INFANT AND UNDER-FIVE MORTALITY IN SOUTH AFRICA: PERSPECTIVES FROM THE 2011 CENSUS AND THE 2012 HSRC SURVEY.

BOBOH KAMANGIRA (Mr)

CENTRE FOR ACTUARIAL RESEARCH (CARe)
UNIVERSITY OF CAPE TOWN

Thesis submitted for the degree of Master of Philosophy in Demography to the Faculty of Commerce.

December 2014

PLAGIARISM DECLARATION

This research is my original work, produced with supervisory assistance from my supervisor. I have used the Harvard convention for citation and referencing. Each contribution to this dissertation from the works of other people has been acknowledged, cited and referenced. In addition, this dissertation has not been submitted for any academic or examination purposes to any other university.

Signature: Date: 04/12/2014

ABSTRACT

This research focuses on estimating infant and under-five mortality in South Africa for the period 1998-2012, both to update previous estimates taking into account new data and to assess the reasonableness of all estimates. Data from the 2011 Census and the 2012 HSRC survey were used for this purpose. The 2011 Census provided data from deaths reported by households as well as the survival of the most recent births. The 2012 HSRC provided full birth history data for women aged 15-49 which were used for direct estimation of childhood mortality.

Deaths reported by households together with census estimates of the number of children under the age of five are used to produce estimates of infant and under-five mortality using the synthetic cohort life table approach. Blacker and Brass's previous birth technique is used to provide an estimate of infant mortality based on the survival of the most recent birth in the 24 months preceding the census after correcting for the bias in the proportion dead among most recent births relative to the proportion dead among all births. The under-five mortality rate corresponding to this infant mortality rate is estimated using the ratio of under-five mortality to infant mortality as observed in the Princeton West level 19 model life tables. The direct method for estimating childhood mortality is applied to the 2012 HSRC full birth history data. After imputing the exact dates of birth and death, locating deaths in time and calculating the exposure to risk of dying, the method of deriving period life tables is then followed to estimate $q(1)$ and $q(5)$.

Results show that the correction for the bias in the proportion dead among the most recent births relative to the proportion dead among all births in the Previous Birth Technique estimates did not work as estimates were still too low relative to those produced by other researchers. Also, the direct estimates from the 2012 HSRC survey were lower than those produced by other researchers owing to too few deaths recorded in this survey and a large proportion of these missing ages at death. Only the estimates from deaths reported by households were found to be useful. Thus, it was concluded that the data from the 2012 HSRC survey and the survival of most recent births from the 2011 Census do not produce reliable estimates of childhood mortality.

It is recommended that a DHS-type survey should be conducted with one of its purposes being to investigate the issues around childhood mortality estimation in South Africa, particularly to improve on the quality of data available for estimation of childhood mortality and to investigate further why the current methods are failing to produce

reasonable estimates of childhood mortality. Further research to investigate the extent of the bias in the proportion of children dead among most recent births relative to the proportion dead among all births is also needed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Rob E. Dorrington, for without his counsel and guidance I wouldn't have made it this far. A lot of challenges were encountered in this research which made me to think of even quitting but he continued to give me hope and thus encouraging me to solder on.

Also, I would like to extend my gratitude to Statistics South Africa, in particular to Diego Iturralde who provided me with the tabulations of deaths reported by households before the 10 per cent sample was released.

I would like to appreciate the Hewlett Foundation for providing me with the scholarship that enabled me to pursue and complete my Master of Philosophy degree.

Lastly, special thanks goes to my parents, siblings and colleagues (Malvern Munjoma and Clara Lemani) whose support and encouragement made me to believe in myself.

Boboh Kamangira

Cape Town, December 2014

1 INTRODUCTION

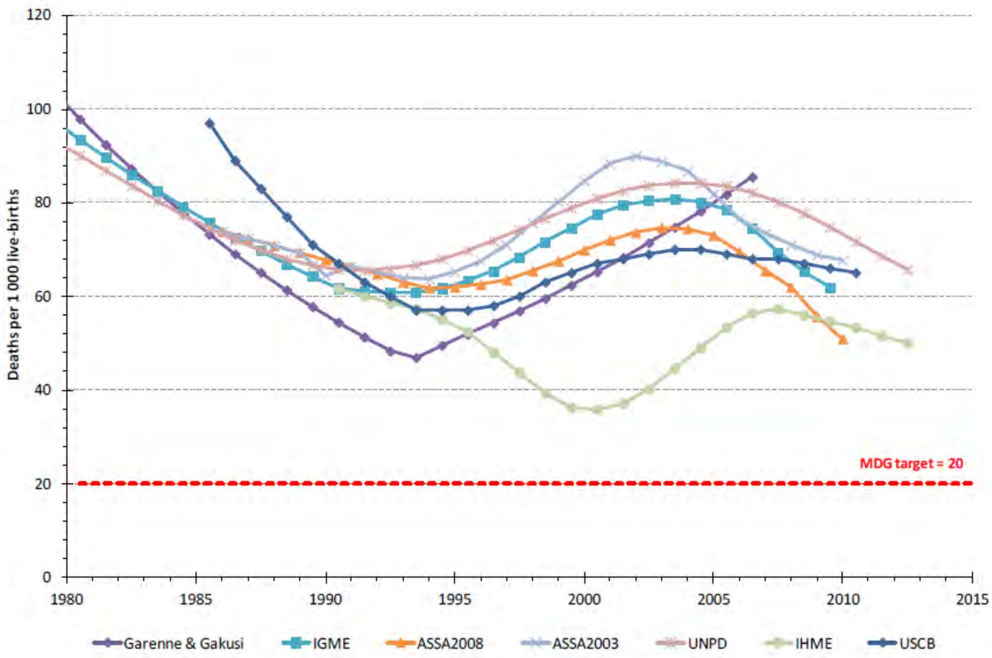
1.1 BACKGROUND AND CONTEXT

Infant and under-five mortality rates are important measures used in the assessment of the progress towards achieving the United Nations Millennium Development Goal number 4 (MDG 4) (Nannan, Timæus, Laubscher *et al.*, 2007) which is to reduce child mortality by two thirds in 2015 from the 1990 level (United Nations, 2000). As such, estimation and constant updating of these measures is needed in every country that is striving to meet this goal. Care should be taken when estimating infant and under-five mortality as data in Sub-Saharan Africa are defective and deficient - vital registration is incomplete and surveys and censuses often produce data with errors. Application of indirect methods to data from countries with high HIV prevalence in other studies failed to produce estimates of child mortality that are robust to the HIV bias even after attempting to correct for the bias resulting in some researchers concluding that these countries should not use indirect methods to estimate child mortality (Mahy, 2003; Ward and Zaba, 2008). As such, estimates produced using indirect methods need to be reconciled with estimates produced from direct methods using full birth history data.

Considerable work on childhood mortality trends and estimates in South Africa has been produced by different researchers (Bradshaw, Dorrington and Laubscher, 2012; Darikwa, 2009; Dorrington, Moultrie and Timæus, 2004; Dorrington, Bradshaw and Laubscher, 2014; Garenne and Gakusi, 2005; Mazur, 1995; Nannan, Dorrington, Laubscher *et al.*, 2012; Nannan, Timæus, Laubscher *et al.*, 2007; Wang, Dwyer-Lindgren, Lofgren *et al.*, 2012; Wang, Liddell, Coates *et al.*, 2014). The estimates produced have made use of data from various sources, i.e. censuses, Demographic and Health Surveys, vital registration, and demographic surveillance and health service data. Independent groups such as the United Nations Inter-agency Group on childhood Mortality Estimates (UN-IGME), the Institute for Health Metrics and Evaluation (IHME) and the Actuarial Society of South Africa (ASSA) have also produced estimates of infant and under-five mortality over time based on models. These scholars and groups generally agree on the trend in childhood mortality before the start of the HIV epidemic, i.e. that child mortality declined rapidly until around 1995, but differ as to the level and trajectory of estimates over time since then. Analysis of the 1996 Census and 1998 DHS shows that estimates of the trends in childhood mortality from these two sources are consistent with each other after adjusting for misreporting of still-births as live-births in the census while the estimates of the levels of childhood mortality are different (Nannan, Timæus, Laubscher *et al.*, 2007) thus giving confidence in the estimates of trends in childhood mortality up to 1996.

For the period after 1996, characterised by high HIV prevalence, estimates of the levels of childhood mortality seem to vary with each researcher and group although they all agree that the trend of declining mortality reversed around 1996. Figure 1.1 illustrates the extent of variation in the estimates of childhood mortality levels and trends in South Africa by showing estimates of the levels and trend in U5MRs produced by various models as extracted from Nannan, Dorrington, Laubscher *et al.* (2012). For the period before 1990, most of the models are consistent with each other with regards to the levels and trend of U5MRs, with the exception of the United States Census Bureau estimates of the levels of U5MRs which are higher than those of others (Figure 1.1). Even allowing for the fact that IHME reworked their estimates (not shown in Figure 1.1) to be more consistent with others (Wang, Dwyer-Lindgren, Lofgren *et al.*, 2012) much of the uncertainty in these model estimates can be seen in the period after the onset of the HIV/AIDS epidemic, particularly after the year 1996. Differences in both the level and trajectory of the trends are much too significant to be ignored as trivial.

Figure 1.1 Model estimates of under-five mortality rates compared with the MDG 4 target for South Africa



Source: Figure 6.1 in Nannan, Dorrington, Laubscher, et al (2012).

Estimation of childhood mortality in South Africa has depended on models in the recent past owing to the lack of good quality data on children's survival which can be used for estimating childhood mortality. Estimates of childhood mortality produced from the 2007 Community Survey indicate that child mortality may have remained constant in the ten years preceding the survey, after adjusting for HIV/AIDS bias (Darikwa, 2009; Nannan, Dorrington, Laubscher *et al.*, 2012). However, Nannan, Dorrington, Laubscher *et al.* (2012) argue that these estimates cannot be fully relied upon as there are some quality issues with the data – about eight per cent of the summary birth history data was excluded from the analysis because it was internally inconsistent. Also, the CEB/CS method tends to 'smooth' estimates over time and thus is not ideal for estimating rates over a period when they are expected to have peaked and fallen (Nannan, Dorrington, Laubscher *et al.*, 2012). The survey is also too small to put too much trust in the estimates it produces, even if the estimates appear reasonable.

Prior to the 2007 Community Survey, South Africa had conducted the SADHS in 2003 and a national census in 2001. SADHS 2003 failed to produce reasonable estimates of child mortality due to "problematic fieldwork" and, to a lesser extent, "the bias introduced by the AIDS epidemic" (Nannan, Dorrington, Laubscher *et al.*, 2012:18). Thus, the estimates produced from the 2003 SADHS were inconsistent with estimates from the 1998 SADHS, with the 2003 SADHS estimates being lower than 1998 SADHS estimates as interviewers omitted births in the 2003 SADHS by simply leaving the questionnaire blank (Department of Health (South Africa) and Macro International Inc (USA), 2007).

The 2001 Census also failed to produce sensible and reliable estimates of child mortality owing to poor quality data which was generally incoherent and inconsistent, partly due to editing (Dorrington, Moultrie and Timæus, 2004). The failure of these surveys and censuses to produce reliable national estimates leaves researchers with only the 2007 Community Survey, HDSSs (e.g. Agincourt and Africa Centre) data and vital registration data to monitor child mortality in South Africa. Though completeness of registration of child deaths has improved, estimated to have reached 85% for infants and 58% for children aged 1-4 by 2011 (Dorrington, Bradshaw and Laubscher, 2014), the extent of completeness is still not certain given the lack of reliable recent estimates of child mortality. Also, data from the Health and Demographic Surveillance Systems (HDSSs) are not sufficiently representative to produce estimates that can be generalized for the whole South African population. However, the HDSSs provide insight into child mortality issues in South Africa as they are located in rural communities of high HIV incidence and prevalence (in the case of Africa Centre) and which are also poor (in the case of both Agincourt and the Africa Centre).

The need for greater certainty about the trend and level of estimates of childhood mortality in South Africa motivates this study. Uncertainty about the level and trend of childhood mortality in the HIV epidemic era provides impediments to monitoring the progress in childhood mortality (Kerber, Tuaone-Nkhasi, Dorrington *et al.*, 2012) and hence the effect of different policies on the health of children. Thus, this research will focus on revisiting childhood mortality estimation in South Africa with a view to update estimates for new data from the 2011 census and the South African National HIV, Behaviour and Health Survey conducted by HSRC in 2012 and to describe the trends in childhood mortality. These data provide an opportunity, and an alternative to models, with which to update childhood mortality estimates.

1.2 THE RESEARCH QUESTION

The main purpose of this research is to use the survival of most recent births data from the 2011 Census, deaths reported by households in the 2011 Census and full birth history data from the 2012 HSRC survey to produce estimates of infant and under-five mortality rates in South Africa by population group for the period 1998-2012, both to update previous estimates taking into account new data and to assess the reasonableness of all estimates.

1.2.1 SPECIFIC OBJECTIVES

The specific objectives of this research are as follows:

- To assess the quality of the 2011 Census and the 2012 HSRC survey data on child survival to see if useful estimates can be produced from these data.
- To produce IMRs and U5MRs from the 2011 Census and 2012 HSRC survey data by population group and sex.
- To adjust estimates produced using full birth history data from the 2012 HSRC survey for HIV/AIDS bias.
- To estimate the magnitude of bias in proportion dead among most recent birth relative to the proportion dead among all births.
- To describe and discuss the levels and trend in childhood mortality implied by the data from the 2011 census and the 2012 HSRC survey, how consistent they are with estimates produced by others.

1.3 STRUCTURE OF THESIS

This thesis is organized as follows. Chapter 2 reviews the literature on recent estimates of childhood mortality in South Africa that have been produced using different data sources as well as models. It also reviews the methods of producing childhood mortality estimates as well as the effect of HIV/AIDS on these methods.

Chapter 3 describes in detail the methods applied in this research to achieve the objectives set in the previous section. It starts by assessing the quality of the 2011 Census and the 2012 HSRC survey data and the steps followed in doing so. The next section of chapter 3 then outline in detail the steps followed in using the Previous Birth Technique (PBT), deaths reported by households in the 2011 Census and the direct method in producing estimates of childhood mortality. The steps followed in estimating the magnitude of the bias in proportion dead among most recent births relative to the proportion dead among all births are outlined.

Chapter 4 presents the detailed results and some analysis of those results. Finally, Chapter 5 reflects on the results obtained and deliberates on the conclusions that can be inferred from the research, limitations of the research and the scope for future research.

2 LITERATURE REVIEW

South Africa has many sources of data available for the assessment of childhood mortality, namely, 1987-1989 HSRC survey, 1996 and 2001 Censuses, 1998 and 2003 DHSs, 2007 Community Survey, the Agincourt and Africa Centre HDSSs and the vital registration data. These data sources provide an opportunity for comparing results produced by different sources (Joubert, Rao, Bradshaw *et al.*, 2012) and for testing the reasonableness of estimates produced using new data. However, problems with the quality of these data encouraged the development of models as an alternative to monitor progress in childhood mortality and to inform public health planning. Several models have been developed by different research groups (e.g. Actuarial Society of South Africa (2011), UN-IGME (2013) and IHME (Lozano, Wang, Foreman *et al.*, 2011; Wang, Dwyer-Lindgren, Lofgren *et al.*, 2012; Wang, Liddell, Coates *et al.*, 2014)). Apart from modelling, different methods for estimating childhood mortality have also been applied to South African data and these methods range from estimating mortality directly using full birth history data and deaths reported by households in censuses to indirect methods using summary birth history data. These methods have to be applied taking into account the effect of HIV on the estimates of childhood mortality because of the HIV/AIDS epidemiological profile of South Africa, which shows a high prevalence of the disease and prevalence which has not remained constant over time. In this chapter, recent estimates of childhood mortality in South Africa are discussed as well as the methods for estimating childhood mortality, taking into account the effects of HIV/AIDS on these methods.

2.1 RECENT ESTIMATES OF CHILDHOOD MORTALITY IN SOUTH AFRICA

Ideally, mortality estimates would be derived from vital registration data. However, in sub-Saharan Africa, these data are characterised by incompleteness such that mortality measurement become dependent on population censuses and national surveys (Joubert, Rao, Bradshaw *et al.*, 2012). Kahn, Garenne, Collison *et al.* (2007) argue that reliability of South African vital registration and census data was problematic due to data quality issues and completeness of death registration, particularly in children, when they reported on a study using Agincourt data for the period 1992 to 2003, and this situation is still apparent in these data sources even today, though completeness in death registration has increased to around 85% in infants and 58% in children aged 1-4 (Dorrington, Bradshaw and Laubscher, 2014).

