total fertility, nationally and by population group from previous national enquiries

	National	African	Coloured	Indian	White
1996 Census	3.24	3.49	2.64	2.45	2.02
1998 DHS	2.89	3.11	2.53	1.80*	1.88
2001 Census	2.84	3.04	2.41	1.98	1.81
2007 CS	2.54	2.71	2.35	1.39	1.45

* The sample size of Indians in the 1998 DHS was exceptionally small, and this estimate cannot therefore be taken as accurate.

Estimates of recent fertility (typically for the year before the census) can be derived from the information collected in a census in which women report the date of their last child's birth. With highly accurate and high quality data, these estimates should provide a good and robust direct estimate of fertility. Unfortunately, however, the data so collected frequently suffer from systematic biases arising from reference period error and a generalised underreporting of recent fertility. Over the years, a number of methods have been proposed to correct for these errors, several of which are presented in the UN *Manual X* (UN Population Division 1983). A recent revision of that work (Moultrie, Dorrington, Hill *et al.* 2013) recommends that the most reliable and best method is the relational Gompertz method, first proposed by Brass in the late 1970s, and formalised by Zaba (1981) and Booth (1984).

The relational Gompertz method can be regarded as a successor to the earlier P/F technique, also pioneered by Brass. It is predicated on two observations about the general errors typically encountered in fertility data collected in developing countries, that recent fertility is generally underreported, but independent of women's age, and that lifetime fertility (the average number of children born to women in a particular age group at a particular point in time) is generally well-reported by younger women, but that lifetime fertility is increasingly underreported with increasing age of the women reporting. The method seeks to establish the shape of the underlying fertility schedule from the reports of recent fertility while correcting for the anticipated underreporting using the information on lifetime fertility. The method is explained in detail in Moultrie (2013b) with a downloadable spreadsheet that implements the method.

Estimating recent fertility in South Africa using the relational Gompertz model

As noted above, the relational Gompertz model requires directly calculated estimates of recent and lifetime fertility to be used as inputs. Recent fertility rates were derived as follows. Births in the last year were identified as all those reported as occurring between the 10 October 2010 and 9 October 2011 and extracted from the 10 per cent sample provided by Statistics SA. However, not all women enumerated in the census were asked questions about their fertility, for example, the question was not asked of those living in institutions. The population of women reporting births are therefore not the entire population of women in the data, but only those who were asked the question.

Further, because fertility differs markedly between population groups in South Africa, fertility rates were derived initially for each population group separately. Women who self-identified their population group as "Other" were distributed proportionally across the four principal categories. Average parities for women in each five-year age group and in each population group were calculated using conventional routines. An el-Badry correction (el-Badry 1961; Moultrie 2013a) to correct for the commonly encountered problem associated with enumerators failing to record a response to questions on fertility asked of childless women was appropriate for only the African population group: the diagnostic plot for other population groups was strongly non-linear. A relational Gompertz model was then applied to the input data. While the exact procedure whereby these estimates were derived are the subject of a monograph (Moultrie and Dorrington, forthcoming), the estimates are not hugely relevant for this exercise, save as a consistency check of the answers derived from the historical reconstruction of the trends in fertility based on the age distribution. However, it is worth noting certain results from the exercise:

- Parity data for Indian and White women were noticeably inconsistent with the reported recent fertility, especially at younger ages. This suggests errors in these data, arising either at the enumeration or editing stage.
- Because one of the aims of the forthcoming monograph is to examine fertility patterns at a provincial level, separate models were fitted to the data for African women in each province. Subsequent investigation suggested that there were no material differences in the resulting age patterns of fertility, so the shape of the national fertility schedule for African women was applied to reproduce the total fertility rate for Africans in each province.

- The generally small numbers of women in each age group and province among the other three population groups, and because - where estimates could be produced - the shape of the fertility schedules suggested that differences were not material, fertility patterns and levels for Coloured, Indian and White women were assumed to be invariant across provinces.
- National estimates of fertility were derived by weighting the age specific fertility rates by the numbers of women in each age group across provinces (for Africans) and by population groups (for non-Africans).

The final estimates of total fertility derived are shown in Table 2.

Table 2 Total fertility rates, by province and population group, 2011 South Africa Census

	Africans	Coloureds*	Indians*	Whites*	ALL
Western Cape	2.34	2.47	1.84	1.73	2.30
Eastern Cape	2.52	2.47	1.84	1.73	2.49
Northern Cape	2.63	2.47	1.84	1.73	2.51
Free State	2.50	2.47	1.84	1.73	2.45
kwaZulu-Natal	2.62	2.47	1.84	1.73	2.54
North-West	2.79	2.47	1.84	1.73	2.71
Gauteng	2.34	2.47	1.84	1.73	2.25
Mpumalanga	2.64	2.47	1.84	1.73	2.58
Limpopo	3.03	2.47	1.84	1.73	3.00
South Africa	2.58	2.47	1.84	1.73	2.49

* National rate assumed for the provinces

The national level of fertility in the year before the census is close to 2.5 children per woman. This is notable, principally, for the confirmation that fertility decline in South Africa would appear to have come close to stalling in the decade before the 2011 census. The national TFR described by the 1998 DHS was 2.9 children per woman; and 2.6 children per woman in the 2001 census[2].

The reconstruction of past trends in fertility using reverse survival methods shows that that story is somewhat more complicated.

