Quality and level of birth registration in South Africa: 1998-2008

Nhlapo Mosidi
Statistics South Africa
170 Andries Street
Pretoria, 0001, South Africa
E-mail: mosidin@statssa.gov.za

Introduction

This study assesses the feasibility of using birth registration data for direct estimation of fertility schedules. Firstly the paper looks at issues of data quality, including trends in current and late registrations, missing data and sex ratios. Estimates of the completeness of birth registration are also provided. Lastly comparative estimates of fertility schedules derived from the register and enumerated data are shown.

Data sources and methodology

This analysis is based on data from the Department of Home Affairs (DHA) on registered births from 1998 to 2008. Censuses 1996, 2001 and Community Survey (CS) 2007 are used to derive comparative fertility indicators. (CS is a large national survey including 246 618 households, the survey included questions on births in the last twelve months which are usually included in the Census).

Issues of data quality

Late registrations, missing data and parental age

Current registration has been improving in South Africa, it was 273 180 in 1998 and increased to 915 674 in 2008. The trends in late birth registrations between 1998 and 2008 show a decline in the number of late registrations, with a fair amount of fluctuation between the years. In all the years analysed except for 2008, birth registered for foreign nationals related to births that occurred about thirty to forty years ago.

For the first time in 2005 the number of current registrations was higher than the number of late registrations. This trend has persisted in subsequent years.

The age of father is one of the most poorly captured variables in birth registration data. About 45% of the births registered in 1998 did not contain information on age of the father, this figure rose to 67 percent in 2008. Paget and Timaeus (1994) note that this might be an indication of high proportions of illegitimate births. The unavailability of the marital status variable in the dataset makes it impossible to confirm this case. The mother age variable has fewer missing values with less than one per cent missing data for all the years.

In addition to missing parental age indicated above, data shows an extensive number of births that occurs to women outside the reproductive ages of 12 to 49. This is notwithstanding the fact that births do occur to women above the age of 50. Age of the mother is a critical variable in demographic analysis. In fertility analysis maternal age has implication for the interpretation of trends and level of fertility.

The data show occurrence of births to children in the age group 1–11 as well as some births to women 50 years and above. The number of births occurring to children aged 1 to 11 years was 648 at the base year with

the highest number recorded in 2004 at 1190, though there seems to be no trend over time. There has been a decline in the number of births to women above 50 years from around 24024 in 2001, to 4 275 in 2008. It is highly likely that birth occurrences to children 1-11 might be the result of capturing errors at DHA.

Estimating birth register completeness

Two approaches are adopted to estimate completeness of birth registration. In the first instance Stat SA estimates of births are used to derive completeness levels. The second approach uses enumerated data from Census 2001 and Community Survey 2007 as denominators to estimate completeness. In addition, in an attempt to see which dataset provide plausible completeness level, three variations of register data are used. The first is data updated for late registration using all late registration for each birth year 1998 to 2009 regardless of the time lag between occurrence and registration. The second data is updated using only those births registered within the first five years of birth /occurrence (This data is referred to as updated data in the rest of the paper). The third data is only births registered within occurrence/birth year (In the rest of the paper this is referred to as current registration birth data). Completeness is also computed using the latter two datasets based on enumeration points

Figure 1. Comparative estimates of completeness level for birth registration,1998- 2008