2.1.1 ESTIMATES FROM EMPIRICAL DATA

The 1996 Census and the 1998 DHS are the first nationally representative inquiries to be conducted in the post-apartheid South Africa that collected useful data on childhood mortality (Nannan, Timæus, Laubscher *et al.*, 2007). Prior to the abolition of the Apartheid system, childhood mortality was analysed using the DHS-type survey conducted by HSRC between 1987 and 1989 and the 1993 Living Standards and Development Survey which was sponsored by the World Bank. Other sources which attempted to collect childhood mortality information were the 2001 Census, the 2003 DHS and the most recent 2011 Census and 2012 HSRC survey.

2.1.1.1 ESTIMATES FROM CENSUSES

The 1996 Census was the first since the abolition of the Apartheid system to attempt to enumerate the whole South African population. Between 1970 and 1996, the censuses were split between 'South Africa' and the so-called independent homelands. However, not all of the 'homelands' were able to undertake censuses and for some of those that could, the data are no longer available. Even the 'South African' censuses in this period were characterized by severe coverage errors. The 2001 Census asked women of reproductive age a question about their lifetime fertility and the survival of children ever born to them (Nannan, Timæus, Laubscher *et al.*, 2007). To assess the levels and trend in childhood mortality from the 1996 Census, Nannan, Timæus, Laubscher *et al.* (2007) applied the Brass CEB/CS technique. The estimates produced using the CEB/CS techniques were found to be too high owing to the errors in the data and the failure of the technique to perform well in the South African data. As such adjustments had to be made to correct the data for excess mortality risks of children born to teenage mothers (Collumbien and Slogget, 2001) and misreporting of still-births as live-births based on the findings of the analysis of the 1996 Census and 1998 DHS (Moultrie and Timæus, 2003). Their conclusions were that the misreporting of still-births were a major concern of the quality of the 1996 Census data since, unlike the 1998 DHS, the census did not have questions that probed women of the actual outcome about their pregnancies (Nannan, Timæus, Laubscher *et al.*, 2007). The 1996 Census estimates suggest that childhood mortality had been declining since the 1980s until around 1993, with the decline more rapid in the 1980s (Nannan, Timæus, Laubscher *et al.*, 2007). Adjusted census estimates yielded estimates that were consistent with the 1998 DHS estimates, thus giving confidence in the estimates produced up to 1996 (Nannan, Dorrington, Laubscher *et al.*, 2012).

The 2001 Census failed to produce estimates of childhood mortality owing to the incoherence and inconsistencies of the data. Dorrington, Moultrie and Timæus (2004) interrogated the 2001 Census data on children surviving and children ever born. Their conclusion was that neither the edited nor the unedited data yielded plausible estimates of CEB/CS ratios and hence they were unable to produce estimates of childhood mortality using these data. Instead, for "illustrative purposes" (Dorrington, Moultrie and Timæus, 2004: 68), they used deaths in the last 12 months reported by households together with 2001 Census estimates of children under the age of five to produce an estimate of the recent probability of dying before the age of five.

2.1.1.2 ESTIMATES FROM DHS-TYPE SURVEYS

Prior to the abolition of the apartheid system in South Africa, the main source of childhood mortality was the 1987-1989 HSRC survey. This survey documented declines in childhood mortality for all population groups, with the greatest decline observed among Asians, Coloureds and Whites (Nannan, Timæus, Laubscher *et al.*, 2007). Towards the end of the apartheid system, the Living Standards and Development Survey was conducted in 1993 (Mazur, 1995). The estimates from this dataset were observed to be higher than in other surveys, but it confirmed that mortality was declining in the 1980s as was observed from the 1987-1989 HSRC survey (Mazur, 1995; Nannan, Timæus, Laubscher *et al.*, 2007).

The 1998 DHS was the first to be conducted after the abolition of apartheid in South Africa and the survey used a two-stage cluster sampling design to draw a nationally representative sample of 11,735 women aged 15-49 years. The survey collected full birth history data which were used to estimate childhood mortality both directly and indirectly. Nannan, Timæus, Laubscher *et al.* (2007) used the CEB/CS technique to produce indirect estimates of childhood mortality, which were consistent with adjusted estimates from the 1996 Census. Nannan, Timæus, Laubscher *et al.* (2007) also produced direct estimates of infant and under-five mortality rates. The direct estimates for U5MRs were found to be lower than the indirect estimates produced both from the 1998 DHS and 1996 Census (Nannan, Timæus, Laubscher *et al.*, 2007). The indirect estimates suggested a similar downward trend in U5MRs to that of the 1996 Census but differed from the 1996 Census in terms of the level of U5MRs. This trend was also confirmed by the direct estimates from these data which suggested that mortality declined between 1975 and 1993 before it started to reverse (Nannan, Timæus, Laubscher *et al.*, 2007).

The 2003 DHS failed to produce plausible estimates of childhood mortality due to problematic fieldwork (Department of Health (South Africa) and Macro International

Inc (USA), 2007). Nannan, Dorrington, Laubscher *et al.* (2012) argue that 11% of under-five mortality in the five years before the survey was estimated to be missed and 17% in the five to ten years before the survey. Generally, the estimates of U5MR from the 2003 SADHS were lower than those from the 1998 SADHS for the same time periods (Bradshaw and Dorrington, 2007). The estimates of the U5MR for the five-year period centred around 2001 are 58 per 1000 live-births from the 2003 DHS compared to 61 deaths per 1000 live-births for the 5-year period centred around 1996 from the 1998 DHS, which is implausible since HIV prevalence rates were still rising and therefore a decrease in child mortality would not be expected. Also, provincial estimates of U5MR from the 2003 DHS are inconsistent with the estimates from the 1998 DHS, with KwaZulu-Natal (KZN) estimated to have the lowest U5MR of 33 deaths per 1000 live-births from the 2003 DHS which is highly unlikely given that KZN has the highest prevalence of HIV/AIDS. The 2003 SADHS was particularly badly done in that province.

2.1.1.3 ESTIMATES FROM THE 2007 COMMUNITY SURVEY

Darikwa (2009) estimated the levels and trends of infant and under-five mortality in South Africa from 1996 to 2006 using the 2007 Community Survey. The survey included questions on children ever born, children dead, survival of previous birth and household deaths in the previous 12 months. With these data, he produced estimates of IMR of 46 deaths per 1000 using the Pervious Birth Technique (PBT) and 53 deaths per 1000 from deaths reported by households as estimates for the year centred on the year before the survey reference point (2007.125). The PBT is known to underestimate infant mortality due to the selection bias of infants deaths that are immediately followed by another birth (Hill, 2012c). To arrive at the final estimate of infant mortality in 2006, Darikwa (2009) took the average of these two estimates and reported an IMR of 49.1 deaths per 1000 live-births. For the U5MR in 2006, the CEB/CS technique was found to give a high estimate of 80 deaths per 1000 live-births after adjusting for the bias due to HIV/AIDS. This estimate was scaled to maintain the ratio of IMR to U5MR of 0.66 which corresponds to the ratio of registered deaths in these ages and the ratio of rates from other sources (e.g. the ASSA2003 population projection model) thus a final U5MR estimate of 74.7 deaths per 1000 live-births was used (Darikwa, 2009).

The final estimates for 1996, taken as the average of direct and indirect estimates from the 1998 SADHS, were reported as 50.5 deaths per 1000 live-births for IMR and 67 deaths per 1000 live-births for U5MR (Darikwa, 2009). For the period 2000 to 2004, the IMRs and U5MRs were derived using the Ward and Zaba (2008) variant of the Brass

CEB/CS technique, after adjusting for the fact that prevalence is not constant in South Africa by scaling the Ward and Zaba (2008) correction factors using the ratio of the prevalence at the time of birth of the child to the prevalence at the time of the survey (Darikwa and Dorrington, 2011). The trend in IMR and U5MRs between 2000 and 2004 from the Ward and Zaba (2008) variant suggests that mortality has been fairly constant in this period. This trend in U5MR and IMR observed by Darikwa (2009) suggest that the indirect method used might have flattened any changes in mortality as the CEB/CS technique tends to average estimates.

2.1.1.4 ESTIMATES FROM VITAL REGISTRATION DATA

The South African Medical Research Council used data from Vital Registration for the period 1996 to 2006, together with completeness of death registration information reported by Darikwa (2009), to estimate the trend in IMR and U5MR (Nannan, Dorrington, Laubscher *et al.*, 2012). The reported estimates of U5MR in this study were 63.9 deaths per 1000 live-births in 1996 and 71.8 deaths per 1000 live-births in 2006. The overall trend from vital registration, adjusted for completeness of death registration, suggests that IMR and U5MR were relatively constant between 1998 and 2006. The MRC observed an unexpectedly large increase in IMR and U5MR between 1997 and 1998 from vital registration data. Analysis of the completeness of the vital registration of deaths suggested that child deaths registration improved during 1997 and 2007 reaching levels of 90% completeness for infants and 60% for children aged 1-4 years (Darikwa, 2009). However, the recent levels of completeness is still uncertain given the lack of recent estimates of childhood mortality that are reliable (Dorrington, Bradshaw and Laubscher, 2014).

Bradshaw, Dorrington and Laubscher (2012) used cause-of-death data from Statistics South Africa (Stats SA) and data from the rapid mortality surveillance (RMS) to estimate the trend in IMR and U5MR between 2000 and 2011. The estimates of IMR and U5MR were updated to include estimates for the year 2012 in the second RMS report (Dorrington, Bradshaw and Laubscher, 2014). The RMS database was established by the Medical Research Council (MRC) and University of Cape Town (UCT) in 1999 using monthly information about deaths in the national population register which is maintained by the Department of Home Affairs. The cause of death data from Stats SA, compiled using death certificates, were available up to the year 2010 (Statistics South Africa, 2013) therefore the data from RMS were used to estimate mortality for the years 2011 and 2012 (Dorrington, Bradshaw and Laubscher, 2014) so as to get, inter alia, the latest estimates of childhood mortality in South Africa using empirical data. The RMS data suffer from under-reporting of deaths due to non-registration on the population register and non-registration of deaths. The RMS data on the number of deaths were consistent in trend with the Stats SA data, although they were relatively lower in number of deaths since the deaths in the RMS are based on registration on the population register and therefore deaths without identity documents are missed by the RMS. As such, the RMS data were adjusted for completeness of death registration using the estimates of q_0 and ${}_5q_0$ derived from different sources including deaths reported by households in the 1996, 2001 and 2011 Censuses (Dorrington, Bradshaw and Laubscher, 2014). The trend in under-five mortality was found to have peaked in 2004 reaching 80 deaths per 1,000 live-births and falling gradually to less than 70 deaths per 1,000 live-births in 2008. The trend in under-five mortality rate was estimated to have decreased markedly between 2008 and 2011 using RMS data adjusted for completeness of death registration. The RMS data suggest that U5MRs were 56, 52, 40 and 41 deaths per 1000 live-births in 2009, 2010, 2011 and 2012 respectively, a decrease of about 15 deaths per 1000 live-births in four years.

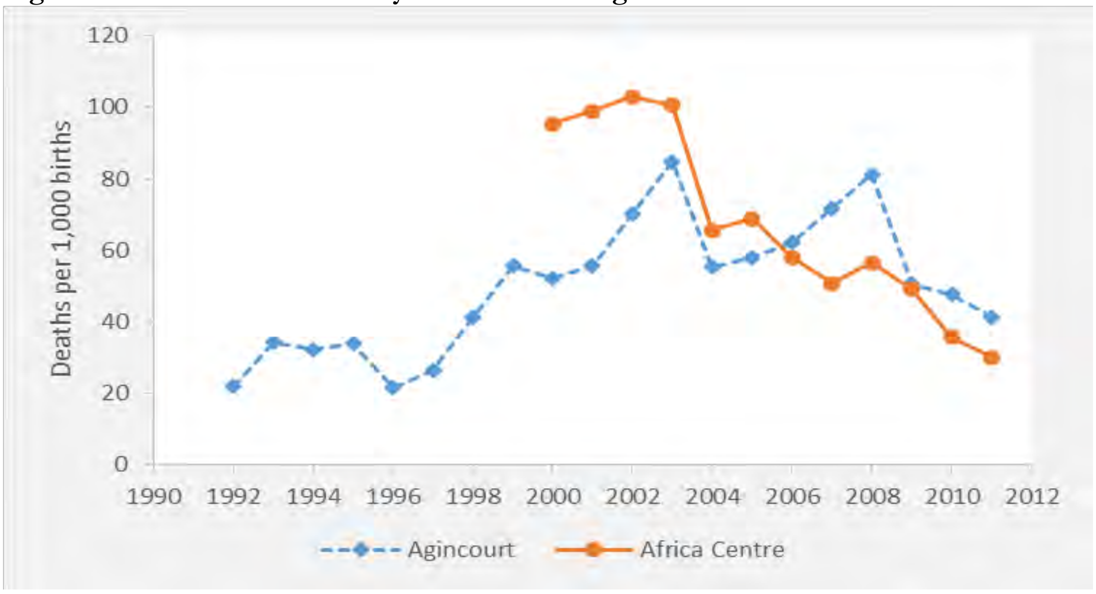
2.1.1.5 ESTIMATES FROM HEALTH AND DEMOGRAPHIC SURVEILLANCE SITES (HDSSs) DATA

Using Agincourt DHSS data, Kahn, Garenne, Collison *et al.* (2007) found a differential increase in under-five mortality by sex after comparing the 1992-93 rates to the 2002-03 rates. They found that the increase in mortality between 1992-93 and 2002-03 was statistically significant in both males and females, with male mortality estimated to have increased by 2.29 times and female mortality to have increased by 1.75 times. However, the difference in the rate of increase of mortality by sex was found to be statistically insignificant at $p = 0.05$.

The under-five mortality in Agincourt increased between 1996 and 2003 and declined between 2003 and 2004 before starting to rise again until 2008 (Figure 2.1). For the period between 2008 and 2011, under-five mortality was shown to have declined in Agincourt (Figure 2.1). For the Africa Centre HDSS (Indepth Network, 2014), under-five mortality was shown to have increased from the year 2000 and reached its peak in 2002 (Figure 2.1). For the period between 2002 and 2011, under-five mortality declined in Africa Centre (Figure 2.1). The mortality increase observed in the Agincourt and Africa Centre data was attributed to the emergence of the HIV/AIDS epidemic (Kahn, Garenne, Collison *et al.*, 2007).

Kahn, Garenne, Collison *et al.* (2007) critique their estimates and findings from Agincourt HDSS citing that the annual visits to households in the Agincourt HDSS limit the recording of pregnancies and therefore might possibly results in the undercount of infant deaths. However, in spite of the possibility of undercount in infants deaths in the Agincourt HDSS and the fact that Agincourt and Africa Centre are not representative of the whole South African population, the trends in childhood mortality shown in Figure 2.1 conform very much to the South African childhood mortality trends reported in other studies.

Figure 2.1 Under-five mortality rates from the Agincourt and Africa Centre HDSSs



Source: Indepth Network, http://www.indepth-ishare.org/indepthstats/index.php/graphs

2.1.2 ESTIMATES FROM MODELS

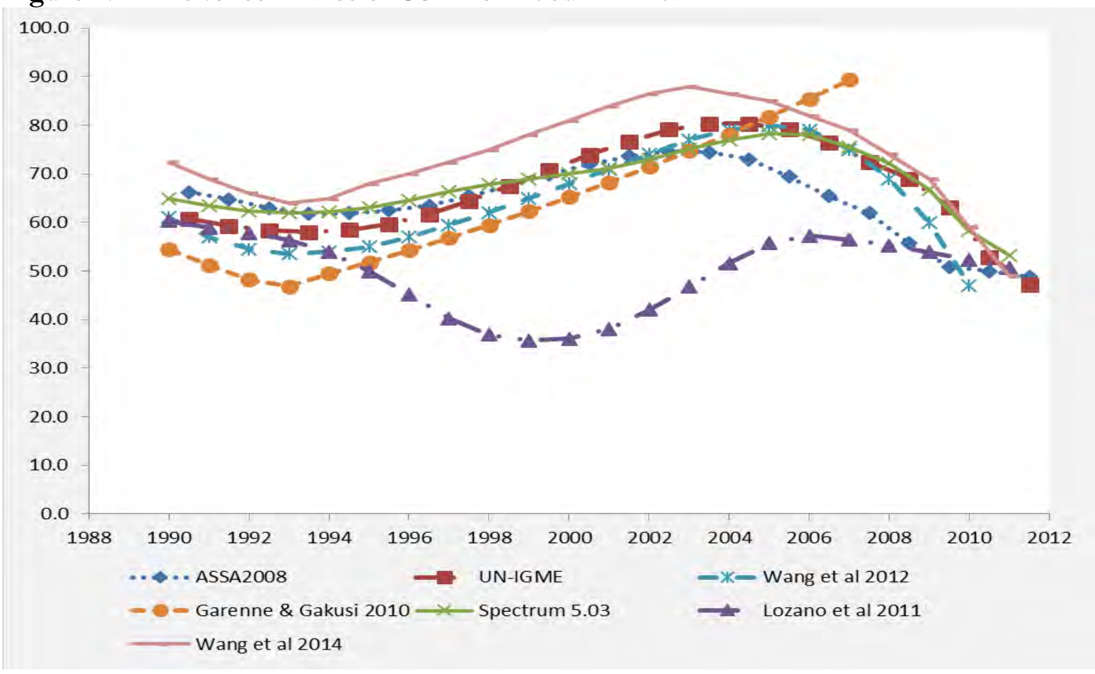
The model estimates produced for South Africa vary on the level and trends due to the difference in assumptions and data used (Bradshaw and Dorrington, 2007) thereby creating uncertainty about the estimates of childhood mortality for South Africa.