Reconstruction of past trends in fertility

The approach adopted to attempt some kind of resolution of the discrepancy identified follows a number of steps. First we project the enumerated population backwards to determine the number of births in the last 15 years that would have produced the enumerated population at each age under the age of 15 in the 2011 Census. We then compare those estimates of the births with data from other sources: reverse-survival of the numbers counted in previous censuses and surveys; the registered births (adjusted for incompleteness); estimates of the births derived from the registered births adjusted for under-registration estimated from the death registration; births recorded by the National District Health Information System (DHIS) plus an estimate of the number of births occurring outside this system; and reported school enrolments (adjusted for level of attendance) to establish whether the time trend in the number of births is plausible.

[2] The estimates for 2001 used here differ from those published earlier (Moultrie & Dorrington, 2004) as a consequence of re-evaluation of the method used to derive the estimates. The previous research used the Relational Gompertz model to correct the shape of the fertility distribution and Feeney's adjustment (Feeney 1998) to determine the level of fertility based on the P/F ratio observed at the mean age of childbearing. Subsequent interrogation of this approach suggests that this adjustment is likely to overstate the level of fertility. Estimates for 1996 and 2007 were not changed because the impact on the estimate in 1996 was not as big and the estimate was consistent with estimates from other data, and because a different method was used to produce the estimate for 2007.

The approach employs a variant of reverse-survival analysis that allows the derivation of estimates of total fertility. Reverse-survival analysis of fertility is a relatively uncommon approach to estimating fertility trends, although it is suited to the task here. The most frequent criticism of the approach is that it is most sensitive to fluctuations and errors in the enumerated population of young children. But, since this distribution is precisely what we want to assess, the method is well suited.

Reverse survival requires estimates of cohort (as opposed to period) survival. However, the method is largely insensitive to this assumption as even large errors in mortality translate into rather small errors in survival[3]. In order to take the time-varying effects of HIV mortality into account, cohort survival probabilities were derived from the ASSA2008 models.

The effects of migration also need to be considered. Young children recorded in the Census as having not been born in South Africa were netted off the enumerated population before estimating the number of births in order to provide an estimate of the survivors of the South African-born population. Although it is possible to make a number of alternative assumptions as to how to handle the migration of young children the numbers of migrants are, by and large, very small relative to the numbers of births, and so the estimated number of births is not particularly sensitive to the approach adopted.

Application of the reverse-survival approach produces estimates of the number of births nationally shown in Figure 4.

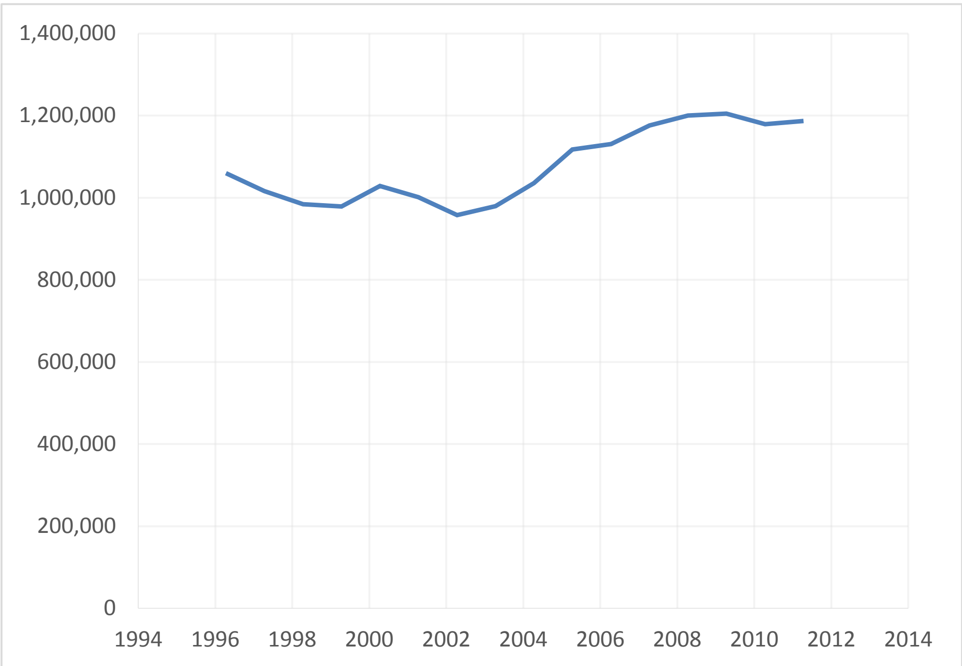

Figure 4 Estimated number of births in South Africa 1996-2011, 2011 South Africa census

The most remarkable aspect of this trend in the numbers of births is that, if correct, it suggests, that the number of births increased from a roughly constant 1 million per annum in the late 1990s and early 2000s to around 1.2 million per annum in the four years prior to the Census. (The clearly visible "hump" in 2001 is probably a reflection of age-heaping of children reported

[3] As a simple example, if under-five mortality was 50 per 1000, the proportion surviving would be 0.95. If that estimate of mortality was overstated by 100 per cent (i.e. an estimated under-five mortality of 100 per 1000), the proportion surviving would be 0.90. Thus, the error in the survival factor arising from over-estimating the level of child mortality by a 100% would translate into an error of a little over 5 per cent in the number of births.

to be aged 10 in the 2011 Census or those attributed a year of birth of 2000 (i.e. aged 11 in the 2011 Census).)