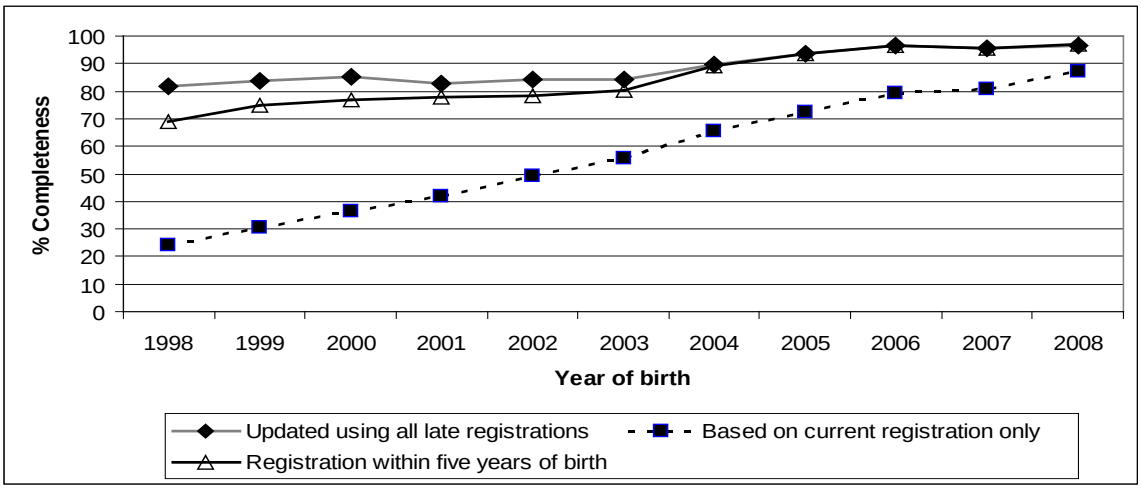
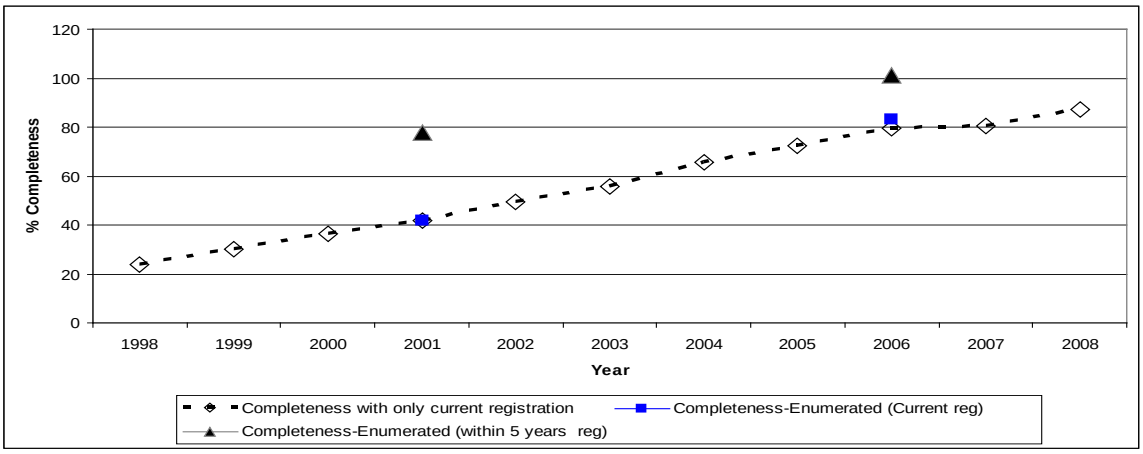


Figure 1 compares completeness estimates obtained using the three different variants of birth registration data, where the denominator is Stats SA official birth estimates. Births registered within five years of births could only be obtained till 2004 after this period the first two methods yield the same estimates of completeness. Data adjusted using all late registrations shows that completeness increases from 77% in 1998 to an estimated 96% in 2008. These estimates however, appear improbably high and bring into question the effects of updated data, given that the dataset is adjusted for late registration regardless of how far off registration occurs to the birth year. Completeness derived from updated data is slightly lower that the latter estimate though it can be seen that by 2004 the estimates are the same. Declining late registrations over time also influence the trends observed in Figure 1. Using current birth registrations increases completeness levels from a low of 24% in 1998 to 87% in 2008. These estimates seem much more plausible than the case when updated data are used and present a much clearer trend over time. It is significant that by 2008 the completeness estimates from the two approaches are moving closer to each other. This is attributable to improvements in the number of births that are registered within occurrence year.