Garenne and Gakusi (2005) reconstructed the trend in under-five mortality rates using the 1998 SADHS and estimate mortality to have declined rapidly from 1968 to 1992, before increasing rapidly from 1992 to 2006 (Figure 2.2). These estimates were updated in 2010 (Garenne and Gakusi, 2010). The reversal in mortality decline was attributed to an increase in HIV/AIDS mortality and stagnation in the health transition (Garenne and Gakusi, 2005). However, Nannan, Dorrington, Laubscher *et al.* (2012) argue that estimates by Garenne and Gakusi (2005) possibly exaggerated the trough in U5MR between 1960s and early 1990s.

The IHME estimates of childhood mortality were produced using the 2010 and 2013 Global Burden of Diseases (GBD) studies (Lozano, Wang, Foreman *et al.*, 2011;

Wang, Dwyer-Lindgren, Lofgren *et al.*, 2012; Wang, Liddell, Coates *et al.*, 2014). Lozano, Wang, Foreman *et al.* (2011) produced estimates of childhood mortality for IHME using an ensemble model which was built using data from vital registration, Health and Demographic Surveillance Sites (HDSSs), summary birth histories from censuses and surveys and deaths reported by households in censuses and surveys included in the 2010 GBD study. The rate of change in the under-five mortality rate (U5MR) was used as an indicator of what countries might achieve by 2015, i.e. to see if they were on track to meet the MDG 4 target. For South Africa, U5MR was estimated to be 50.7 deaths per 1000 live-births in 2011. Under-five mortality was estimated to have decreased between 1990 and 1999 and to have increased between 2000 and 2006 (Figure 2.2). The reversal in previous mortality gains was estimated to have peaked in 2006 before mortality started to fall again (Lozano, Wang, Foreman *et al.*, 2011). This trend in childhood mortality was not consistent with other trends produced elsewhere using empirical data or models as shown in Figure 2.2 above. The progress towards achieving MDG 4 in South Africa was estimated to be lagging behind by 25 years, i.e. the target will be achieved in 2040.

Figure 2.2 Model estimates of U5MRs in South Africa



The IHME estimates produced by Lozano, Wang, Foreman *et al.* (2011) were reworked by Wang, Dwyer-Lindgren, Lofgren *et al.* (2012) and these are more consistent with other estimates produced by other researchers. In calculating the number of deaths and exposure to risk of dying, Wang, Dwyer-Lindgren, Lofgren *et al.* (2012) considered calendar years by dividing each one-year birth cohort into 52 birth week cohorts and following them through to age five. The results for South Africa, shown in Figure 2.2, suggests that mortality reversed around 1993 and continued to increase up to around 2006 before it began to fall again. This is a completely different trend in U5MR to that produced by Lozano, Wang, Foreman *et al.* (2011) in a space of one year. The major difference can be attributed to the difference in the South African data used in the IHME model by these two groups of researchers. Wang, Dwyer-Lindgren, Lofgren *et al.* (2012) dropped some of the data that were used by Lozano, Wang, Foreman *et al.* (2011), particularly the 2001 Census data on summary birth histories which appear to be suggesting a downward trend when mortality was actually increasing which other researchers had known to be problematic (Dorrington, Moultrie and Timæus, 2004).

Wang, Liddell, Coates *et al.* (2014) updated the IHME estimates using data from vital registration, Health and Demographic Surveillance Sites (HDSSs), summary birth histories from censuses and surveys and deaths reported by households in censuses and surveys included in the 2013 GBD study. They used the Gaussian Process Regression (GPR) with adjustments for bias and non-sampling error to synthesize the data. Their results, shown in Figure 2.2, suggests a similar trend in U5MR to the one produced by Wang, Dwyer-Lindgren, Lofgren *et al.* (2012), although their levels of under-five mortality were higher than those produced by Wang, Dwyer-Lindgren, Lofgren *et al.* (2012).

The UN-IGME (2013) also recently published estimates of infant and under-five mortality up to the year 2012. The method used involves projecting non-AIDS mortality from a pre-HIV epidemic period using the Spectrum population projection package and then calculating the proportion of mortality attributable to HIV/AIDS so as to adjust estimates for the bias due to HIV (Walker, Hill and Zhao, 2012). The estimate of the trend in under-five mortality is produced from the available national data using a Bayesian B-spline statistical model (Alkema and New, 2013). Since the emergence of HIV/AIDS epidemic do not appear to substantially modify the pre-existing balance between infant and child mortality (Guillot, Gerland, Pelletier *et al.*, 2012), infant mortality rates are estimated by transforming the under-five mortality rates using model life tables, for South Africa, the Coale and Demeny (1966) West model life tables were used. This method resulted in a trend in childhood mortality that suggested a downward trend in mortality up to around 1995 before mortality reversed to peak at around 2004, which is a similar trend to that produced by other models with different assumptions (e.g. the ASSA2008 population projection model). The UN IGME estimates suggest that under-five mortality declined at an annual rate of 1.4 percentage points between 1990 and 2011, with an uncertainty interval of 0.4 percentage points to 3.4 percentage points, reflecting the extent of uncertainty in these estimates.

The data sources included in the IHME and UN-IGME models were different, with UN IGME excluding data sources with values that were consistently below or above those of other sources or if data quality issues were reported, while IHME excluded outliers based on a "discussion" of pervious mortality estimates. Though differences exist partly because of differences in methods used, Alkema and You (2012) argue that much of the differences can be attributed to unreliable data and these differences are likely to decrease as more and better data become available.

A comparison of the model estimates of childhood mortality (Figure 2.2) suggest that mortality reversed around the middle of the 1990s. It can be noted from Figure 2.2 that there is uncertainty as to when the reversal of mortality peaked and the actual levels of mortality. The estimates of the levels of mortality are as varied as there are models. One important difference in these estimates and the trends can be attributed to the assumptions made by these models on the course of the HIV epidemic which emerged in the mid of the 1990s and had a great impact on the mortality profile of South Africa. These models make different assumptions on the treatment rollout and the impact of treatment interventions, use different data, but the big difference probably lies in the way the models model the impact of HIV/AIDS.

2.2 METHODS OF ESTIMATING CHILDHOOD MORTALITY

The second half of the 20th century has been characterised by the development of different methods of estimating childhood mortality. Estimating childhood mortality correctly is important in policy recommendations (Adetunji, 1996). The different methods of estimating childhood mortality can be broadly categorized into two: indirect and direct methods. The indirect methods are based on the number of children ever borne, the number of these dead/alive reported by women and the survival of previous births (Adetunji, 1996). Direct methods are based on vital registration, deaths reported by households in censuses or dated vital events from full birth history data (Adetunji, 1996). In contrast to direct methods, indirect methods depend on model life tables and fertility schedules of a population (Adetunji, 1996). This section describes and discusses the indirect methods and then the direct methods used in this research before proceeding to discuss the limitations and suggested adjustments to be made to these methods in populations with generalized HIV/AIDS epidemics.

2.2.1 INDIRECT METHODS: UNADJUSTED FOR THE BIAS DUE TO HIV/AIDS

Indirect methods of estimating childhood mortality are based on summary birth history data and information on survival of the previous birth collected in a census or survey. There are two such methods, both proposed by William Brass, and these methods are the Children Ever Born/Children Surviving method (Brass and Coale, 1968) and the Previous Birth Technique (Brass and Macrae, 1984). These methods have been described and also discussed recently in the tools for demographic estimation (Moultrie, Dorrington, Hill *et al.*, 2012). The following sub-section describes the previous birth technique.

2.2.1.1 THE PREVIOUS BIRTH TECHNIQUE (PBT)

The method was first proposed by Brass and Macrae (1984). The idea is to use data provided by women at the time of maternity to estimate child mortality. With some simple birth interval distributions and a variety of mortality models, Brass and Macrae (1984) showed that the proportions dead amongst the previously born children provides a good approximation to the probability of dying before the second birthday, $q(2)$. Extending these ideas to the survival of the second-to-last born children and taking $I = 30$, where I is the mean birth interval in months and is appropriate only in non-contracepting populations, Hill and Aguirre (1990) showed that the proportions dead among the second-to-last born children at the current maternity will approximate $q(2.I)$, i.e. $q(5)$.

The method by Hill and Aguirre (1990) of using the current birth as an index birth and analysing the survival of a birth preceding the index birth suffers from selection bias since data will be collected from women of higher parities (women who have given birth to two or more children). The proportions of mothers who give birth in hospitals vary widely due to different completeness in maternity coverage thus the data lacks the total numbers exposed to risk unless all the mothers are asked about the survival of their previous births. As such, these data cannot be used to estimate child mortality rates (Aguirre, 1994).

To solve the problem of bias due to incomplete coverage by health facilities, Aguirre (1994) suggested the use of a cross-sectional survey which takes a sample from the whole population including mothers who give birth outside of health facilities. Today, questions on the survival of previously born children are widely included in censuses and surveys and Blacker and Brass (2005) have developed a method to estimate infant mortality rate based on these data. Having realised that results of infant mortality rates calculated using the proportion dead among those born in the past twelve months have been widely unacceptable due to rounding up of the age of children born nearly twelve months prior to the survey, Blacker and Brass (2005) suggested the use of births occurring 24 months prior to the survey and an adjustment factor to calculate infant mortality rate.

Using life table notation, the proportion of children surviving from births occurring in the last 12 months is ${}_1L_0$ and that from the 12 months before that is ${}_1L_1$. Thus the proportion dying among births in the last 24 months will be

$$D = 1 - \frac{{}_2L_0}{2}$$

$$D = 1 - \frac{1}{2}\int_0^2 l(x)dx$$

Assuming that the probability of survival from birth up to age 10 can be described by the following curve

$$l(x) = (1 + \alpha x)^{-\beta}$$

where x is age and alpha and beta are constants which give the level and shape of mortality, the infant mortality rate will then be defined by

$$q(1) = 1 - l(1) = 1 - (1 + \alpha)^{-\beta}$$

Thus the adjustment factors needed to convert D into $q(1)$ can be calculated given values of alpha and beta (Blacker and Brass, 2005). Using 120 simulated survival curves, the adjustment factor was found to be unrelated to the level parameter and is essentially determined by the age pattern of mortality as defined by alpha (Blacker and Brass, 2005). An adjustment factor of 1.09 was proposed for populations about which nothing is known, and a lower adjustment factor of not less than 1.04 for populations with low mortality, say less than 50 deaths per 1000 births, or where child mortality is unusually high/low relative to infant mortality (Blacker and Brass, 2005). The errors due to the use of an inappropriate adjustment factor will usually be trivial as compared to those due to data errors (Blacker and Brass, 2005).

The method proposed by Blacker and Brass (2005) has limitations. One such limitation peculiar to the Blacker and Brass (2005) variant of the PBT arises from the assumption that most recent births in the last 24 months approximates all births in the last 24 months. However, analysis of DHS data reveal that, for women with more than one birth in the past 24 months, the earlier births are much less likely to have survived than the most recent birth (Hill, 2012c) and thus those dead children are less likely to be reported. Therefore, the proportion dead among most recent births in the 24 months is biased downwards relative to the proportion dead among all births. The bias has been estimated to be about 20 per cent on average (Hill, 2012c). Hill (2012c) concluded that analysis of information based on survival of most recent birth as proposed by Blacker and Brass (2005) is not recommended due to selection bias for births occurring in the last 24 months before the census or survey. However, if the bias in the proportion dead among most recent births occurring in the 24 months before the census or survey relative to all births occurring in the 24 months before the census or survey could be estimated and allowed for in the method, it is possible to produce useful estimates of infant mortality rate.

2.2.2 DIRECT ESTIMATION: UNADJUSTED FOR THE BIAS DUE TO HIV/AIDS

Mortality rates can be estimated directly from deaths reported by households in censuses and full birth history data reported in surveys such as the DHS. Direct estimation of mortality, as highlighted earlier, is different from indirect estimation in that it does not depend on model life tables and fertility schedules (Adetunji, 1996). Direct estimation also has an advantage in that it enable the estimation of child mortality rates that are more recent than those estimated from indirect methods. This section describes and discusses direct estimation of childhood mortality based on deaths reported by households and full birth history data. The effects of HIV/AIDS on these methods are discussed in a later section.

2.2.2.1 DEATHS REPORTED BY HOUSEHOLDS

The method, in its simplest form, involves calculating the probability of dying between ages x and $x+1$ by using the number of children dead between ages x and $x+1$ in the 12 months before the census, for $x<5$, and the census estimate of the population aged between x and $x+1$ at the time of the census. Given the central mortality rate as

$$M_x = \frac{D_x}{\overline{P}_x}$$

where D_x is the number of children dying between x and $x+1$ and $\overline{P}_x$ is the population aged between x and $x+1$ that is exposed to the risk of dying in the year before the census. The population exposed to the risk of dying between ages x and $x+1$ can be calculated as the geometric mean of $P_x(1)$ and $P_x(2)$ where $P_x(1)$ is the population aged between x and $x+1$ a year before the census and $P_x(2)$ is the census estimate of the population aged between x and $x+1$. The population aged between x and $x+1$ a year before the census, $P_x(1)$, can be estimated for $x=1,2,3,4$ by

$$P_x(1) = P_x(2) + D_x^{'}$$

where $D_x^{'} = 0.5 \times (D_x + D_{x+1})$ are deaths aged x to $x+1$ one year before the census. The probability of dying between ages x and $x+1$, for $x=1,2,3,4$, can then be estimated as

$$q_x = \frac{m_x}{1+(a_x \times m_x)}$$

assuming that $m_x = M_x$ (Preston, Heuveline and Guillot, 2001) and where a_x is the average number of person years lived by those who die between ages x and $x+1$. Preston, Heuveline and Guillot (2001) provide a table of estimating a_x given a value for m_x, adapted from Coale and Demeny (1966) model life tables. The probability of dying before the age of one, q_0, can be estimated as

$$q_0 = \frac{D_0}{B}$$

where B is the number of births occurring in the 12 months before the census. The probability of dying before the age of five, ${}_5q_0$, is then given by

$${}_5q_0 = 1 - \prod_{x=0}^{4}(1-q_x)$$

The accuracy of estimating childhood mortality rates from deaths reported by households in a census has not been thoroughly tested in developing countries (Dorrington, Moultrie and Timæus, 2004). In South Africa the method has been applied to the 2001 Census data (Dorrington, Moultrie and Timæus, 2004) and to the 2007 Community Survey data (Darikwa, 2009). Dorrington, Moultrie and Timæus (2004) noted that deaths reported by households have a poor record of estimating child mortality rates accurately, citing the Swaziland 2007 Census as an example where these data failed to produce reasonable estimates. They thus caution researchers against using these data in isolation but advised that these data should be used in combination with data from other sources such as vital registration systems.

The major limitations of this method in estimating child mortality rates arises from the incompleteness of death reporting by age, possible over-reporting of deaths especially of young adults if the young adults have recently moved out of the family home and, for infant mortality, the undercount of births occurring 12 months prior to the census. For example, Darikwa (2009) observed an undercount of births in the South African 2007 Community Survey and thus, had to obtain births from the ASSA2003 population projection model. As was discussed in the section on the previous birth technique, the undercount of births occurring 12 months before the census is due to age heaping on the date exactly 12 months before the census (Blacker and Brass, 2005).

2.2.2.2 FULL BIRTH HISTORY DATA

The method makes use of the month and year of birth for each child born to a woman (recorded in century-month-code (CMC)), the vital status of that child and the age at death of that child if dead. Cohort estimates of childhood mortality are calculated as the number of children dead divided by the number of births (Hill, 2012a). For IMR, one needs data for children in the period 12 to 23 months before the survey and for U5MR, one needs data for children in the period 5 to 10 years before the survey. Thus, the cohort estimates of IMR reflects mortality risks for the 2 year period before the survey and U5MR reflects mortality risks for the 10 year period before the survey thus there is no period interpretation of these estimates (Hill, 2012a).