What this comparison does not permit, however, is the detailed assessment of whether this pattern is real (and thereby reflects some combination of rising fertility rates and population momentum) or merely a reflection of systematic errors in the data. However, data from other sources can be used to shed light on this question. In order to validate the pattern of the number of births declining through to 2001, increasing between 2001 and 2009 and then (seemingly) beginning to fall again thereafter.

To gain further insight into this matter, we consider a number of alternative data sources to determine the reasonableness of these numbers of births, namely:

- The TFRs from censuses (1996, 2001 & 2011), the CS and 1998 DHS estimated from the numbers of births reported by women;
- The estimates of the numbers of births produced by reverse survival of the populations from previous censuses/surveys;
- The level of registration of births implied by comparison of deaths on the population register (for which births need to be registered) to all registered deaths[4] (which include deaths of births which have not been registered). This level of registration can then be applied to the number of registered births to provide an estimate of all births in each year;
- The number of births in state facilities captured by the Department of Health's District Health Information System (DHIS) plus estimates of the number of births outside these facilities (private facilities and home births);
- The number of children enrolled in school, adjusted by estimates of the proportion attending school from the 2011 Census.

Fertility rates

Until recently, the reverse survival method of fertility estimation as described in *Manual X*, or in Bogue (1993), produced only (and at best) estimates of the General Fertility Ratio. As inputs, the method requires the number of births implied each year by the current age distribution of the census, as well as estimates of the number of women in each five year age group at the midpoint of the year defined by the births. The number of women at reproductive ages was estimated by interpolation between the numbers counted at each census. One conceptual issues, however, is that while the births are determined from those born in South Africa, the process of accurately deriving the resident population of women in each age group in the country in each year is more complicated, and the relatively small proportion of women who are immigrants does not merit the assumptions that would be required to derive a more accurate estimate. Taking account of the number of women allows one to consider how much of the change in the estimated numbers of births is due to fertility alone.

More recently, Timæus and Moultrie (2013), have shown how – with the assumption of a (possibly time-varying) shape of the fertility distribution, the calculation of the implied TFR for any given year is an algebraically simple extension of the standard approach to reverse survival estimation. In applying this refined approach to the South African data, we draw on the shape (but not the level) of fertility derived from the application of the relational Gompertz model to the data from the three censuses and the Community Survey. Since the shape of a fertility distribution generally changes only slowly over time, linear interpolation to any year between

---

[4] i.e. deaths for which a Death Notification form has been completed

those four points will provide a passable estimate of the shape of the fertility distribution at any point in time. Applying this approach to the data from the 2011 census therefore permits the estimation of the level of fertility in the country for the fifteen years before the census. The results are plotted in Figure 5.

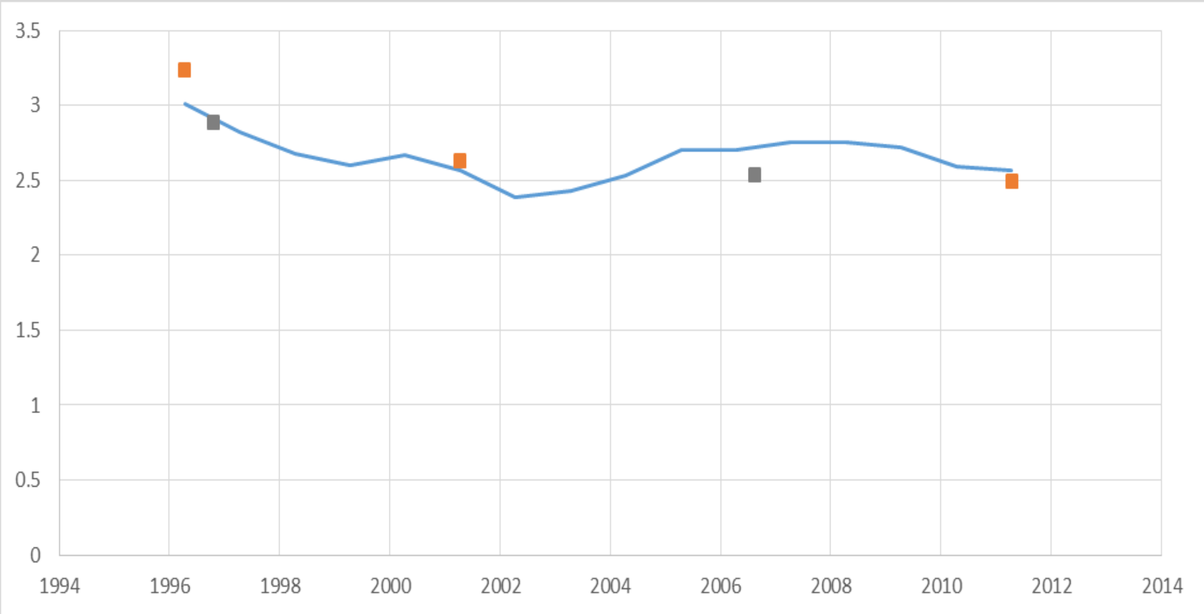

Figure 5 Comparison of estimates of total fertility using reverse survival applied to the 2011 Census age distribution and those derived based on reports of recent births from the 1996, 2001 and 2011 Censuses, the 1998 DHS and the 2007 Community Survey

For three of the five data points (the 1998 DHS, and the 2001 and 2011 censuses) there is a close correspondence between the estimates of fertility derived from data on recent births and the reverse survival estimate of fertility based on the 2011 Census age distribution. This correspondence is all the more remarkable given the relatively poor quality of the recent fertility data in all enquiries other than the 1998 DHS.