Figure 2 present the result of the analysis based on second approach. The trend line is based on current registration and similar to the previous section. The square points represent 2001 Census and 2006 based on CS (2007), using only current registration data. The triangular points are based on updated data (The 2006 point only represents data updated for births that were registered between 2006 and 2009, only three years after occurrence). Of interest in figure 2 is that completeness based on current registration only is identical to the computation based on enumerated data, indicating that both Census 2001 births and the CS 2007 births were adjusted using current birth registration data. Completeness level were around 24% in 2001 due to high levels of under-registration of births, this means that the adjustment factors used were much larger than in subsequent years. Larger adjustment factors are usually associated with higher uncertainty levels. When completeness is estimated using updated data, the estimates are higher than those obtained using only current registrations for both 2001 and 2006. In 2006 the number of births reported in the register was higher than the number obtained using CS. This approach thus yields completeness levels above 100% in 2006.

Figure 2 Comparative estimates of completeness level based on enumerated data, 2001 and 2007



Further analysis of the late registration data was done by cumulating births from late registrations for each year after birth year over time. Completeness levels were then computed for each year from additional registrations. Table 1 shows the percentage contribution of each year of late registrations to completeness levels for each year 1998 to 2006. It shows two things; firstly that for each year the contribution of late registrations to completeness decline over time. Secondly that the contributions made by births registered one year after birth has bee declining between 1998 and 2006. For births that occurred in 1998, registering one year after occurrence contributed about 10% more to completeness levels, by year nine contributions of late registration resulted in just one percent difference in completeness levels. By 2006 births registered one year after birth contributed only about 4% to completeness levels. This suggests that in subsequent years late registration will make minimal or no contribution to completeness levels.

Table 2. Percentage contribution of late registration to completeness levels for each year of registration

	Year One	Year two	Year three	Year four	Year five	Year six	Year seven	Year eight	Year nine
1998	0.10	0.08	0.08	0.06	0.04	0.02	0.02	0.01	0.01
1999	0.10	0.09	0.06	0.04	0.02	0.02	0.01	0.01	
2000	0.10	0.07	0.04	0.02	0.02	0.01	0.01		

2001	0.08	0.04	0.02	0.02	0.01	0.01
2002	0.06	0.02	0.02	0.01	0.01	
2003	0.04	0.03	0.02	0.01		
2004	0.05	0.02	0.02			
2005	0.04	0.02				
2006	0.04					

Estimating Fertility Rates

This section looks at comparative estimates of fertility derived from the birth register and enumeration. The t hree points in Figure 3 shows estimates of TFR from enumeration, while the trend line is derived fro m the register. For each year the register data is adjusted for missing data and for births occurring to women outside the 12-49 age range, these are then redistributed proportionally across the seven age gro ups based on the observed age pattern of fertility.

The register estimated 3.2 in 1998 which is similar to the Census 1996 estimate. The 2001 census estimates is the closest of the three periods, where the register estimated TFR of 3 and Census had an estimate of 2.9. Given the observation from the previous section about the high uncertainty around the adjustment factors in earlier years it is highly unlikely that 1998 and 2001 estimates from the register are believable. The 2006 estimate is much closer to an estimate by Stats SA (2010) of 2.6 unadjusted and an adjusted 2.8 based on the CS (2007).

Figure 3 Comparison of Total fertility rates estimated from current registration and enumerated