Where period estimates are needed, these are derived using the synthetic cohort concepts of life tables as explained in the section on using deaths reported by households. Preston, Heuveline and Guillot (2001) also give detailed explanations on the derivation of period life tables. The only difference between the method of using death reported by households and full birth history data is the derivation of the population exposed to risk. In using full birth history data to derive the population exposed to risk of dying in a given period one needs specific dates of birth and specific ages at death to avoid heaping at boundaries and these have to be imputed using random numbers since they are not reported by mothers (Hill, 2012a). Due to the lack of reproducibility when a true random number generator is used, the use of a variable that are not correlated to child mortality such as household number is recommend. Hill (2012a) proposed the use of the last digit of the day of interview and the household number to generate the random numbers to be used to impute dates of birth and ages at death is proposed. Once the specific dates of birth and ages at death have been imputed, the central mortality rate by each month of life can be defined by

$$M(x,j)=\frac{\sum_{i=1}^{N}D(i,x,j)}{\sum_{i=1}^{N}E(i,x,j)}$$

where $M(x,j)$ is the central mortality rate for age x months and year j, $D(i,x,j)$ is a binary variable indicating the death of child i aged x months in year j and $E(i,x,j)$ is the exposure-time of child i aged x months in year j (Hill, 2012a). Once the central mortality rates have been calculated, the probabilities of dying can then be computed as

$$q(x,j)=\frac{\frac{M(x,j)}{12}}{1+\frac{M(x,j)}{24}}$$

and the probability of dying before the age of five as

$${}_5q_0=1-\prod_{x=0}^{59}(1-q(x,j))$$

The underlying assumptions of this method are that completeness of reporting is the same for children alive and children dead, dates of birth and ages at death are reported with reasonable accuracy, mortality of children is independent of mortality of mothers and that mortality of children is independent of mother's age. These assumptions are violated particularly due to the correlation of mothers' mortality and children's mortality in populations with high HIV prevalence. Such violations due to the correlation of mortality of mothers to that of children are examined in a later section.

Other limitations of the method arise due to truncation bias in the data as one looks further back in time if an upper age limit is applied to the age of the mothers. Data for earlier periods will be increasingly based on the experience of younger women and thus an over-representation of the first births among younger women, meaning that child mortality is likely to be increasingly overestimated for earlier time periods (Hill, 2012a) since children born to younger women are known to experience mortality that is above average. However, overestimation due to truncation of data is counter-balanced by underestimation due to recall bias and selective omission of children dying further back in time (Hill, 2012a) thus the bias may be insignificant.

2.3 EFFECTS OF HIV/AIDS ON THE ESTIMATION OF CHILDHOOD MORTALITY

The indirect and direct methods are both based on the assumption that the mortality experience of children is independent of that of their mothers. In high HIV/AIDS prevalence populations, the mortality of mothers and mortality of children become much more highly correlated as dead mothers cannot report on the survival of their children who might have experienced higher mortality. Mortality of children born to HIV-positive women is higher than that of children born to HIV-negative women (Mahy, 2003) and thus the Brass CEB/CS method and the direct estimation using full birth history data underestimate child mortality rates unless adjusted for the bias in mortality estimates due to HIV/AIDS.

The mortality of mothers becomes much more highly correlated to that of their children due to mother-to-child transmission (MTCT) of HIV. Mother-to-child transmission of HIV causes infection of children during pregnancy, at birth or through breast feeding. This is sometimes referred to as vertical transmission. In the absence of prevention, the vertical transmission among HIV-positive mothers is estimated to be around 32 per cent, and it is estimated that without treatment about 60 per cent of the children who are infected during pregnancy or at birth die before their fifth birthday (Mahy, 2003). However, with treatment of mothers the bias on mortality estimates due to HIV/AIDS is expected to be much less. In South Africa, the prevention of mother-to-child transmission (PMTCT) coverage is now almost complete, with coverage estimated to be 96 per cent (Mayosi, Lawn, van Niekerk *et al.*, 2012), with less than two per cent becoming infected in the first six weeks. This has reduced the bias due to HIV/AIDS on the estimates of mortality in the most recent past (Bradshaw, Dorrington and Laubscher, 2012).

Section 2.3.1 describe and discuss the methods for adjusting for bias in estimates produced by the direct method due to HIV/AIDS, with more attention being given to the method that will be applied in this research.

2.3.1 THE EFFECT OF HIV/AIDS ON DIRECT ESTIMATES

Direct estimation is based on the assumption that mortality of mothers and their children are independent. However, in direct estimation of child mortality, this assumption is less important than in the indirect methods, especially with respect to the most recent estimates of child mortality (Mahy, 2003). The overall bias due to HIV rarely exceed five per cent for the most recent five-year period (Artzrouni and Zaba, 2003), and thus estimates from direct methods are more reliable in the most recent five-year time period (Mahy, 2003). This is because most births in the most recent five-year period occur to young mothers who would have been recently infected and are likely to be still alive to report the survival of their births.

2.3.1.1 THE HALLETT, GREGSON, KURWA ET AL. (2010) APPROACH

Hallett, Gregson, Kurwa *et al.* (2010) used data from a cohort study in Manicaland, Zimbabwe, to estimate the bias due to HIV/AIDS and through this to inform the construction of a model to estimate the bias in child mortality estimates due to HIV/AIDS. The cohort was interviewed between 1998 and 2000, with follow-up interviews at three and five years. At each interview round, child deaths since the previous interview were recorded. In the final interview in 2005, a full birth history data was collected from the surviving women and the U5MR was estimated for the period 1998 to 2005. True levels of the U5MR were calculated by adding back the mortality experience of children born to mothers who had died by the time of interview in 2005. The bias, calculated as the estimates from reports by surviving mothers divided by the estimates for all mothers was 6.7 per cent for infant mortality and 9.8 per cent for the U5MR. The bias in the direct estimates increased with the duration of the epidemic (Hallett, Gregson, Kurwa *et al.*, 2010). However, their model did not allow for the extensive provision of treatment, which could be expected to reduce the bias in future.

An individual-based stochastic model using parametric probability distributions to describe the relationships between age, fertility, risk of acquiring HIV infection, time since HIV infection, AIDS-related mortality rates and background mortality was created to produce three DHS-type time series data designated as 'DHS analogous', 'DHS continuous' and 'corrected' data (Hallett, Gregson, Kurwa *et al.*, 2010). The estimate of the extent of the bias was taken to be the difference between estimates from the 'DHS continuous' and the 'corrected' time series data since the 'DHS continuous' time series did not include reports of women who had died of AIDS before the survey, whereas the 'corrected' time series included these reports (Hallett, Gregson, Kurwa *et al.*, 2010). This mathematical model was applied to data from six other countries in sub-Saharan countries with HIV epidemics of differing intensity and having different background mortality of infants. The model requires HIV prevalence rates, which are taken from UNAIDS prevalence data, as well as estimates of mortality and fertility rates, which were taken from DHSs conducted in pre-HIV epidemic periods. The model estimates the bias directly for the five-year period before the survey.

2.3.1.2 THE UN-IGME APPROACH FOR ADJUSTING FOR HIV BIAS

The UN IGME approach uses a cohort component projection model that starts with the latest projection of a national population and its HIV epidemic from UNAIDS (Walker, Hill and Zhao, 2012). HIV/AIDS projections are made using the Spectrum Estimation and Projection Package (EPP) (Walker, Hill and Zhao, 2012). Assumptions are made about the survival times for infected adults with or without treatment and age- and sex-specific prevalence rates so as to produce an estimate of the complete course of the epidemic from its start to the current date (Walker, Hill and Zhao, 2012).

The annual number of births, the annual number of women in need of PMTCT (taken to be the number of births occurring to HIV-positive mothers) and the number of HIV-positive infants are taken directly from Spectrum. Births are then divided into three categories: HIV-negative births occurring to HIV-negative women, HIV-negative births occurring to HIV-positive women and HIV-positive births occurring to HIV-positive women.

Deaths under the age of five are then calculated for births in each year and each category. The probability of dying before the age of five for HIV-negative children is estimated using a model life table, assuming that the risks of dying for HIV-negative children is the same for children born to HIV-positive mothers and children born to HIV-negative mothers (Walker, Hill and Zhao, 2012). The method assumes that all births to HIV-positive women occur four years after infection. The proportion of HIV-positive mothers dying before the time of the survey and the proportion of their births are estimated using a survival curve from four years after infection (Walker, Hill and Zhao, 2012).

This series of steps produce the true and reported births, and the under-five deaths for each of the five-year periods before the survey. The bias due to HIV/AIDS is then estimated by subtracting from one the reported ratio of under-five deaths to births, PD_i^r, to the corresponding true ratio, PD_i^t, as follows

$$B_i = 1 - \frac{PD_i^r}{PD_i^t}$$

where B_i is the bias in five year period *i* before the survey. Child mortality estimates are then adjusted by dividing the estimates from full birth history data by $1 - B_i$.

2.4 CONCLUSION

The lack of reliable up-to-date data poses a challenge for monitoring childhood mortality in South Africa, resulting in estimates being produced using models based on earlier empirical data (Bradshaw and Dorrington, 2007), a situation which is not desirable (Kerber, Tuaone-Nkhasi, Dorrington *et al.*, 2012). Childhood mortality rates in South

Africa also vary with each model (Figure 2.2). Differences in both the level and trajectory of the trends, as highlighted earlier, are much too significant to be ignored as trivial.

However, the problems with estimation of childhood mortality rates cannot be entirely attributed to lack of reliable and timely data and the quality of such data. The methods used in estimating these rates also have limitations. It is impossible to draw conclusions about the short term fluctuations in the mortality estimates produced from the summary birth histories since no information is available about specific dates or ages in the data (Hill, 2012b). The best summary birth histories can offer is a broad indication of an average trend over the past. Apart from HIV/AIDS, other potential biases in the methods arise from high migration rates. High in-migration rates results in women reporting survival of children born and raised elsewhere, while high out-migration will remove responses about children who were born and raised in the community being surveyed.

Full birth history data is generally regarded as producing the best estimates, particularly in the most recent past (Adetunji, 1996). Estimates from these data become less reliable as one moves further back in time, due to recall errors.

However, both the direct and indirect methods provide measures that can closely monitor the progress in childhood mortality rates, assuming that data are accurate. The literature reviewed in this chapter shows that significant progress have been made in trying to deal with the bias due to HIV/AIDS in estimates produced by both direct and indirect methods, thus giving us confidence in our ability to produce reasonable estimates that are close to the true level of mortality given good quality data.

3 DATA AND METHODS

The 2011 Census and the 2012 HSRC survey data are used in this research to assess the levels and trends of childhood mortality in South Africa. Direct and indirect demographic techniques are used to estimate levels and trends of childhood mortality. This chapter first describes and assesses the quality of the data used and then describes how the demographic methods were applied to derive estimates of the level and trend of childhood mortality.

3.1 THE 2012 HSRC NATIONAL HIV, BEHAVIOUR AND HEALTH SURVEY DATA

The 2012 National HIV, Behaviour and Health Survey, conducted by HSRC, is the fourth of its kind, with other HSRC surveys having being conducted in 2002, 2005 and 2008. The survey had a sample of 15,000 households across the country. The 2012 survey is different from previous HSRC surveys in that it was expanded to include information on demographic and health indicators, including fertility, morbidity and mortality with a goal to estimate infant, under-five and maternal mortality rates (Human Sciences Research Council, 2012). As such, the survey included questions similar to those from Demographic and Health Surveys which ask each woman in the household aged 15 and above to report on their full birth history data. The full birth history data are used for direct estimation of infant and under-five mortality rates. The next sub-section briefly assess the quality of the full birth history data collected in the 2012 National HIV, Behaviour and Health Survey.

3.1.1 ASSESSMENT OF THE FULL BIRTH HISTORY DATA

Before cleaning the data, there were 22,617,296 pregnancies after weighting the data that were recorded among women aged 15-49. Of these weighted pregnancies, 20,907,032 were reported as live-births with the rest being reported as either still-births or miscarriages. Among the reported live-births, 1.3 per cent had their vital status missing. Among those reported to be dead, 23.4 per cent had their age at death missing.

As these data exhibited quite a number of inconsistencies and irregularities, they were 'corrected' using logical imputations. The irregularities and inconsistencies that were found in these data as well as the logical imputations applied are described in detail in Appendix A.

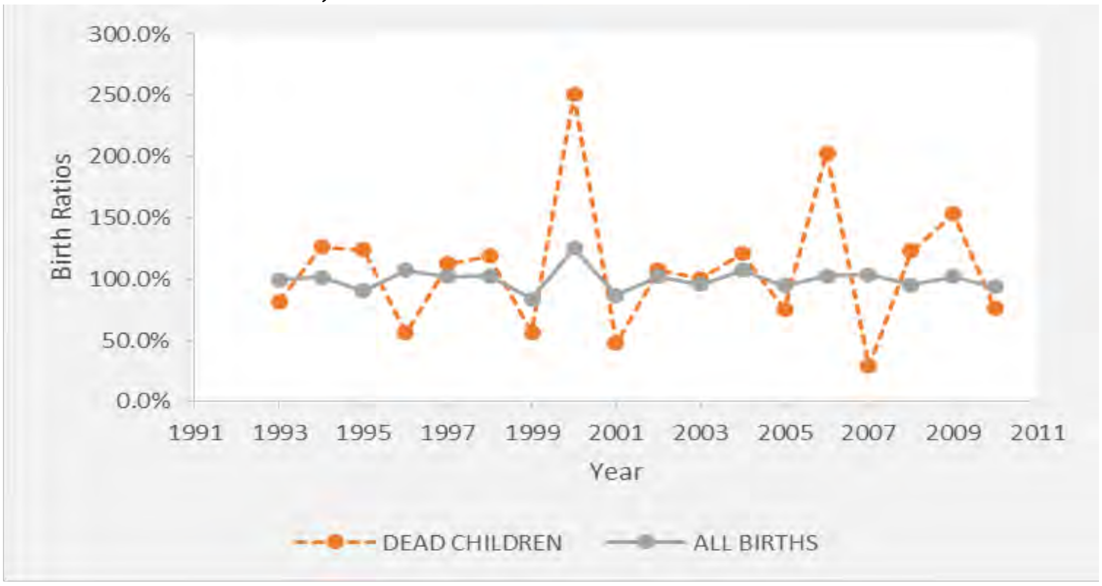
After cleaning the data, there were 23,238,228 pregnancies after weighting the data that were recorded among women aged 15-49. Of these weighted pregnancies, 21,602,306 were reported as live-births with the rest being reported as either still-births or miscarriages (**Error! Reference source not found.**). Among the reported live-births, 4.4 er cent had their vital status missing. Among those reported to be dead, 23.3 per cent had their age at death missing.

Table 3.1 Full birth history data for women aged 15-49 after applying logical edits, HSRC 2012

Category	Number of cases	Weighted Number
Total women	11,668	14,576,986.7
Pregnancies		
Total	16,142	23,238,228
Miscarriages	796	1,093,477.88
Still-births (Born Dead)	308	358,006.66
Live-births (Born Alive)	14,861	21,602,306
Alive	*13,579*	*19,805,668*
Dead	*582*	*844,080.19*
Survival Status Missing	*700*	*952,557.81*

To investigate the possibility of birth transference in the 2012 HSRC survey, birth ratios were calculated by calendar year, two-year and five-year periods. Birth ratios for dead children are significantly different from 100% for the period 1992 to 2012. Notable heaping of births among those dying is seen in years 1995, 2000, 2003, 2006, 2008 and 2009 with the peak in year 2000 (Figure 3.1). However, such fluctuations may be largely due to random fluctuations since few deaths were recorded in the survey. It can also be noted from Figure 3.1 that there is an indication of possible birth transference from the second year of each two-year period to the first year of the adjacent two-year period and this is also true for five-year periods. For example, the birth ratio for children surviving for the second year of the first two-year period before the survey is significantly lower than 100 per cent while that of the first year of the second two-year period before the survey is significantly above 100 per cent. This is also noticed for the second year of the second two-year period before the survey and the first year of the third two-year period before the survey. Birth ratios by five year periods before the survey also indicated a possible transference of births from the period 0-4 years before the survey for both surviving and dead children, with much of the transference being for those who died. However, since there is no motivation for interviewers to transfer births out of a given period, the fluctuations in the birth ratios observed in Figure 3.1 can be explained by simple random fluctuations rather than birth transference.

Figure 3.1 Birth ratios by calendar-years for births and children dead, born between 1992 and 2012, HSRC 2012.