The estimates from the 1996 census (too high relative to the reverse survival estimate) and the 2007 Community Survey (too low) fit the pattern less well, although there are no strong grounds with either dataset for asserting that the estimates calculated using the data on recent births are necessarily more robust than those produced by the reverse survival method.

Furthermore, Figure 5 indicates that the TFR implied by the age distribution in the 2011 Census appears to have continued the trend from the late 1960s of gradually declining fertility in South Africa (Moultrie and Timæus 2002) up to 2002, before increasing sharply (by about 15%) from 2002 to 2009, and then possibly starting to fall again.

While the estimates for 2011 and 2001 point to at least some reliability in the data, the heaping in the 2011 Census distributions at age 10 and 11 identified earlier make the correspondence appear somewhat closer than it actually may be after correcting for age-heaping.

Numbers of births produced by reverse survival of the populations from previous censuses/surveys

Figure 6 compares the estimates of the numbers of births in census/survey years (that is, periods of 12 months centred on the date ½, 1 ½, 2 ½ etc. years before the census) implied by the South African-born population counted in each of the censuses (1996-2011) and the Community Survey. With the exception of the 1996 South African Census (which, at the time, was argued to have undercounted, particularly, children under five – even after adjustment on the basis of the PES) and the general undercount in censuses of those under five, there is a fair degree of support for the level and trend over time for the period 1992-2002, the period where the annual number of births was declining rather slowly. Some indication of a rise in the number of births in the period 2002-2006 is evident in the data collected in the 2007 CS.

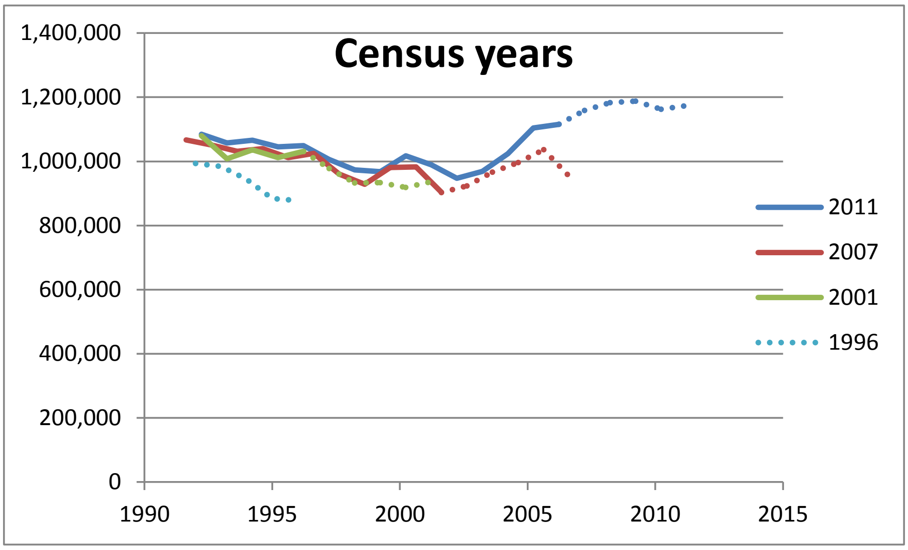

Figure 6 Comparison of number of births estimated using reverse-survival of census counts

Registration implied by comparison of deaths on population register to all registered deaths

A second approach to validating the number of births implied by the 2011 age distribution is to consider data from the vital registration system.

Individuals whose births are registered are added to the Population Register. Deaths for which a death notification form is completed are first processed by the Department of Home Affairs and if they are on the Population Register (i.e. those whose birth has been registered) the death is also recorded on the Register. All death notification forms (including those of people whose birth was not registered) are then passed on to Stats SA who capture details to allow the production of the cause-of-death dataset. Thus comparison of the numbers of deaths in any given year by age, other than those in the first year or two of life where early death may have prevented the birth being registered, from two sources gives an estimate of the completeness of birth registration. Applying these estimates of completeness to the number of registered births to that point provides, for each year of death registration, a series of estimates of the number of births by year which can be compared to those from reverse-survival of the 2011 census count. (A more detailed description of the derivation of these estimates is presented in the Appendix.) This comparison for deaths for the years 2006-2011 is shown in Figure 7.