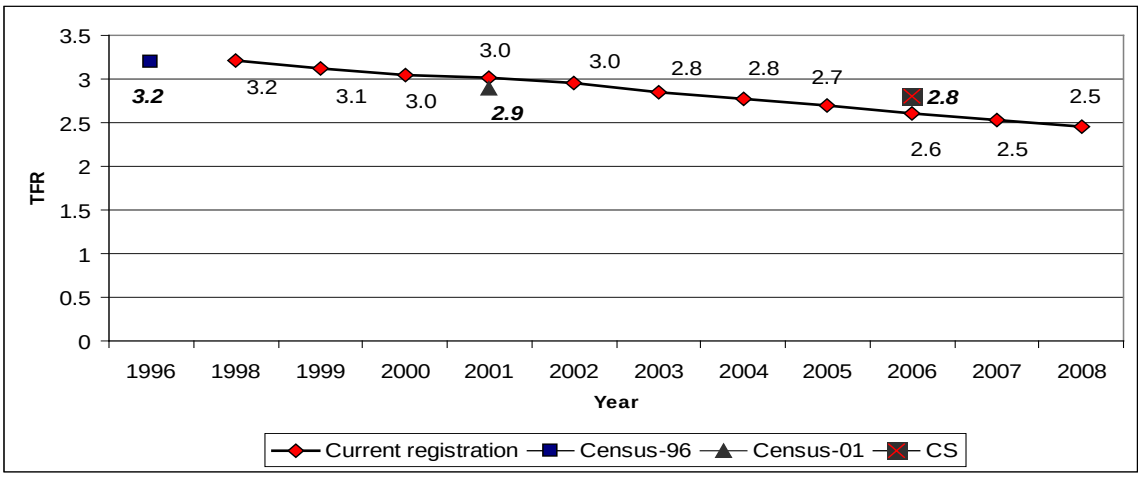


Figure 4 and 5 compares the age pattern of fertility derived from the Census 2001, CS 2007 with both updated and current registration data corresponding to the same years.

Figure 4 Age specific fertility rates from updated registration data and enumerated data

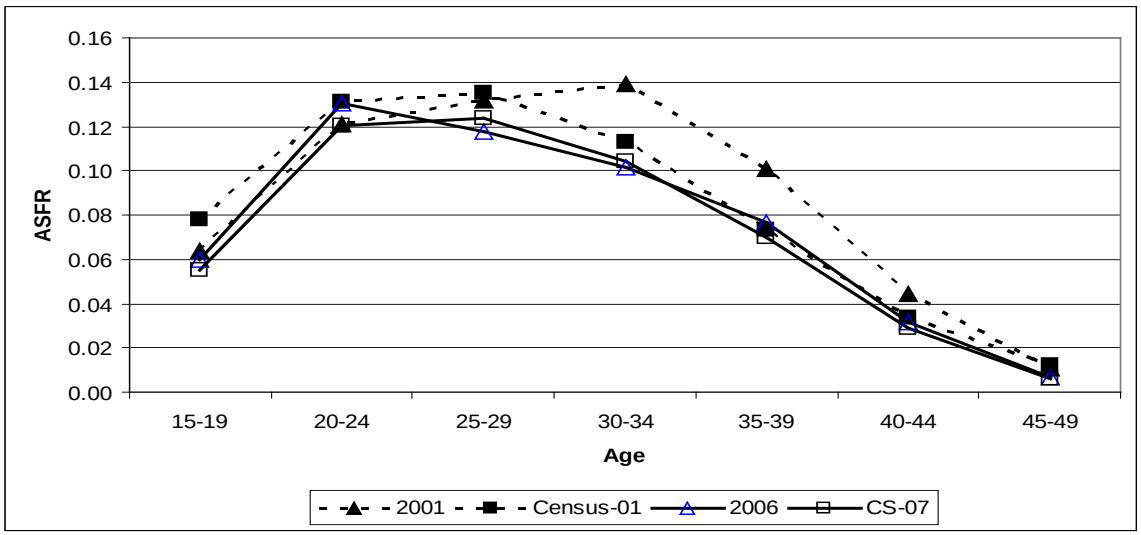
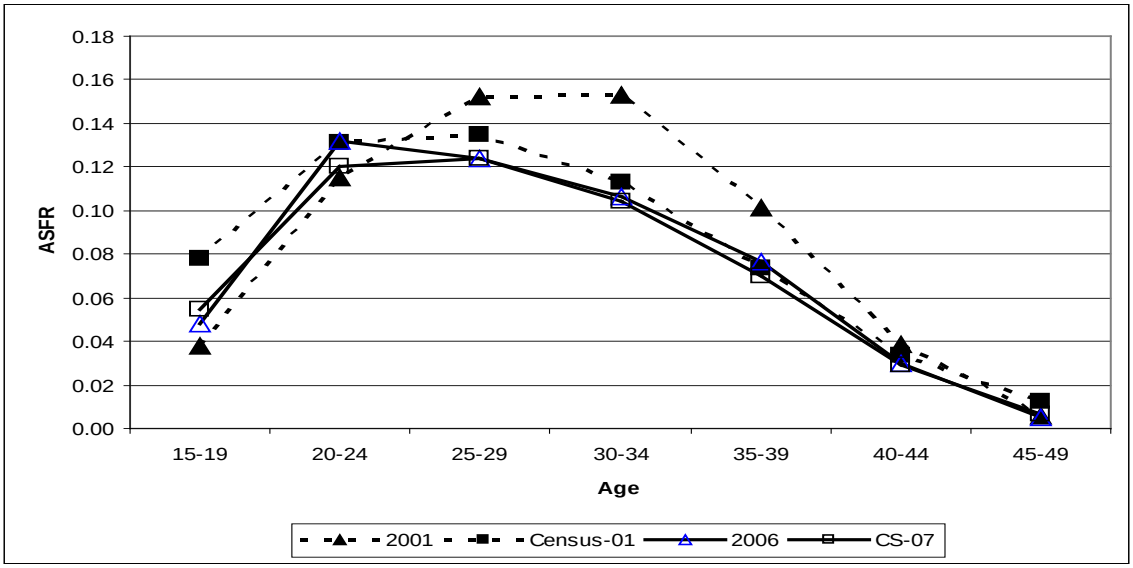