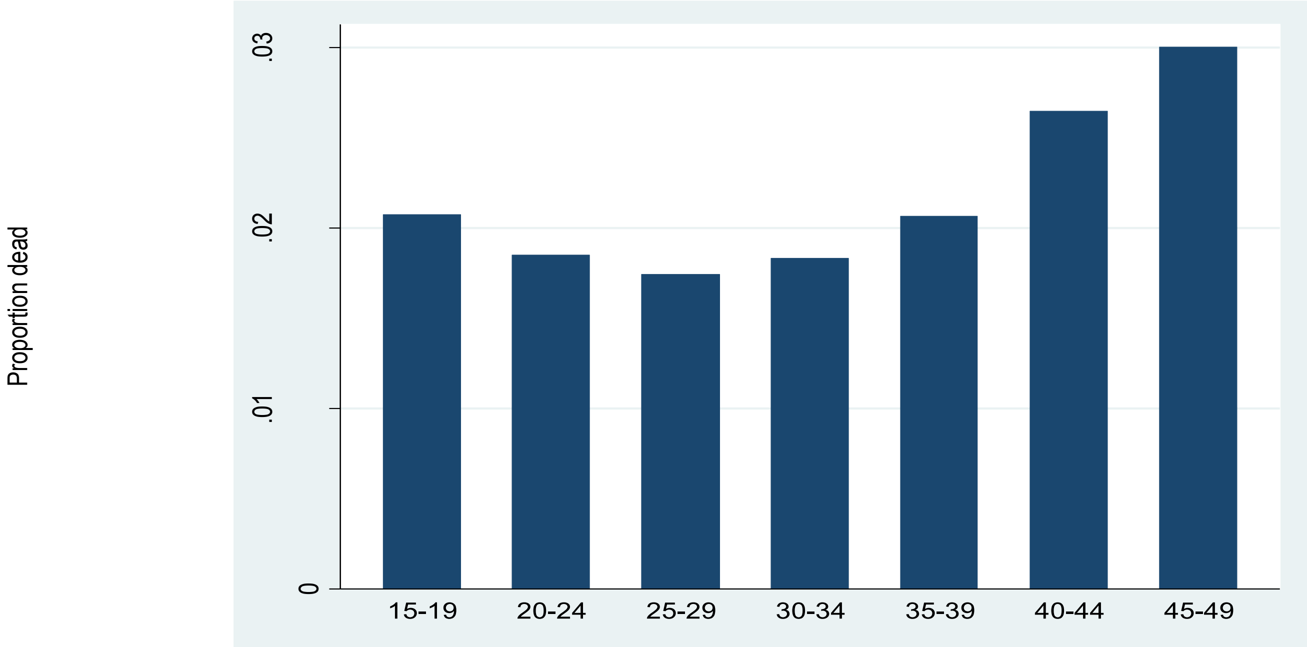
3.2 THE 2011 CENSUS DATA

The 2011 Census is the third census in post-apartheid South Africa and was conducted from the 10th to the 31st of October 2011, with the census night being the night of 9-10 October 2011. Section G of 2011 Census questionnaire A collected summary birth histories of women aged 12-50 by asking the respondent to report, for each woman in a household aged between 12 and 50, the number of children ever born alive (CEB), the number of children still surviving and living in the household (CS^h), the number of children still surviving and living elsewhere (CS^e), the number of children dead (CD), the date of her most recent birth and the survival status of this birth, all by sex of child. Section I of the same questionnaire collected information on deaths in the last 12 months in households including the age and sex of each deceased. The children surviving data were not released to the public owing to poor quality of these data (Statistics South Africa, 2014b). The following sub-sections assess the quality of the PBT and deaths reported by households data collected in the 2011 Census.

3.2.1 ASSESSMENT OF THE PBT DATA

The 2011 Census collected information about the survival of last births among women aged 12-50. There were 2,028,014 (weighted 10 per cent sample) most recent births reported to have occurred in the 24 months preceding the 2011 Census. The proportion dead among last born children born 24 months before the census was found to be 0.0191 for the country as a whole. The proportion of children dead among women aged 15-19 is higher than the proportions of children dead among women aged 20-24, 25-29 and 30-34 (Figure 3.2) as might be expected.

Figure 3.2 Proportion dead among the most recent births occurring in the 24 months prior to the 2011 Census by age group of mother



3.2.2 ASSESSMENT OF THE HOUSEHOLD DEATH DATA

Households reported on deaths that occurred in the 12 months before the census by the sex and age of the deceased. Among the total reported deaths, an overall 2.57 per cent have unspecified age at death and 0.1 per cent of all deaths reported had both sex and age unspecified. The proportion of deaths with missing ages is less than 2.57 per cent in all population groups except for Africans, which had 2.71 per cent with unspecified ages. For provinces, KwaZulu-Natal (KZN) has the highest proportion of deaths with missing ages (3.66 per cent) while Western Cape (WC) has the lowest proportion of deaths with missing ages (0.48 per cent). The proportion of deaths with both sex and age unspecified is consistently less than one per cent across all population groups and all provinces. Also, the proportions of deaths with either sex or age unspecified is less than five per cent in all population groups and all provinces. The details of the percentages of deaths with missing age, sex or both age and sex for each population group and each province as well as nationally are shown in Appendix B.

Apportioning deaths with missing ages to other ages increase the numbers of deaths at those ages but the scaling makes very little difference to the proportions if distributed to all ages. Though there is no consensus as to whether or not deaths with missing ages should be apportioned to ages less than 15 or 20, relative to the 2011 vital registration deaths, as shown by Figure 3.3 and Figure 3.4, the comparison of the 2011 Census and the 2011 vital registration seems to suggest that deaths with missing ages could be concentrated at adult ages. Thus, in this research, deaths with missing ages will be assumed to be aged above the age of five.

Figure 3.3 Age distribution of male deaths – registered and those reported by households.

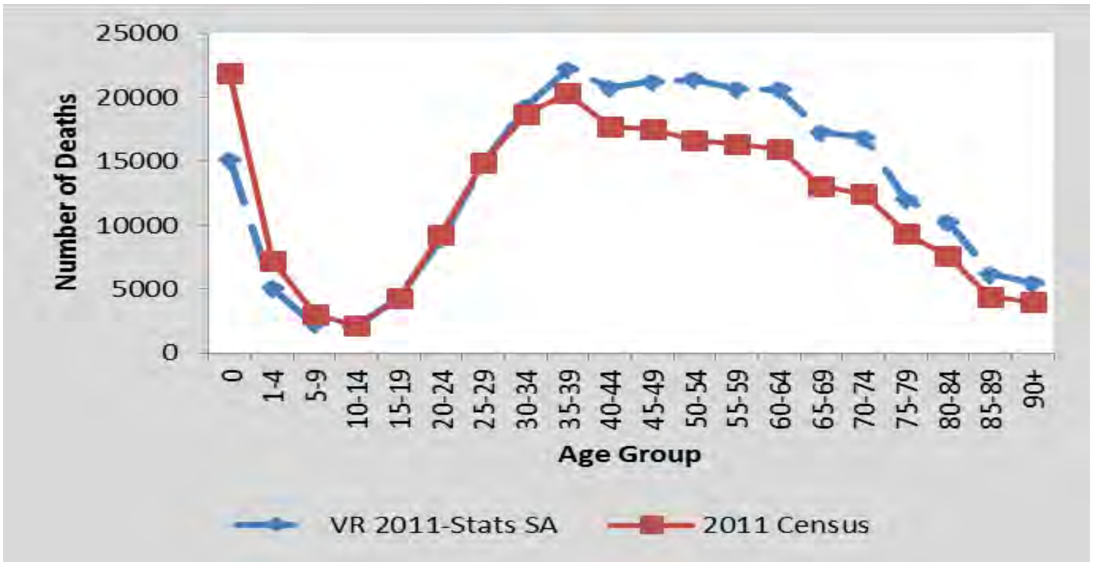
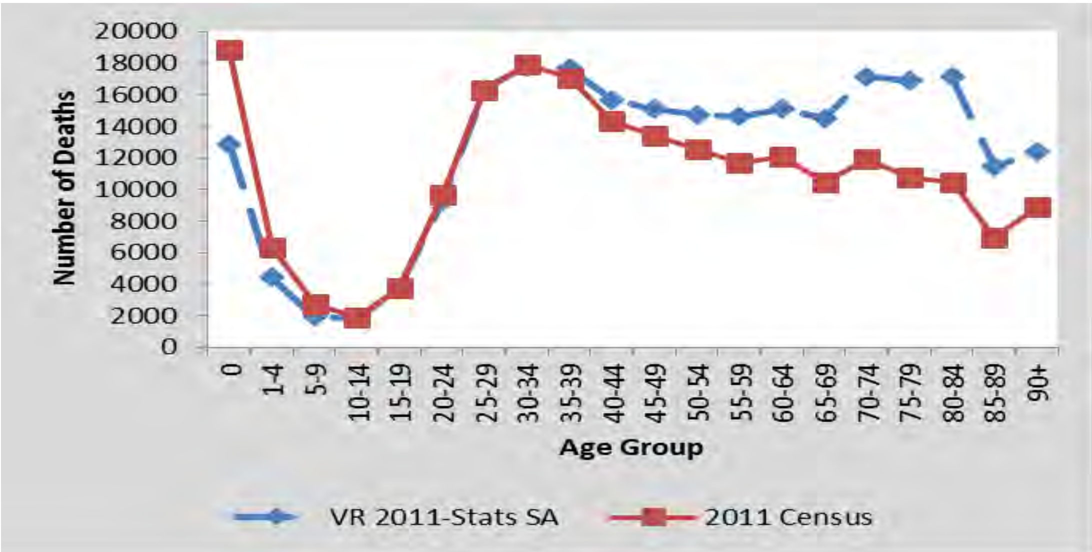


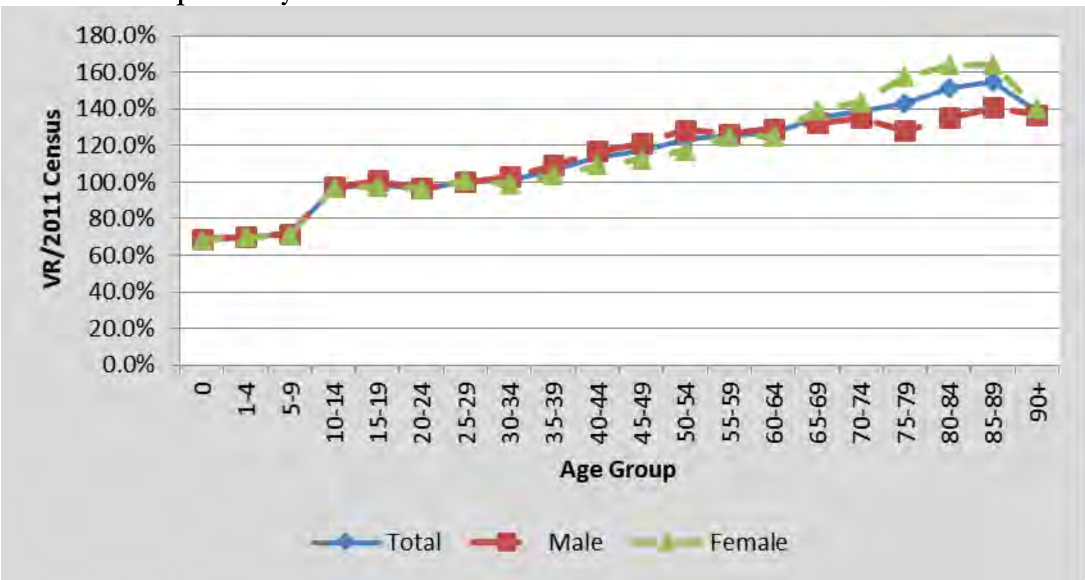
Figure 3.4 Age distribution of female deaths – registered and those reported by households.



The distribution of the number of deaths by age at death is also examined and this is compared to the distribution of deaths by age at death from the vital registration system for 2011 as shown in Figure 3.3 and Figure 3.4. The vital registration data appear to show a shortage of deaths under the age of 10 when compared to the 2011 Census data but this is to be expected since the vital registration data are known to be incomplete at all ages

(Joubert, Rao, Bradshaw *et al.*, 2012). It can also be noticed from Figure 3.3 and Figure 3.4 that the 2011 Census recorded far fewer deaths than the 2011 vital registration deaths for ages 35 and above for males and 40 and above for females. This under-reporting of deaths in the 2011 Census relative to the 2011 vital registration could be the result of household disintegration on the death of a person in this age range, which is not uncommon at these ages.

Figure 3.5 The age-specific ratios of 2011 vital registration deaths to the deaths reported by households in the 2011 Census



The age-specific ratios of the 2011 vital registration deaths to the deaths reported by households in the 2011 Census is also examined. This implies a completeness of the 2011 vital registration of deaths relative to the deaths reported by households in the 2011 Census that is between 68 per cent and 100 per cent for ages less than 35 for males and 40 for females (Figure 3.5). For infants, the implied completeness of the 2011 vital registration deaths relative to the household deaths reported in the 2011 Census is lower than the 88 per cent estimated for 2006 by Darikwa (2009) suggesting an unexpected fall in the completeness of registration of infant deaths. This might be a result of mis-editing of the 2001 Census data to produce too many deaths. For children aged 1-4 the implied completeness of the 2011 vital registration deaths relative to the household deaths reported in the 2011 Census is higher than the 60 per cent estimated for 2006 by Darikwa (2009) suggesting an improvement in the registration of child (1-4) deaths which is quite possible.

Thus, relative to the Community Survey, the household deaths reported in the 2011 Census appear to be missing infant deaths, but not child (1-4) deaths, when the implied completeness of the 2011 vital registration deaths relative to the deaths reported by households in the 2011 Census is compared.

Further investigation of under-five deaths reported by households is carried out by calculating ratios of deaths under age one last birthday to deaths under age five last birthday, for each province (Table 3.2) and for each population group and nationally (Table 3.3). The ratios of deaths under age one last birthday to deaths under age five last birthday from the ASSA2008 population projection model (Actuarial Society of South Africa, 2011) and from the 2001 Census are used for comparison with the ratios from the household deaths reported in the 2011 Census. It can be assumed that the ASSA2008 population projection model captures the mortality profile of South Africa reasonably well in this period as the model is calibrated to reproduce the observed prevalence rates and allows for the best estimates of mortality of infected children. In addition the results of the model are broadly consistent with those from other researchers (UN-IGME, 2013; Wang, Liddell, Coates *et al.*, 2014).

Thus, the ASSA2008 population projection model is a reasonable benchmark to use for comparison with the 2011 Census ratios. The ratios from ASSA2008 reveal that if the level of under-five mortality is increasing, the ratio of deaths under age one last birthday to deaths under age five last birthday is expected to decrease and vice versa. As such, the ratios derived from the household deaths reported in the 2011 Census are expected to be higher than the ratios derived using the 2001 Census since childhood mortality has fallen in the recent past, in particular between 2009 and 2011 (Dorrington, Bradshaw and Laubscher, 2014) and the level of mortality in 2011 would be expected to be lower than in 2001 due to the increased rollout of free antiretroviral therapy which eased access to treatment and thereby increased the life expectancy of those infected by HIV/AIDS.

Table 3.2 The ratios of deaths under age one to deaths under age five by province, both sexes combined

	WC	EC	NC	FS	KZN	NW	GT	MP	LP
Census 2011	0.75	0.81	0.73	0.74	0.76	0.75	0.76	0.75	0.74
Census 2001	0.58	0.62	0.72	0.71	0.65	0.69	0.66	0.70	0.62
ASSA2008 (2011)	0.76	0.76	0.74	0.75	0.73	0.71	0.69	0.73	0.75

It can be noted from Table 3.2 that the ratios of deaths under age one last birthday to deaths under age five last birthday from the 2011 Census are not significantly different from those derived from ASSA2008 population projection model in all provinces except

Gauteng (GT) and Eastern Cape (EC) which have ratios that are 0.07 and 0.05 higher in the 2011 Census than in the ASSA2008 model respectively. Although similar ratios are observed from ASSA2008 population projection model and the 2011 Census in the provinces, there is no high correlation between the ratios from ASS2008 population projection model and the 2011 Census because there is not a lot of differences between the levels of mortality and hence ratios by province and the rates for the provinces are probably not accurate enough to capture these subtleties.

Ratios by population groups for 2011 Census and ASSA2008 are also compared and these are generally found to be similar except for the White population group which had a higher ratio in the 2011 Census than ASSA2008 by 0.07 suggesting that the 2011 Census under recorded deaths at ages 1-4 relative to those under age one.

However, the ASSA2008 model may not be entirely correct since it is a model based on assumptions and changing these assumptions can result in different ratios than presented here. For example, changing the assumptions about treatment rollout and the impact of interventions in the ASSA2008 model can result in different ratios since the number of children assumed to survive to age one and age five depends on these assumptions. Thus, the ratios observed from the 2011 Census cannot be regarded to be entirely wrong based on the ratios derived from the ASSA2008 model. As such, the ratios derived from the ASSA2008 model are only used for 'illustrative purposes' to get an indication of how well the ratios observed from the 2011 Census compare with those derived from ASSA2008.