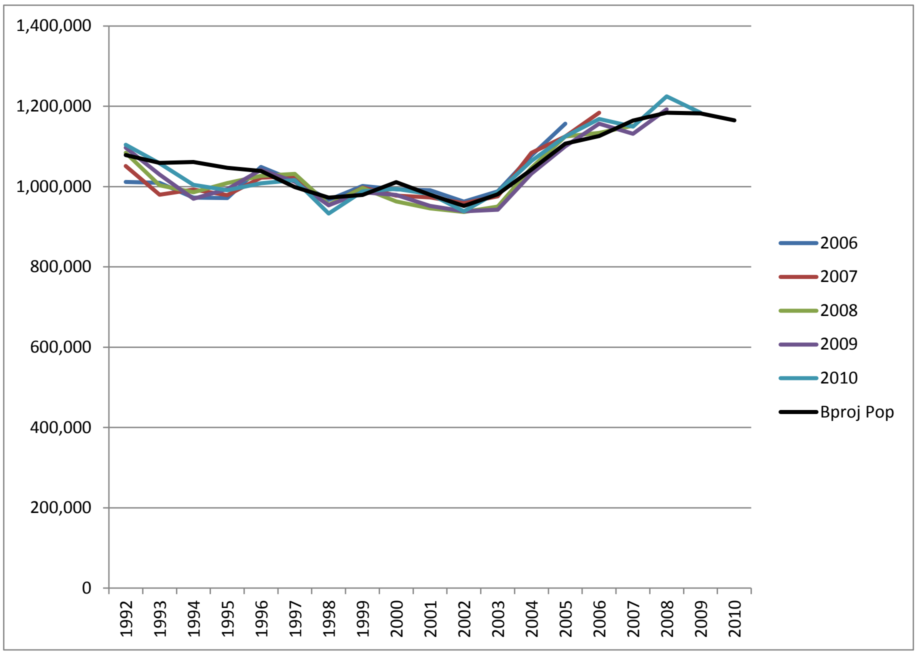

Figure 7 Comparison of the number of births to those estimated from the death registration

Figure 7 shows a remarkably close correspondence between the numbers of births estimated by the two methods. Since not all deaths are registered there is the possibility that the estimate of births derived from the registered deaths could underestimate the true numbers of births to the extent that birth registration is lower for these deaths, however, since, with the exception of the very young, incompleteness of reporting of deaths is relatively low (estimated to be currently about 7% of adult deaths (Dorrington, Bradshaw and Laubscher 2014)) this bias is likely to be small.

The number of births occurring in state facilities, private facilities and 'at home'

Owing to delays in registration of births and deaths, and the frequently encountered problem of particularly poor enumeration of infants in censuses, neither of the above comparisons are useful for confirming the trend in births in most recent years. For this we turn to the record of the number of births recorded by the DHIS as occurring in state facilities since 2004[5]. However, this system does not capture births in private facilities or births occurring at home. The proportion of births occurring in the private sector was estimated as the proportion of under-one year olds recorded as being covered by a medical aid or health insurance in the General Household Surveys from 2004-2006. The proportion of births occurring at home were estimated by linear extrapolation between the estimate of 6.6% from the 2003 DHS and the 4.5% for 2012 from the HSRC survey. Together these suggest that some 16.7% of births occur in private facilities or at home.

Figure 8 compares the total of births in facilities and at home to the births estimated from reverse-survival of the 2011 Census count (interpolated/extrapolated to calendar years). Also included on Figure 8 are the number of deaths registered up to the end of the registration year following the year of death (VR+1), the DHIS deaths and the births estimated from the Census population including foreigners (Census). Once again there appears to be a remarkable consistency, but this time to estimate of births in the most recent years.

[5] The system is known to be incomplete prior to this.

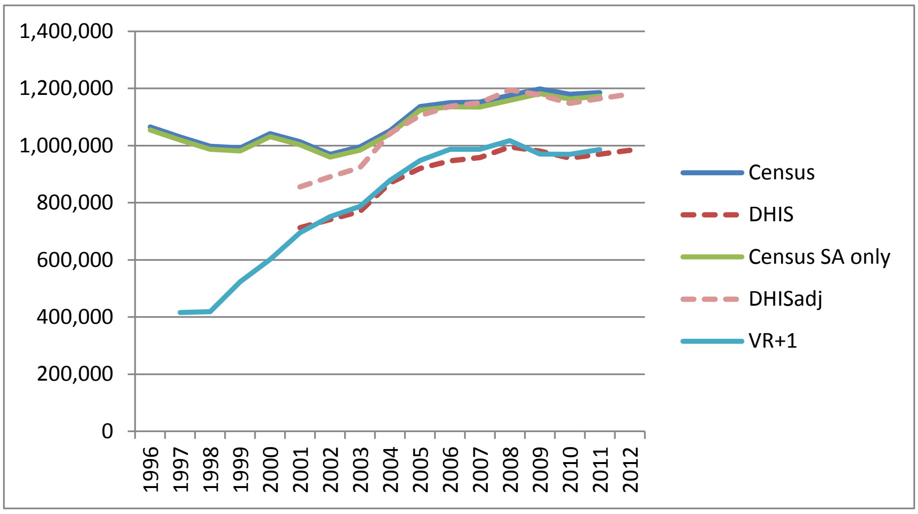

Figure 8 Comparison of the numbers of births with those estimated to have occurred in and out of facilities

The number of children enrolled in school

Figure 9 compares an estimate of the population based on the number of children enrolled in school by age divided by the corresponding proportions of children reported to be attending school according to the census[6]. The comparison appears to support the count, at least from ages 7 to 17 (and hence, at least approximately, the number of births between 2004 and 1994). The drop-off outside this age range is probably the result of respondents reporting, mistakenly, that the person is attending school because of the person's age, or perhaps they were attending an institution not captured by the Department's system.