Figure 5 Comparative estimates of age specific fertility rates from current registration data and CS data



In 2001 for both updated and current data fertility peaks at ages 30-34 while census data indicate fertility peak at ages 25-29. Both the ASFRs based on updated data and current data show that in 2001 births by women below age 29 and 25 respectively were grossly under-registered. Above age 29 for updated data and 25 years for current data the schedule from the register reflect more births than census. By 2006 there was much agreement between the two datasets as they show fertility peak at ages 20-24, while the CS indicate fertility peak at ages 25-29, in line with census.

Quite importantly by 2006 the under-registration observed in 2001 at ages below the peak had declined extensively, and discrepancies observed in ages above the peak had also reduced somewhat. Although there are some slight variations between the estimates produced from the two data sources, the patterns and level by age appear much closer in 2006 than was the case in 2001. The discrepancy might be attributed to age misreporting in the register and/ or enumerated data, a point which require further research.

Discussion and conclusions

The results of this study suggest improvements in some aspects of birth registration data in South Africa. The number of births registered within the year of birth increased from 273 180 in 1998 to 915 674 in 2008, an increase of 70%. However the data still presents with a number of challenges, mainly related to quality. Three major types of errors have been identified in this analysis. Firstly, the continued presence of late birth registrations. The second major error is the number of missing data that is associated with age. The age of the father has the highest percentage of missing cases of the two parental age variables. Lastly is the incorrect reporting of parental age particularly the mother's age, which is an important variable used in the computation of most demographic indices. Births occurring to women outside the 12-49 age range might be the result of data capturing errors at DHA.

Various approaches used in this study to estimate completeness levels indicate that using updated data produces improbably high estimates of completeness. The analysis also revealed that over time the contribution of late registration to completeness levels become insignificant.

Fertility schedules derived from completeness adjusted data compares favourably with enumerated data, but large adjustment factors in earlier years attracts some uncertainty about fertility estimates derived from the register in earlier years. Both updated and current registration data yield the same estimate of 2.6 Total Fertility Rate in 2006. This is largely due to increasing current registrations over time. Comparisons of age pattern of fertility from the two data sources in 2006 indicate that after adjustment for completeness the register data could be used to provide reasonable evaluation of **pattern** of fertility derived from enumerated data in the near future. However, ASFR for 2001 provide indications of under-registration of births at younger ages and possibly age misreporting.

It is also unclear at which stage of the registration process are some of the noted deficiencies introduced in the data. This could be either during data collection stages or at processing stages. The findings in this study also signal that challenges facing DHA have shifted from coverage to data quality.

References

Paget J. and Timaeus I. M. 1994. A Relational Gompertz Model of Male fertility: Development and Assessment. Population Studies, 48(2) pg 330-340

Stats SA. 2007. The coverage and quality of birth registration data in South Africa,
1998-2005. Report No. 06-03-01. Pretoria. Statistics South Africa.

Stats SA. 2009. *Recorded live births 2009* Statistical Release, P0305. Pretoria. South Africa.