Table 3.3 Ratios of deaths under age one to deaths under age five by population group, both sexes combined

	Total	African	Coloured	Indian/Asian	White
Census 2011	0.75	0.75	0.80	0.72	0.72
Census 2001	0.66	0.66	0.67	0.63	0.50
ASSA2008 (2011)	0.74	0.73	0.79	0.73	0.65

Ratios of deaths under age one last birthday to deaths under age five last birthday in the 2011 Census are also compared to ratios of deaths under age one last birthday to deaths under age five last birthday in the 2001 Census for provinces, population groups and nationally. The ratios in the 2011 Census are consistently above those in the 2001 Census for all population groups and provinces and in line with what would be expected when mortality is declining. However, there is not much difference between the 2011 Census ratio and the 2001 Census ratio for the Northern Cape (NC) province and Free State. Considering that the level of childhood mortality have fallen between 2001 and

2011 (Dorrington, Bradshaw and Laubscher, 2014; McKerrow and Mulaudzi, 2010), one would expect that the difference between the ratios from the 2011 Census and the ratios from the 2001 Census would be significant as observed in Table 3.2 and Table 3.3. However, the percentage change in the fall in childhood mortality between 2001 and 2011 is too small (McKerrow and Mulaudzi, 2010) to warrant large differences in the ratios between 2001 and 2011.

In addition to the above, the 2001 Census data have their own problems and thus cannot be regarded as being entirely correct. Interrogation of the age at death data for deaths reported by households in the 2001 Census showed that close to eight per cent of the data were imputed using hot decking, giving rise to distributions of the numbers of deaths by age that are inconsistent with the numbers observed for any of the population groups and the numbers resulting from the edited data also "tends to exaggerate deaths below age 40" (Dorrington, Moultrie and Timæus, 2004: 10). However, if the extent of exaggeration in the deaths under age one is the same as the extent of exaggeration of deaths between ages one and four, the errors will cancel out when the ratios of deaths under age one to deaths under age five are calculated.

Also, it is worth noting that the crude death rates derived using the deaths reported by households in the 2011 Census without correcting for incompleteness of death reporting produced the expected ranking of rates by population group, with rates being higher in Africans, followed by Coloureds then Indians and lowest in Whites (Department of Health (South Africa) and Macro International Inc (USA), 2002). The ranking of crude rates by province was also in line with the expected ranking, with rates being lighter in Western Cape, followed by Gauteng, and being heavier in Free State, followed by KwaZulu-Natal (CoMMiC, 2012; McKerrow and Mulaudzi, 2010). Comparison of crude rates depends on the population having similar age-structures across all population groups and across all provinces. For the 2011 Census data, the age-structure was the similar across all population groups and provinces as evidenced by the ratios of the population under one to the population under-five being 0.2 in all population groups and ranging from 0.19 to 0.21 in all provinces.

Taking all of the above into account, the deaths reported by households in the 2011 Census can be used to investigate national childhood mortality as the data did not show any irregularities. However, the estimates for population groups and for provinces need to be interpreted with caution as there is evidence of irregularities in deaths reported by households by age group when the data are disaggregated by provinces or population group, particularly for Gauteng, Eastern Cape, Northern Cape and the White and Indian/Asian population groups.

3.3 METHODS FOR ESTIMATING INFANT AND UNDER-FIVE MORTALITY

This research applied three different methods of estimating infant and under-five mortality rates using the survival of most recent birth and deaths reported by households from the 2011 Census and full birth history data from the 2012 HSRC survey. The direct estimates from full birth history data are corrected for HIV/AIDS bias to give estimates of childhood mortality that are closer to the true measures of infant and under-five mortality. In the sub-sections that follow, the previous birth technique and direct techniques used are described as they are applied in this research.

3.3.1 THE PREVIOUS BIRTH TECHNIQUE (PBT)

The previous birth technique was applied to the 2011 Census data on the survival of the last born child. The method is fairly straight forward to apply. However, the method is argued to produce biased estimates of infant mortality due to the proportion dead among most recent births being lower than the proportion dead among all births (Hill, 2012c). The possible magnitude of this bias in the South African context is estimated in the following sub-section.

The proportions of children who have died, by sex of child, among the most recent births occurring in the 24 months before the survey or census were calculated as the ratio of the number of children who have died among the most recent births occurring in the 24 months before the census or survey to the number of most recent births occurring in the 24 months before the census or survey. The Blacker and Brass (2005) multiplying factor of 1.09 was used to convert the proportion of children who have died among the most recent births occurring in the 24 months before the census or survey into an estimate of the probability of dying before the age of one, $q(1)$. The Princeton West level 19 model life tables were then used as standards in estimating the probability of dying before the age of five, $q(5)$, from the probability of dying before age one, $q(1)$, using the Brass' relational logit system. The estimates were assumed to apply 12 months before the census or survey.

3.3.1.1 ESTIMATION OF BIAS DUE TO USING THE MOST RECENT BIRTHS AND DEATHS

The Blacker and Brass (2005) method for estimating IMR using survival of the most recent births in 24 months before the census/survey assumes that the most recent births occurring in 24 months before the census/survey approximate all the births occurring in the 24 months before the survey. However, according to Hill (2012c), the proportion of children dead among most recent births occurring in the 24 months before the survey is said to be biased downwards relative to the proportion dead among all births occurring in the 24 months before the survey by 20 per cent on average.

Information on the full birth history data of each woman allows one to estimate all births and the deaths arising from them as well as the most recent births and the deaths arising from them occurring in the 24 months before the survey. Comparing the proportions dead among the most recent birth and among all births, one can then proceed to estimate the selection bias of the proportion dead among most recent births relative to the proportion dead among all births. The 1998 and 2003 SADHSs and the 2012 HSRC Survey were used to investigate this bias for South Africa.

Births and deaths for the 10 years before the survey are extracted in five 24-month intervals using the STATA code shown in Appendix C. The exact dates of birth, death and interview are imputed using random numbers as proposed by Hill (2012a). These are expressed in century month code (CMC). Century months are defined as the number of months since the beginning of the century, calculated as $CMC = (Year - 1900)*12 + Month + (Day/30)$. The period between the date of birth (in century months) and the interview date (in century months) gives the number of months before the survey when the birth occurred. Similarly, for deaths, the period between the imputed date of death (in century months) and the interview date (in century months) gives the number of months before the survey when the death occurred. Extracting all births and deaths in the five 24-month periods becomes fairly straightforward. However, to extract the most recent births in a 24-month period before the survey, one needs to create a binary variable that tags a woman's birth which occurred in a relevant 24-month period, and use this to extract most recent births and the deaths arising from them in each 24 month period for the past 10 years. Proportions of children dead among all births occurring in each 24-month period were calculated and compared to the proportions of children dead among most recent births in each 24-month period for the 1998 SADHS, 2003 SADHS and the 2012 HSRC survey. The biases in the proportion dead among most recent births relative to the proportion dead among all births for the three surveys are shown in Table 3.4.

Table 3.4 Estimated selection bias in the proportion dead among most recent births relative to the proportion dead among all births in each 24-month period

PERIOD	1998 SADHS	2003 SADHS	2012 HSRC SURVEY
0-23	9.6%	8.6%	27.5%
24-47	11.8%	3.5%	4.5%
48-71	11.5%	11.9%	-1.6%
72-95	6.7%	13.5%	5.1%
96-119	12.3%	5.8%	39.1%

The proportion of children dead among most recent births occurring in the 24 months before the survey is biased downwards relative to the proportion dead among all births by circa 10 per cent for the 1998 and 2003 SADHSs (Table 3.4). The bias in the proportion dead among most recent births relative to all births in the first 24-month interval for South Africa is significantly lower than the average of 20 per cent suggested by Hill (2012c) probably due to long inter-births intervals in South Africa (Moultrie, Sayi and Timæus, 2012). The 1998 SADHS showed that 0.56 per cent of all births in the 0 to 23 months interval followed a death that occurred in that interval. This proportion is less than the average of 2.5 per cent in all the five 24-month intervals before the survey. Thus, the bias in the proportion dead among the most recent births relative to all births in a 24 month interval is not likely to be due to the omission of a death that preceded a most recent birth in the 24 month interval. Also, as one moves further away from the survey, the data become more affected by recall bias and as well as missing mothers who would have died due to HIV/AIDS before being able to report on the mortality of their children. This has an effect of biasing the mortality estimates downward, particularly for children born to HIV positive mothers more than five years before the survey as most of the mothers of these children, assuming no treatment, would have died to report on the survival of their children. However, 1998 was early in the HIV epidemic of South Africa and thus the direct estimates of childhood mortality are less likely to be biased downwards as HIV prevalence was still low, especially for more distant periods. The larger bias in the proportion dead among most recent births relative to the proportion dead among all births found in the 2012 HSRC survey (bias of 27.5 per cent) is difficult to explain in the context of the 1998 and 2003 SADHSs, as well as in the evidence from research on birth intervals by Moultrie, Sayi and Timæus (2012), although it could be due to deficiencies in the 2012 HSRC survey - too few deaths were recorded and among these a larger proportion (23.4%) had their ages at death missing.

The Blacker and Brass (2005) estimates of IMR were also calculated for each 24-month period using a correction factor of 1.09. Direct estimates of IMR for each 24-month period were calculated using the method by Hill (2012a) which has been described in the literature review chapter. The STATA code used to calculate direct estimates of IMR is shown in Appendix D. Direct estimates and the Blacker and Brass (2005) estimates are compared to establish the bias in IMR estimated from most recent births relative to IMR estimated directly (Table 3.5). Comparison of the Blacker and Brass (2005) estimates of IMR and direct estimates of IMR produced using the method by Hill (2012a) revealed that the bias in estimates produced from survival of recent births generally increase in magnitude for the five 24 month periods as one moves further back in time for the 2003 SADHS and fluctuate for the 1998 SADHS (Table 3.5). However, a bias in the opposite direction than expected is observed with the 2003 SADHS data. This is probably due to the data quality problems inherent in the 2003 SADHS data (Department of Health (South Africa) and Macro International Inc (USA), 2007). The South African department of health documented that more than 21 per cent of deaths among children born within 15 years before the survey date had no age at death reported. This has the effect of biasing downwards the mortality estimates produced using the method by Hill (2012a).

The biases for the first three 24 month periods for the 1998 and 2003 SADHSs are significantly lower than what has been ascertained by Hill (2012c) while those for the last two 24 month periods are consistent with what has been ascertained by Hill (2012c). However, for the 2012 HSRC, the biases fluctuate between positive and negative ranging from as high as 64 per cent to as low as -21 per cent (Table 3.5). The biases observed in the 2012 HSRC survey are quite implausible and may be attributed to possible birth transference and random fluctuations due to few deaths being reported and a large proportion of these missing age at death.

Table 3.5 Bias in IMR estimates from survival of most recent births relative to direct estimates of IMR.

PERIOD	1998 SADHS	2003 SADHS	2012 HSRC SURVEY
0-23	5.0%	-8.6%	38.5%
24-47	7.0%	3.5%	-10.1%
48-71	5.3%	11.9%	39.8%
72-95	24.3%	13.5%	-21.0%
96-119	21.8%	5.8%	64.5%

In conclusion, it is quite difficult to establish the magnitude of the selection bias in the proportion dead among the most recent births relative to the proportion dead among all births. The theory from the 1998 and 2003 SADHSs combined with analysis of birth intervals in South Africa by Moultrie, Sayi and Timæus (2012) suggests that this selection bias is lower than the average of 20 per cent observed in other studies. However, data from the HSRC survey seem to suggest otherwise, with the bias in the proportion dead among the most recent births relative to the proportion dead among all births being 27.5 per cent. Although this bias is consistent with what has been observed by other researchers, the data cannot be trusted as the number of deaths recorded in this survey are too few for a national survey and a large proportion of these have their age at death missing, thus these results might be due to random fluctuations. The data quality issues found in the 2012 HSRC survey are more or less the same as those found in the 2003 SADHS thus one cannot trust the estimates of the bias of the proportion dead among the most recent births relative to the proportion dead among all births that is derived from the 2003 SADHS. On the other hand the 1998 SADHS is too outdated to use its estimate of the bias in the proportion dead among most recent births relative to the proportion dead among all births. There is nothing known about how this bias changes over time or whether it remains constant. As such, in this research, an average of the three estimates of this bias derived from the 1998, 2003 SADHSs and the 2012 HSRC survey will be used to adjust the estimate of the proportion dead among most recent births derived from the 2011 Census. The average bias is estimated to be 15.2 per cent and is approximately consistent with what has been observed in other countries with long birth intervals (Hill, 2012c).

3.3.2 DEATHS REPORTED BY HOUSEHOLDS

This method was applied to data from the 2011 Census on deaths occurring in a household in the past 12 months before the survey. The estimates produced by this method are assumed to apply over the year centred around six months before the census night, i.e. they apply over the year centred on 2011.273 (which is 9-10 March 2011). To apply the method, data of the population by age and sex together with data on the deaths in the past 12 months by age and sex were required. Defining the population aged x last birthday at the time of the census date, t, as $P_x(t)$, the population aged x one year before the census date as $P_x(t-1)$ and the deaths aged x in the twelve months before the census date as D_x, the population aged x one year before the census, assuming that there is little or no migration at ages less than five, was estimated as

$$P_x(t-1) = P_{x+1}(t) + 0.5 * D_x^{'},\ for\ x = 1, 2, 3, 4.$$

where $D_x^{'} = D_x + D_{x+1}$ approximates the deaths during the year aged x last birthday at time $t-1$.

There are three ways in which the population exposed to risk of dying before age one in the 12 months before the census could be estimated from the census data. One way is to use the number of the births occurring in the 12 months before the census as reported by mothers when responding to the question on the survival of most recent births in the census. This approach uses answers to different questions for births and deaths which could be inconsistent for various reasons including multiple births to one woman in a period of 12 months. However, the proportion of multiple births in 12 months is usually small thus, most recent births in the 12 months before the census could be a reasonable approximation of all births occurring in the 12 months before the census. Another way is to back-project the population aged zero at the time of the census using $\frac{L_0}{l_0}$ estimated from a suitable source, for example ASSA2008. However, this approach uses an estimate of mortality that may be inconsistent with the rates being estimated, thus it could also produce a biased estimate. The third way is to estimate births in the 12 months preceding the census date as

$$B = P_0(t) + (0.75 \times D_0)$$

In this approach, 0.75 is a crude approximation that 75 per cent of deaths at age zero were born in the 12 months before the census (the other 25% having been born the year before that). The 0.75 approximation is chosen since, as a rule of thumb, the mean number of person years lived by those dying before the age of one in a single year is 0.3 years thus, taking into account the effect of HIV/AIDS on mortality, 0.75 is a reasonable approximation that 75 per cent of deaths at age zero were born in the 12 months before the census. This approach was adopted in this research as it uses the same data for which rates are being estimated from.

The population aged $x = 1, 2, 3, 4,$ exposed to the risk of dying in the twelve months before the census, $P_x(t-0.5)$, is calculated as

$$P_x(t-0.5) = \sqrt{P_x(t-1) \times P_{x+1}(t)}$$

where $P_x(t-1)$ is the population aged x last birthday at time $t-1$ and $P_{x+1}(t)$ is the population aged $x+1$ last birthday at time t. The central mortality rate between ages x and $x+1$ for $x = 1,2,3,4$, M_x, is then calculated as

$$M_x = \frac{D_x}{P_x(t-0.5)}$$

For age $x = 0$, the infant mortality rate (IMR) was estimated as the deaths aged zero in the past 12 months before the census, D_0, divided by the births in the 12 months before the census, B. It was assumed that this is approximately the same as the probability of dying before the age of one, q_0.

The central mortality rates, M_x, for $x = 1,2,3,4$, were then converted into probabilities of dying between ages x and $x+1$, q_x, for $x = 1,2,3,4$, using the following formula

$$q_x = \frac{M_x}{1+0.5*M_x}$$

The person-years lived by those dying between ages x and $x+1$ were assumed to be 0.5 years for ages $x = 1,2,3,4$. The probability of dying before the age of five, ${}_5q_0$, was then calculated as

$${}_5q_0 = 1 - \prod_{x=0}^{4}(1-q_x)$$

Infant and under-five mortality rates from deaths reported by households were estimated with and without adjustment for deaths with missing ages since there is no consensus as to what proportion of those missing age are children. There was little difference between estimates with and without adjustment for deaths with missing ages thus estimates produced without adjusting for deaths with missing ages will be reported, effectively assuming that all deaths with missing ages are above the age of four.

3.3.3 FULL BIRTH HISTORY DATA

This method is applied to the 2012 HSRC survey. It requires the month and year of birth and the age at death of the child. The date of birth is expressed in century months and the age at death, which is reported in days for neonatal infants, in months for children 1-23 months old and in years for children two years old and above, is expressed in months by converting those reported in days or years to months. With the date of birth and age at death, one can then calculate the date of death by adding the date of birth and age at death. The exact date of birth is imputed using random numbers generated using the variable for enumeration area and the day of interview. These two variables are assumed not to be correlated with one another and are both not related in any way to the mortality of children.