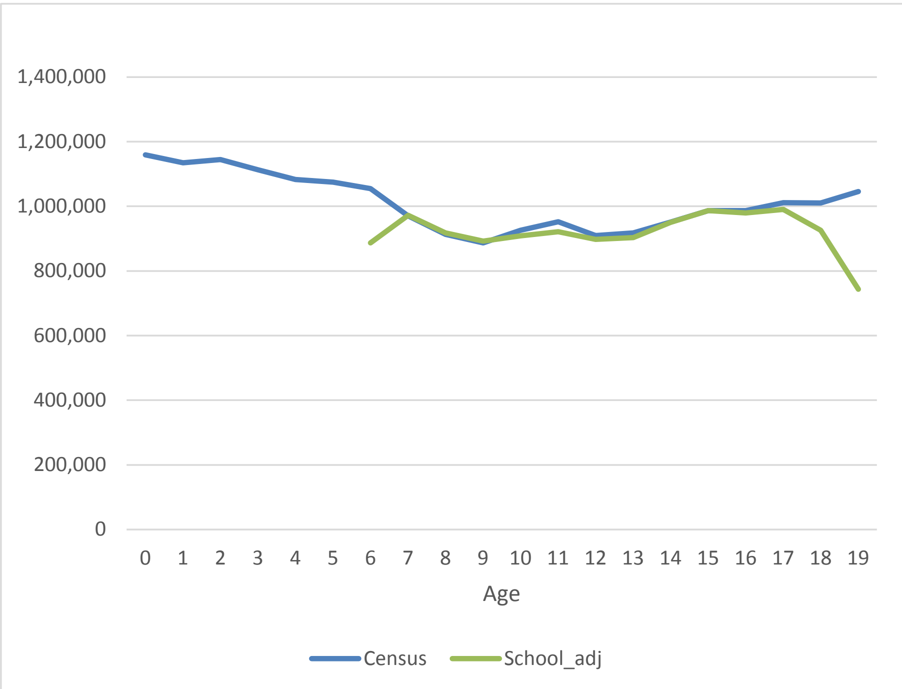

Figure 9 Comparison of the 2011 Census numbers with estimates of the population derived from the numbers enrolled in schools

[6] For this purpose 'did not know', 'unspecified' and missing responses were not attending school. Since the concern with the Census count at this age is of an undercount, this is the most extreme estimate to compare.

Discussion and Conclusions

This paper set out to explore the recent trend in South African fertility using data from the 2011 Census. At the time of the release of the data, a number of demographers, ourselves included, expressed concerns about the accuracy of the census given the implications of the age distribution of the population for the pattern of recent births and fertility in the country. Not only did it imply that the fertility decline had stalled in the country in recent years, but that fertility had risen quite appreciably over a fairly short period of time before again falling, gradually, immediately prior to the Census. The sociological reasons for such a pattern are still not clear. The rise in fertility predates the widespread roll-out of antiretroviral therapy for HIV-infected women, so the rise cannot be attributed to HIV-infected women having children that they might not otherwise have had. Similarly, the rise does not coincide with any sudden increase in the value of the Child Support Grant, or its extension to older children.

Further work is still required to understand and explain these phenomena. For the present, however, the results confirm the decline in fertility until 2002, after which fertility appears to rise and then fall. As we have already seen, this rather odd pattern is also borne out in a number of other datasets and investigations into the number of births, and thus there is a strong probability that the outcome is, indeed real.

Although the results appear to confirm the rather odd trend in fertility implied by the census count, perhaps with some minor adjustments at the national level, this is certainly not the case for all of the population group and provincial populations. It is not the purpose of this paper to analyse and interpret the results by population group or by province. This work is ongoing and will appear in the forthcoming monograph, but suffice it to say that, as expected, there is strong consistency between the TFRs for the African population.

There is also consistency for the Coloured population, but not for the Indian population (which might be due in part to small numbers used to produce point estimates of total fertility) and the White population (where the reverse-survival estimates of fertility are much lower than those derived from recent births as reported in the 2001 and 1996 censuses, possibly due, but only to a limited extent, to emigration of some of the children born in this period).

As far as the provinces are concerned the correspondence between census and school enrolment data is not as good as it is nationally with three provinces being noticeably different (North West suggesting the census count is 5% too high, Mpumalanga that the count is 9% too low and Limpopo that the count is 6% too high). The TFRs match reasonably well for five provinces (Western Cape, Free State, kwaZulu-Natal, North West and Mpumalanga) and of those that don't match there is support for the reverse-survival estimates of recent births in Limpopo from the DHIS data, births from 1995 to 2005 in Gauteng from the school enrolment, and approximate support for recent births in Northern Cape from the DHIS data. Interestingly in the case of the other instances of uncertainty in comparison of the TFRs the school enrolment data suggests either that the census is correct, or where not, that the differences between the TFRs should be even greater!

It is unclear whether these inconsistencies at a subpopulation level reflect inconsistencies in the subpopulation age distributions in the census, perhaps brought about by deficiencies in the post-PES adjustments for undercounts, or through errors in the estimates of recent fertility and other data sources.

Two other, less obvious, conclusions emanate from the results presented. First, the reverse-survival (if correct) implies that after a rapid increase in completeness of birth registrations within a year after birth up to 2004, completeness has increased only slightly since then. The second is that the problems in enumerating children under the age of 5 experienced in the 1996 and 2001 Censuses would appear to have been avoided if the census age distribution is accepted. However, by accepting the census age distribution, a further implication is that the count of children under the age of 5 in the 1996 Census was particularly severe.