Table 3.6 Derivation of exposure to risk of dying

Defining rules	Exposure for survivors	Exposure for the deceased
Age at the beginning of the period (x_{t1}) is greater than the highest age (x_h).	0	0
Age at the beginning of the period (x_{t1}) is greater than the lowest age (x_l) but less than the highest age (x_h). Age at the end of the period (x_{t2}) is less than the highest age (x_h).	$x_h - x_{t1}$	$x_d - x_{t1}$
The lowest age (x_l) is greater than age at the beginning of the period (x_{t1}) and age at the end of the period (x_{t2}) is greater than the highest age (x_h)	$x_h - x_l$	$x_d - x_l$
The lowest age (x_l) is greater than age at the beginning of the period (x_{t1}) but age at the end of the period (x_{t2}) is less than the highest age (x_h)	$x_{t1} + (t2 - t1) - x_l$	$x_d - x_l$
Age at the end of the period (x_{t2}) is less than the lowest age (x_l)	0	0

Note: Adapted from Hill (2012a), Table 2. t1 is the beginning of the time period, t2 is the end of the time period.

When the dates of death have been imputed, it becomes straight-forward to locate the year in which a death aged between x and $x+1$ months occurred, for $x = 0,1,2,...,59$. After locating deaths in time, the next step is to calculate the exposure to risk of dying for those surviving and those dying. The calculation of the exposure to the risk of dying is summarized in Table 3.6.

After calculating the exposure to risk of dying for surviving and dead children, the central mortality rate for each month of life is calculated as deaths at age x last month divided by the exposure at age x last month, for all $x = 0,1,2,...,59$ months. The probability of dying at age x months is then calculated as

$$q_x = \frac{\frac{m_x}{12}}{(1+\frac{m_x}{24})}$$

The probability of dying between age zero and y years is calculated as the cumulative product of the probabilities of surviving each month of life until age y years. Appendix D shows the STATA code used to perform the calculations.

4 RESULTS

Estimates of the levels in infant and under-five mortality rates were derived from the 2011 Census and the 2012 HSRC survey. Only the estimates produced using full birth history data were able to produce a time trend in infant and under-five mortality rates. This chapter presents these results. The chapter is divided into three sections. The first section presents estimates of infant and under-five mortality rates from the 2011 Census data and the second section presents estimates of infant and under-five mortality rates from the 2012 HSRC survey. Finally, estimates from the 2011 Census and the 2012 HSRC survey are compared.

4.1 ESTIMATES OF CHILDHOOD MORTALITY FROM 2011 CENSUS

Two methods were used to produce estimates of childhood mortality from the 2011 Census. Estimates from each method are presented in the subsections below. Only point estimates of childhood mortality could be produced using the deaths reported by households and applying the Blacker and Brass (2005) variant of the previous birth technique to data on survival status of the most recent births in the 24 months prior to the census.

4.1.1 ESTIMATES PRODUCED FROM THE PREVIOUS BIRTH TECHNIQUE

The Blacker and Brass (2005) variant of the previous birth technique was applied to the 2011 Census data. There were slightly over two million most recent births in the 24 months before the census and 38,816 children born in the 24 months before the census are estimated to have died, representing a proportion of 0.0191 (Table 4.1). This proportion of children dead was adjusted for the bias in proportion dead among most recent births relative to proportion dead among all births using an estimated bias of 15.2 per cent. After adjusting for the bias in the proportion dead among most recent births relative to the proportion dead among all births, the Blacker and Brass (2005) adjustment factor of 1.09 was used to convert the proportion dead into estimates of infant mortality rates.

Under-five mortality rates that correspond to the infant mortality rates estimated using the previous birth technique were estimated using the ratio of $q(5)/q(1)$ from the Princeton West level 19 model life tables. The Princeton West level 19 model life tables were chosen as they have a similar shape (rates by age) of mortality to that of South Africa for ages 0-20 when compared to life tables from the ASSA2008 population projection model as shown in Figure E 1 presented in appendix E. The ratio of $q(5)/q(1)$ from the Princeton West level 19 model life tables is the same to that observed from the ratio of $q(5)/q(1)$ from the deaths reported by households in the 2011 Census (a ratio of circa 1.3 in all population groups using deaths reported by households). The results are shown in Table 4.1 by sex and Table 4.2 by population group. The estimates are applicable to the 24 months centred on 2010.723.

Table 4.1 IMR and U5MR estimated using the PBT by sex, Census 2011

	Births	Deaths	Unadjusted Estimates			Adjusted Estimates		
			PD24	q(1) per 1000	q(5) per 1000	PD24	q(1) per 1000	q(5) per 1000
Male	1,019,166	21,228	0.0208	23	30	0.0246	27	36
Female	994,606	16,726	0.0168	18	25	0.0198	22	30
Total	2,028,014	38,816	0.0191	21	28	0.0226	25	33

Note: PD24 stands for proportion dead mong most recent births in the 24 months before the census.

The sex ratio at birth was 1.02 and is consistent with what would be expected from populations in sub-Saharan Africa (Hill, 2012b). The sex-specific infant and under-five mortality rates show that male infant and under-five mortality rates are, consistent with expectations, higher than female infant and mortality rates (Table 4.1). However, even after adjustment for the bias in the proportion dead among most recent births relative to the proportion dead among all births as described in section 3.3.1.1, the mortality rates presented in Table 4.1 are low in comparison with estimates from other sources for the same reference period (24 months centred on 2010.723), such as the rapid mortality surveillance report (Dorrington, Bradshaw and Laubscher, 2014) as well as deaths reported by households in the 2011 Census.

Table 4.2 PBT estimates of IMR and U5MR by population group, Census 2011

	Births	Deaths	Unadjusted Estimates			Adjusted Estimates		
			PD24	q(1) per 1000	q(5) per 1000	PD24	q(1) per 1000	q(5) per 1000
African	1,723,184	36,141	0.0210	23	31	0.0247	27	36
Coloured	170,170	2,035	0.0120	13	18	0.0141	15	21
Asian	30,962	185	0.0060	7	9	0.0071	8	10
White	94,796	419	0.0044	5	7	0.0052	6	8

Note: PD24 stands for proportion dead mong most recent births in the 24 months before the census.

Estimates by population group presented in Table 4.2 are ranked as expected, with the highest rates among the Black Africans followed by the Coloured, Asian and White population groups, in that order. Racial differences in childhood mortality still exist between the population groups with the African population group showing an under-five mortality rate that is more than four times that of the White population group. However, the sample for the White population is small and thus estimates disaggregated by population group will be subject to random fluctuation.

4.1.2 ESTIMATES PRODUCED FROM DEATHS REPORTED BY HOUSEHOLDS

Estimates of infant and under-five mortality rates were produced using deaths reported by households. The synthetic cohort life table approach was used in estimating the infant and under-five mortality rates as described in the methods chapter. The results are shown in Table 4.3 by sex and Table 4.5 by population group. These estimates apply for the year centred on 2011.275.

Table 4.3 IMR and U5MR estimated using deaths reported by households by sex

	MALE	FEMALE	TOTAL
IMR	36	32	34
U5MR	49	43	46

Generally, mortality was observed to be higher in males than in females (Table 4.3) as might be expected. When compared to estimates of under-five mortality rates produced from other sources, the estimates of under-five mortality rates produced from deaths reported by households are consistent with estimates of under-five mortality rates produced from ASSA2008 and by UN-IGME but are significantly higher than those observed from the RMS report for the same reference time period (12 months centred around 2011.275) (Table 4.4).

Table 4.4 Comparison of HHD estimates of under-five mortality rates to those from ASSA2008, UN-IGME and RMS 2012

	HHD	ASSA2008	UN-IGME	RMS (2012)
q(5) per 1000	46	49	47	40

Table 4.5 IMR and U5MR estimated using deaths reported by households by population group

	AFRICAN	COLOURED	ASIAN	WHITE
IMR	38	21	12	9
U5MR	50	26	16	13

The estimates produced for population groups from the deaths reported by households in the 2011 census showed the expected ranking with estimates being highest in the African population group and lowest in the White population group (Table 4.5). Racial difference still exists in the infant and under-five mortality rates between the population groups with the African population group showing an under-five mortality rate that is three times more than that of the White population group. When compared to estimates of the population group from the PBT method, the PBT estimates are lower than those from deaths reported by households for all population groups. The estimates for the African and Coloured population groups are consistent with estimates derived from the ASSA2008 full model while those for the Indian/Asian and White population groups are significantly different from those derived from the ASSA2008 full model (Table 4.6).

Table 4.6 Comparison of HHD estimates to ASSA2008 estimates

	HHD Estimates		ASSA2008 Estimates	
	q(1) per 1000	q(5) per 1000	q(1) per 1000	q(5) per 1000
African	38	50	37	54
Coloured	21	26	19	25
Asian	12	16	8	12
White	9	13	5	9

4.2 ESTIMATES OF CHILDHOOD MORTALITY FROM THE 2012 HSRC SURVEY

The 2012 HSRC survey was used to produce estimates of infant and under-five mortality rates for the 15 years before the survey. Estimates were produced using the direct method for the three successive five-year periods prior to the survey and by sex. Estimates for the population groups were not produced as the data were too sparse to be disaggregated by population groups. Indirect estimates using summary birth history were not produced owing to the poor quality of such data. There were internal inconsistencies in the summary birth history data collected in this survey – the sum of children surviving and children dead did not add up to total children ever born. Also, the summary birth history data derived from the full birth history data were inconsistent with the collected summary birth history data. In light of these data quality issues, no further insights can be added by the estimates from the summary birth history data than those already provided by the full birth history data. The following sub-section discusses the results from the 2012 HSRC survey data.

4.2.1 ESTIMATES PRODUCED FROM FULL BIRTH HISTORY DATA

The direct method proposed by Hill (2012a) was applied to the full birth history data from the 2012 HSRC survey. The estimates derived are presented in Table 4.7 by sex and five-year periods, and the under-five mortality rates adjusted for HIV/AIDS bias are presented in Table 4.8.

Table 4.7 Direct estimates of IMR and U5MR from 2012 HSRC survey by sex

PERIOD	REF YEAR	MALE		FEMALE		TOTAL	
		IMR	U5MR	IMR	U5MR	IMR	U5MR
0-4	2010.5	16	21	23	26	19	23
5-9	2005.5	25	30	21	25	23	28
10-14	2000.5	20	25	22	30	22	28
1-5	2009.5	19	23	20	24	19	23
6-10	2004.5	27	33	24	31	25	32
11-15	1999.5	17	22	23	27	20	24

Estimates were produced for three five-year periods in the 0-14 years before the survey and for three five year periods in the 1-15 years before the survey. Typically, the estimates for the 1-15 years before the survey are derived as a way for adjusting for birth displacement common in DHS-type data. However, too few deaths were recorded in the 2012 HSRC survey and a large proportion of these had their ages at death missing (23.4 per cent) thus the discrepancies in the estimates for the 0-14 years and the 1-15 years can be attributed to random fluctuations in the data rather than to birth transference. Thus missed deaths are of more concern than birth transference, although both these issues affect the estimates of mortality produced from these data. The estimates apply for the five year period centred on the reference year given in column two of Table 4.7.

There is an anomaly in both the infant and under-five mortality rates by sex as these are higher in females than males, specifically for periods 0-4, 1-5, 10-14 and 11-15 years. This could be due to misplacement of male deaths. However, the lower male mortality rates can be attributed to male deaths being missed completely, thereby causing random fluctuations in the data. The levels of infant and under-five mortality rates by sex and for both sexes combined are lower than would be expected from other sources. However, apart from the estimates of under-five mortality for all three periods in the 0-14 years before the survey, the general trend in the childhood mortality is consistent with what has been documented in other studies, with both infant mortality and under-five mortality having declined for the most recent five years before the survey when compared to the second five-year period estimates.

Table 4.8 U5MR adjusted for HIV/AIDS bias

PERIOD	REF YEAR	U5MR	CORRECTION FACTOR	CORRECTED U5MR
0-4	2010.5	23	0.9070	25
5-9	2005.5	28	0.7421	37
10-14	2000.5	28	0.7328	38
1-5	2009.5	23	0.8737	27
6-10	2004.5	32	0.7207	44
11-15	1999.5	24	0.7606	32

The estimates of under-five mortality rates were corrected for HIV/AIDS bias using the method by Walker, Hill and Zhao (2012) described in section 2.3.1.2. Annual number of births, number of mothers in need of PMTCT and newly infected infants from Spectrum 5.03 are used as inputs in the spreadsheet by Walker, Hill and Zhao (2012). The mothers in need of PMTCT are assumed to approximate the number of HIV positive women and the new infant infections are assumed to approximate the new infections to new born babies in the year. It can be seen that an adjustment of close to 10 per cent is needed for the 0-4 years before the survey and that for the 1-5 years before the survey is close to 13 per cent (Table 4.8). However, even after adjusting for these biases, the under-five mortality rates from the 2012 HSRC survey were still too low relative to adjusted RMS estimates, which – under these circumstances – can be regarded as the gold standard, and could not be relied upon to give useful insights into the levels of under-five mortality in South Africa.

4.3 COMPARISON OF ALL ESTIMATES PRODUCED

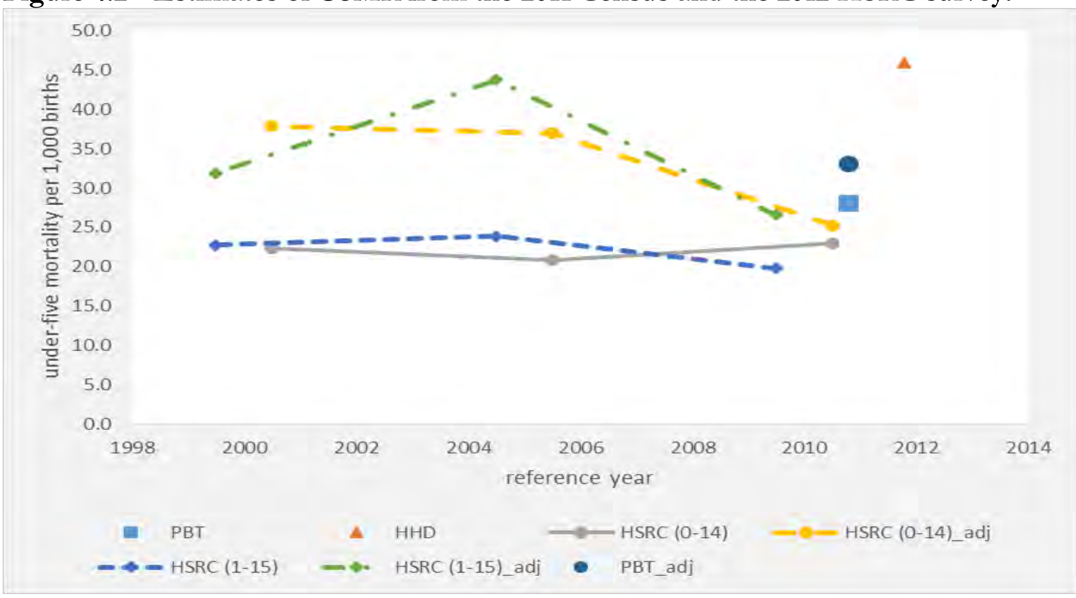
Point estimates from the 2011 Census data were produced using the PBT and deaths reported by households. The 2012 HSRC survey produced estimates for 15 years before the survey by five-year periods corrected for HIV/AIDS bias and thus gave a time trend in the levels of childhood mortality. All these estimates are compared and contrasted in this section. Comparison with estimates from other sources is discussed in Chapter 5.

The estimates of infant and under five mortality rates shown in Figure 4.1 and Figure 4.2 for the 0-4 years before the survey are not far from those produced using data on the survival status of most recent births from the census before adjusting for the bias in the proportion dead among most recent births relative to the proportion dead among all births. However, it should be noted that deaths reported by households produced an estimate that is much higher than those produced using the survival of most recent births, both adjusted and unadjusted, as well as the unadjusted and adjusted estimates for the period 0-4 years and 1-5 years from the 2012 HSRC survey full birth history data. Though it is not certain which one of these is correct, it should be noted that the deaths reported by households produced estimates that are higher than would be expected from other sources such as the Rapid Mortality Surveillance report (Dorrington, Bradshaw and Laubscher, 2014) while those from the previous birth technique and direct estimation from the full birth history data were too low.

Figure 4.1 Estimates of IMR from the 2011 Census and the 2012 HSRC survey

Note: PBT represents previous birth technique, PBT_adj represents adjusted estimates from the previous birth technique, HHD represents household deaths.

Figure 4.2 Estimates of U5MR from the 2011 Census and the 2012 HSRC survey.



Note: PBT represents previous birth technique, PBT_adj represents adjusted estimates from the previous birth technique, HHD represents household deaths.

Even after adjusting for HIV/AIDS bias, the estimates of under-five mortality rate remained lower than estimates produced from the 2011 Census data. Both unadjusted estimates of infant mortality and adjusted estimates of under-five mortality are lower than would be expected from the 2011 Census, suggesting that the 2012 HSRC survey estimates are not useable.