The study suffers from some limitations, most of which are not in our control. It would, for example, have been preferable if Statistics South Africa had followed its previous practice of making available edit flags as well as the rules used to edit the data with the 10% sample, as well as a dataset containing (a sample of) the raw data, the edit flags, and the edited values to bona fide researchers. This would have allowed researchers to assess the extent to which the patterns being observed are influenced by the editing and imputation applied to the data.

Appendix: Estimating the total number of births from the number of births registered and an estimate of the percentage of the population on the population register (derived from the proportion of registered deaths which appear on the population register).

Once a birth is registered the person is added to the Population Register (PR). For the purpose of this exposition, we define a death as being registered if a death notification form has been completed in respect of that death. These death notification forms are the responsibility of the Department of Home Affairs. The population register records the deaths (the office where it was registered and whether it was due to 'external' or natural causes) of all registered deaths of people on the population register (i.e. whose birth has been previously registered). The MRC receives monthly updates of the deaths recorded on the population register (PR) and input for its Rapid Mortality Surveillance (RMS) project with UCT.

Independently, Statistics South Africa captures and processes all the death notification forms to provide more detail, particularly as to causes of death. This process usually takes about two years. Stats SA releases unit record data (excluding some fields, e.g. population group, some detail of dates, etc., to ensure that individuals cannot be identified).

Thus for a given calendar year, *y*, we have the following records of deaths age x last birthday at death, $D^{vr}(x,y)$ from Stats SA's record of vital registration (VR) of deaths for which a death notification form was completed and $D^{pr}(x,y)$ from the Rapid Mortality Surveillance project (being deaths of people for whom a death notification form was completed and they are on the population register). Thus all deaths in $D^{pr}(x,y)$ should be included in $D^{vr}(x,y)$ and the ratio of $D^{pr}(x,y)$ to $D^{vr}(x,y)$ gives a measure of completeness of births of those who have died aged *x*.

To be more precise a proportion, *p*, of deaths in calendar year *y* aged *x* last birthday at death were born in calendar year *y-x* and the remaining proportion, 1-*p*, born in calendar year *y-x*-1. Assume that for $x > 0$, *p* is approximately 0.5. Thus $R(x,y) = \frac{\left(0.5D^{pr}(x,y) + 0.5D^{pr}(x,y+1)\right)}{\left(0.5D^{vr}(x,y) + 0.5D^{vr}(x,y+1)\right)}$ for $x > 0$ gives a measure of the level of registration of births (of registered deaths) in calendar year *y-x* including late registration up to and including year *y*.

It is not possible to use the deaths in the first year of life to estimate the completeness of birth registration because many of these deaths (nearly half?) occur before the birth is registered. Although, to some extent this bias also exists for deaths aged 1 last birthday, the impact is far lower, particularly in recent years, given that it appears as if less than 15% of deaths are registered in the second year of life and since 2005 over 70% of births are registered in the year of birth.

The completeness of birth registration of registered deaths is probably higher than the completeness of all births because:

- It is likely that the completeness of registration of births of deaths that are not registered is, if anything, lower than that of registered deaths (if, for example, they both occur in areas with poor registration).
- Unregistered births that die will never be registered.

In the case of the second of these points, the bulk of this occurs in the first year of life. We can get an estimate of the cumulative extent of this up to a specific age for the registered deaths by multiplying the proportion of the births of registered deaths who weren't registered in year *y*, $(1 - R(x,y))$, by the mortality rate at age *x* in year *y*. Summing these proportions over the cohort up to a given age gives the cumulative proportion of births never registered because of prior death up to a given age ($p^{bnr}(x,y)$). For 2010, for example, this is 2% for deaths under age 1 rising to 3.3% for deaths between the ages 18 and 19 (mainly because the mortality rate under age 1 of this cohort was 40% higher than in 2010).

The first point is a little more difficult to deal with. The most extreme assumption (producing the lowest estimate of completeness of registration) is to assume that none of the births of the deaths that were not registered, were registered. In this case the completeness of birth registration of all deaths is estimated simply by multiplying the completeness of registration of births of registered deaths computed above ($R(x,y)$) by the completeness of reporting of the deaths, $p^{dr}(x,y)$, (which has been estimated elsewhere (Darikwa and Dorrington (2011), HDACC Report (2012), RMS Report (2012) and most recently for the South African Burden of Disease work for the MRC) and multiplying this by multiplied by 1 minus the proportion of births never registered because they died before the birth could be registered (i.e. $1-p^{bnr}(x,y)$).

However, this method produces implausibly low estimates of ultimate completeness of birth registration for deaths at the young ages (because it is too extreme to assume that none of the births of deaths which were not registered, were registered, when death registration is as low as it is at the young ages, or as uncertain as it is for those aged 5 – 14). A better estimate of the lower limit of ultimate completeness of birth registration is probably to assume that few births are registered more than 19 years after birth. Thus, since the completeness of registration of deaths above age 15 is about 93%, this suggests that a lower limit of the completeness of deaths would be $0.93R(x,y)(1-p^{bnr}(x,y))$.

References

Bogue, DJ. 1993. Fertility measurement by reverse survival of age-sex distributions, in D Bogue, J., EE Arriaga, and D Anderton, L. (eds), *Readings in Population Research Methodology*, Volume 3. New York: United Nations Fund for Population Activities, pp. 11/39 - 11/42. doi:

Booth, H. 1984. Transforming Gompertz' function for fertility analysis: The development of a standard for the relational Gompertz function, *Population Studies* 38(3): 495-506. doi: http://dx.doi.org/10.2307/2174137

Caldwell, JC and P Caldwell. 1993. The South African fertility decline, *Population and Development Review* 19(2): 225-262. doi: http://dx.doi.org/10.2307/2938436

Coale, AJ. 1964. How a population ages or grows younger, in R Freedman (ed.) *Population - The Vital Revolution*. Garden City NY: Anchor Books, pp. 47-58. doi:

Dorrington, RE. 2013. *Alternative South African mid-year estimates, 2013*. Centre for Actuarial Research Monograph 13. Cape Town: University of Cape Town. Available: www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/mono13.pdf

Dorrington, RE, D Bradshaw, and R Laubscher. 2014. *Rapid Mortality Surveillance Report 2013*. Cape Town: South African Medical Research Council. Available: http://www.mrc.ac.za/bod/RapidMortality2011.pdf

el-Badry, MA. 1961. Failure of enumerators to make entries of zero: errors in recording childless cases in population censuses, *Journal of the American Statistical Association* 56(296): 909-924. doi: http://dx.doi.org/10.1080/01621459.1961.10482134

Feeney, G. 1998. *A New Interpretation of Brass' P/F Ratio Method Applicable when Fertility is Declining*. Available http://www.gfeeney.com/notes/pfnote/pfnote.htm. Accessed 22 June 2008.

Mostert, WP and JM Lötter (eds.). 1990. *South Africa's Demographic Future*. Pretoria: Human Sciences Research Council.

Moultrie, TA. 2013a. The el-Badry correction, in TA Moultrie, RE Dorrington, AG Hill, K Hill, IM Timæus, and B Zaba (eds), *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population. http://demographicestimation.iussp.org/content/el-badry-correction. Accessed 26/09/2014.

Moultrie, TA. 2013b. The relational Gompertz model, in TA Moultrie, RE Dorrington, AG Hill, K Hill, IM Timæus, and B Zaba (eds), *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population. http://demographicestimation.iussp.org/content/relational-gompertz-model. Accessed 26/09/2014.

Moultrie, TA and R Dorrington. 2004. *Estimation of fertility from the 2001 South Africa Census data*. Cape Town: Centre for Actuarial Research for Statistics South Africa.

Moultrie, TA, RE Dorrington, AG Hill, K Hill, IM Timæus, and B Zaba (eds.). 2013. *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population. Available at demographicestimation.iussp.org.

Moultrie, TA and IM Timæus. 2002. *Trends in South African Fertility between 1970 and 1998: An Analysis of the 1996 Census and the 1998 Demographic and Health Survey*. Cape Town: Medical Research Council. Available http://www.mrc.ac.za/bod/trends.pdf. Accessed 1 May 2011.

Moultrie, TA and IM Timæus. 2003. The South African fertility decline: Evidence from two censuses and a Demographic and Health Survey, *Population Studies* 57(3): 265-283. doi: http://dx.doi.org/10.1080/0032472032000137808

Statistics South Africa. 2013a. *Mid-year population estimates, 2013*. Report P0302. Pretoria: Statistics South Africa. Available: http://beta2.statssa.gov.za/publications/P0302/P03022013.pdf

Statistics South Africa. 2013b. *Recorded live births, 2012*. Report P0305. Pretoria: Statistics South Africa. Available: http://www.statssa.gov.za/publications/P0305/P03052012.pdf

Statistics South Africa. 2014. *Mid-year population estimates, 2014*. Report P0302. Pretoria: Statistics South Africa. Available: http://beta2.statssa.gov.za/publications/P0302/P03022014.pdf

Statistics South Africa. 2015. *Mid-year population estimates, 2015*. Report P0302. Pretoria: Statistics South Africa. Available: http://www.statssa.gov.za/publications/P0302/P03022015.pdf

Timæus, IM and TA Moultrie. 2013. Estimation of fertility by reverse survival, in TA Moultrie, RE Dorrington, AG Hill, K Hill, IM Timæus, and B Zaba (eds), *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population. http://demographicestimation.iussp.org/content/introduction-fertility-analysis. Accessed 26/09/2014.

Udjo, EO. 1997. *Fertility and Mortality Trends in South Africa: The Evidence from the 1995 October Household Survey, and Implications on Population Projections*. Pretoria: Statistics South Africa. Available: https://www.statssa.gov.za/census01/Census96/HTML/Metadata/Docs/fertility_udjo.htm

Udjo, EO. 2005. Fertility levels, differentials and trends, in T Zuberi, A Sibanda, and EO Udjo (eds), *The Demography of South Africa*. Armonk NY: M.E. Sharpe, pp**.** 40-64. doi:

UN Population Division. 1983. *Manual X: Indirect Techniques for Demographic Estimation*, 81. New York: United Nations, Department of Economic and Social Affairs, ST/ESA/SER.A/81. Available http://www.un.org/esa/population/techcoop/DemEst/manual10/manual10.html

Zaba, B. 1981. *Use of the Relational Gompertz Model in Analysing Fertility Data Collected in Retrospective Surveys*. Centre for Population Studies Research Paper 81-2. London: Centre for Population Studies, London School of Hygiene & Tropical Medicine.