5 DISCUSSION AND CONCLUSIONS

The overall objective of this research was to produce infant and under-five mortality rates by population group for the period 1998 to 2012 with a view to update estimates for new data and to assess the reasonableness of all estimates. The 2011 Census and the 2012 HSRC survey were used to attempt to meet this objective.

Prior to the 2011 Census and the 2012 HSRC survey, the 1996 Census, 1998 SADHS and the 2007 Community Survey were used to produce reasonably reliable estimates of childhood mortality for the post-apartheid South Africa (Darikwa, 2009; Department of Health (South Africa) and Macro International Inc (USA), 2002; Dorrington, Timæus, Moultrie *et al.*, 2004). The 2001 Census could not be used to produce estimates of childhood mortality using the CEB/CS method owing to poor quality of these data particularly after editing and imputation (Dorrington, Moultrie and Timæus, 2004). Also, the 2003 SADHS could not be used to produce reliable estimates of childhood data owing to problematic fieldwork which resulted in poor quality data from this survey (Department of Health (South Africa) and Macro International Inc (USA), 2007).

This chapter is divided into five sections. The first section discusses the quality of the data from the 2011 Census and the 2012 HSRC survey. The second section compares the estimates produced in this research for reasonableness in comparison with estimates produced from other sources. Limitations of this research are discussed in the third section. The fourth section gives suggestions for further research based on the outcome of this research. Conclusions on the usefulness of the data and reasonableness of the estimates produced in this research are drawn at the end of the chapter.

5.1 DATA SOURCES AND DATA QUALITY

Data on the survival of most recent births recorded a number of births for the 24 months preceding the census that is consistent with the number of births expected according to, for example, the ASSA2008 model. Also, the trend in the proportion of children dead among the most recent births in the 24 months preceding the census decreased by five-year age group of the mothers between 15 and 29 before starting to increase by five-year age group until the age of 49, as would be expected. On the other hand, the overall proportion of children dead among most recent births seems to be too low. This is because the most recent births are a subset of all births as some women might have more than one birth in a 24 month period and thus, the proportion dead among most recent births will be biased downwards relative to the proportion dead among all births (Hill, 2012a). This then begs the question of how big is the magnitude of this bias and in this research we attempted to quantify this bias using the 1998 and 2003 SADHSs and the 2012 HSRC survey which enabled us to get an estimate of all births in the 24 months preceding the survey as well as the most recent births. Although the average bias (15.2 per cent) in the proportion dead among most recent births relative to all births used in this research is in line with other estimates of the bias in the proportion dead among most recent births relative to all births observed in populations with long birth intervals (Hill, 2012c), the adjustment factor still gives a lower estimate of the proportion dead. However, the data on the survival of most recent births did not show any irregularities suggesting that the low estimate of the proportion dead could be either a result of the method failing to work or that the bias in the proportion dead among most recent births relative to the proportion dead among all births could be higher (say above 20 per cent). The low proportion dead among most recent births, even after attempting to correct for the bias in proportion dead among most recent births relative to the proportion dead among all births, implies that the estimates produced using the PBT method are not useful.

The 2011 Census data also collected information about deaths that occurred in a household in the past 12 months preceding the census date. These data were interrogated for usefulness in producing estimates of infant and under-five mortality rates. Less than five per cent of these data were missing age at death. Initial results of the 2011 Census released in 2012 had over 20 per cent of deaths missing age at death (Statistics South Africa, 2014a). According to Statistics South Africa (2014a) most of the errors in recording information on deaths reported by households in the census can be attributed to scanning errors during data capturing thus 115, 221 deaths were recaptured manually and most of these were found to be inconsistent with the data that were previously captured. Although only cases with missing ages at death were recaptured, it also raises concerns as to whether cases with ages at death were not erroneously scanned thus, estimates produced from these data cannot be divorced from these data quality issues and need to be interpreted in light of these quality issues.

Less than one per cent had their sex unspecified and generally, when disaggregated by population group, the data were found to exhibit some irregularities suggesting that the data might be missing some deaths aged broadly under-five. Although these data may be suitable for producing national estimates of childhood mortality, mortality estimates by population group should be interpreted with caution, particularly for the White and

Indian/Asian population groups as the data for these population groups are too sparse and thus are likely to produce estimates that are too low.

The 2012 HSRC survey collected full birth history data which were assessed for suitability in producing infant and under-five mortality rates. Inconsistencies and irregularities which were found in these data were corrected using logical imputations as detailed in Appendix A. Even after correcting for the inconsistencies, these data had two major problems, which are that the survey recorded far too few deaths and among the recorded deaths 24 per cent had their age at death unspecified. This has the effect of biasing mortality estimates downwards as more deaths will be missed from the calculations while the births which would have died are included in the exposure to risk of dying. Even assuming that deaths with missing age have the same age pattern as those with age at death specified and thus excluding the deaths with missing age from the analysis, the estimates produced were not significantly different from those produced including deaths with missing ages.

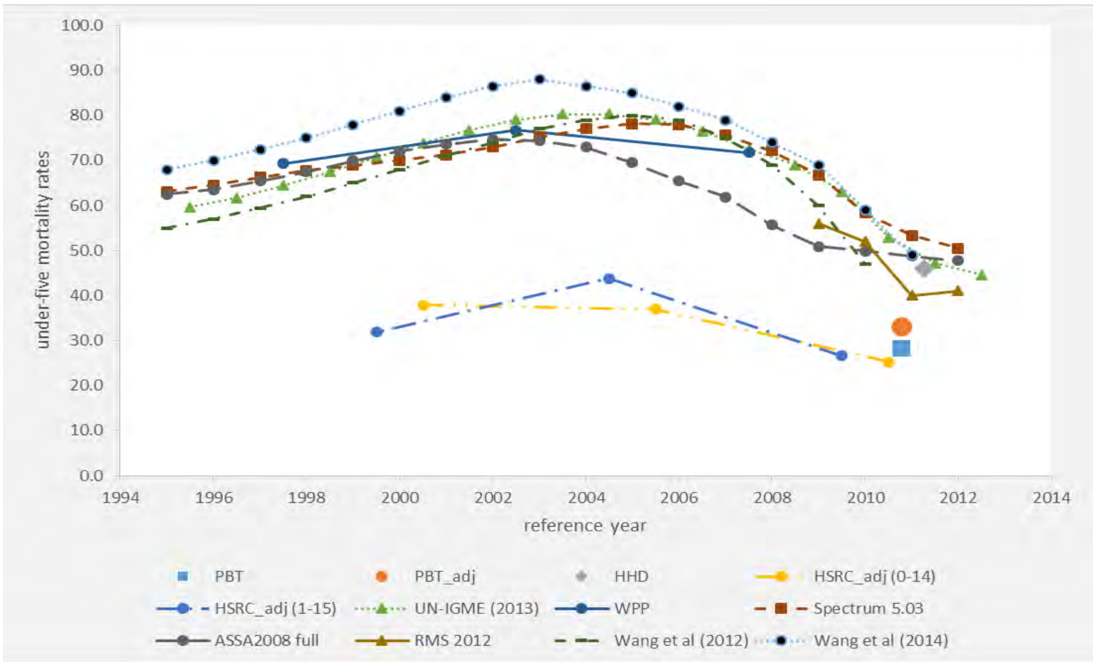
Thus, missed child deaths are a major concern than the deaths reported without ages at death, although these two issues both affect mortality estimation and bias the results downwards. Also, in an attempt to check for any birth transference from one period to another, birth ratios were calculated for calendar, two-year and five-year periods. Although the birth ratios were found to be different from 100 per cent, particularly for dead children, this observation could not be attributed to birth transference rather than to random fluctuations in the data as the survey recorded too few deaths with a large proportion of these missing ages at death. Generally, these data are not good enough to provide reliable estimates and bias mortality estimates downwards due to a large proportion of deaths missing age at death.

5.2 COMPARISON OF ESTIMATES PRODUCED WITH ESTIMATES FROM OTHER SOURCES

Estimates produced from the 2011 Census and the 2012 HSRC survey data are compared with estimates from other sources to see how consistent they are with what has been estimated by other researchers. Figure 5.1 presents the estimates of levels and trends of under-five mortality from seven different researchers including the estimates produced by this research.

Clearly, the estimates of childhood mortality from the 2012 HSRC survey are lower than would be expected from other sources (Figure 5.1). Although the trend shown by the 2012 HSRC survey data are consistent with the trends observed from other sources, the 2012 HSRC survey data is producing estimates that are about half of what is observed from other sources. This is due to the quality of these data which were found to be poor as a large proportion of reported deaths were missing age at death, making the denominator of the mortality estimates large without including such deaths in the numerator.

Figure 5.1 Comparison of under-five mortality estimates from the 2011 Census and 2012 HSRC data with estimates from other studies



Note: PBT_adj is the PBT estimate adjusted for the bias in the proportion dead among most recent births relative to the proportion dead among all births. HSRC_adj estimates are HSRC estimates adjusted for HIV/AIDS bias using the method by Walker, Hill and Zhao (2012).

Also, Figure 5.1 shows that the PBT estimates of under-five mortality are underestimated relative to estimates of others, using different data and methods. This might be expected as the data assessment showed that the survival of most recent birth data produced a low estimate of the proportion dead. However, it should be noted that the magnitude of underestimation in the under-five mortality estimate shown in Figure 5.1 is also dependant, to some extent, on the model life table used to derive the relationship between $q(1)$ and $q(5)$.

The estimate of under-five mortality produced from deaths reported by households in the 2011 Census appear to be consistent with estimates from other sources. Figure 5.1 shows that estimates from the rapid mortality surveillance report (Dorrington, Bradshaw and Laubscher, 2014) are lower than those produced from the 2011 Census data using deaths reported by households. However, the under-five mortality estimate produced using deaths reported by households lie very close to those produced by Wang, Liddell, Coates *et al.* (2014) for IHME using the GBD 2013 study, by UN-IGME (2013) and those produced from the ASSA2008 full model (Actuarial Society of South Africa, 2011).

Generally, one can conclude that estimates produced from the 2012 HSRC survey are not reasonable and these data do not produce reliable estimates of under-five mortality. Also, the estimates from the survival of most recent birth reported in the 2011 Census are not usable as they underestimate relative to what would be expected from other sources. Only the estimates of under-five mortality produced using deaths reported by households in the 2011 Census were found to be reasonable on the basis of the comparison made with estimates from other sources thus, these data can be used to produce estimates of childhood mortality. However, estimates produced by disaggregating the data by population group need to be interpreted with caution, particularly for the Indian/Asian and White population groups as the data for these population group are too sparse.

5.3 LIMITATIONS OF RESEARCH

The 10 per cent sample of the 2011 Census was released in April 2014, about two and half years after the census was conducted, compared with the promised release date of 31 March 2013. This meant that we had to wait for these data to be released as efforts to access these data earlier were fruitless. Apart from the late release of the 10 per cent sample of the 2011 Census, the data excluded the responses to a number of questions, including those on the survival status of children ever born owing to the poor quality of these data (Statistics South Africa, 2014b), making it thus impossible to make use of these crucial data for estimating infant and child mortality. Several requests to be provided with the data on the survival status of children ever born for deeper interrogation, as was done by Darikwa (2009) for the 2007 Community Survey, failed. Thus, the data on survival of most recent births and the deaths reported by households were the only data from the census that could be used to produce point estimates of infant and under-five mortality rates.

The major limitations of using deaths reported by households in a census arises from the incompleteness of death reporting by age and, for infants, possible undercount of births occurring 12 months before the census which then biases the estimate of infant mortality upwards. Relative to the comparison of the 2011 vital registration and the 2011 Census data, the 2011 vital registration appeared to suggest that infant deaths were under recorded in the census.

The major limitation of the Blacker and Brass (2005) variant of the PBT method is that, for women with more than one birth in the 24 months before the census, the earlier birth is less likely to survive than the most recent birth (Hill, 2012c) and thus the dead child from the earlier birth will not be reported. Therefore, the proportion dead among most recent births in the 24 months is biased downwards relative to the proportion dead among all births.

In attempting to estimate a correction factor for the bias in the proportion dead among most recent births relative to the proportion dead among all births in the South African context using available full birth history data from the 1998 and 2003 SADHSs and the 2012 HSRC survey, it is apparent that the most recent data are of poor quality while the data which could be trusted were collected far back in time. Since there is no evidence as to how the bias in proportion dead among most recent births relative to the proportion dead among all births behaves, it was difficult to estimate the correction factor with a high degree of confidence. Thus, for illustrative purposes, an average of the estimates of the correction factors produced from the 1998 and 2003 SADHSs and the 2012 HSRC survey was used in this research to produce estimates of infant mortality from the data on survival of most recent births reported in the 2011 Census. It should be noted, though, that the average of the estimates of the correction factors produced from the 1998, 2003 SADHS and the 2012 HSRC surveys appear to underestimate the bias in the proportion dead among most recent births relative to the proportion dead among all births. Thus, this average can be treated as a broad indication of the actual bias.

The 2012 HSRC survey is based on a sample of around 15,000 households (Human Sciences Research Council, 2012). The number of deaths recorded in this survey were found to be few, suggesting that the survey missed a number of child deaths. There was nothing that could be done to correct for missed deaths. Also, among the few recorded deaths, a large proportion were missing age at death. This compounded the issue of missing deaths and had the effect of biasing the mortality estimates downwards. These limitations in the 2012 HSRC survey data made it impossible to produce useful estimates of childhood mortality from these data.

5.4 SCOPE FOR FUTURE RESEARCH

The problem with the quality of data produced from the Censuses and surveys in South Africa, particularly for indirect estimation of childhood mortality, is still pertinent. Although experts are consulted on census and survey methodologies, poor training of field workers, field work and supervision of field workers seem to be making this problem to be persistent. These were identified in earlier surveys as the major problems affecting the quality of data on childhood mortality estimation (Department of Health (South Africa) and Macro International Inc (USA), 2007). The issue of data quality was also observed in the 2012 HSRC survey data, especially the data collected for childhood mortality estimation. If better data are to be produced, institutions collecting data on demographic indicators should review the training manuals for field workers and field supervisors. The review would also include tools to continuously monitor and evaluate the processes in data collection so as to insure that high quality data are produced.

In addition to the above, the 2015 SADHS can be used to further research into the methods for estimating childhood mortality, particularly to identify why these methods are not working in the South African context. This will provide a basis for the strategy for deriving new methods for childhood mortality estimation if such need is identified. Also, further research into the magnitude of the bias in proportion dead among most recent births relative to proportion dead among all births in the South African context is recommended. Although our research attempted to estimate this bias, the results were inconclusive. Also, the behaviour of this bias, whether it increases, decreases or remain constant with time, need to be investigated as well.

5.5 CONCLUSIONS

With the 2011 Census data and the 2012 HSRC survey data, we have noted the same problems as those seen in the 2001 Census and the 2003 SADHS, suggesting that the problems of poor data quality continue. Thus it can be concluded that the 2012 HSRC full birth history data and the 2011 Census data on survival of most recent birth cannot be used to produce reliable estimates of childhood mortality. Only the deaths reported by households in the 2011 Census are found to be useful, and estimates produced from these data can add insight into the levels of infant and under-five mortality in South Africa.

REFERENCES

Actuarial Society of South Africa. 2011. *ASSA2008 AIDS and Demographic Model*. South Africa: ASSA. http://aids.actuarialsociety.org.za. Accessed: 21 January 2013.

Adetunji, J. A. 1996. "Infant mortality levels in Africa: Does method of estimation matter?", *Genus* **52**(3-4):89-106.

Aguirre, A. 1994. "Extension of the preceding birth technique", *Genus* **50**(3-4):151-169.

Alkema, L. and New, J. R. 2013. "Global estimation of child mortality using a Bayesian B-spline bias-reduction model", *arXiv*:1309.1602 [stat.AP].

Alkema, L. and You, D. 2012. "Child Mortality Estimation: A comparison of UN IGME and IHME Estimates of Levels and Trends in Under-Five Mortality Rates and Deaths", *PLoS Med,* **9**(8):e1001288.

Artzrouni, M. and Zaba, B. 2003. "HIV-induced bias in the estimation of child mortality using birth history reports," Paper presented at Technical Meeting on HIV and Child Mortality. London, March 2003.

Blacker, J. and Brass, W. 2005. "The estimation of infant mortality from proportions dying among births in the past 24 months", *Southern African Journal of Demography* **10**(1 & 2):25-42.

Bradshaw, D. and Dorrington, R. E. 2007. "Child mortality in South Africa-We have lost touch", *South African Medical Journal* **97**(8):582-583.

Bradshaw, D., Dorrington, R.E. and Laubscher, R. 2012. *Rapid Mortality Surveillance Report 2011*. Cape Town: South African Medical Research Council. http://www.mrc.ac.za/bod/RapidMortality2011.pdf. Accessed: 18 January 2013.

Brass, W. and Coale, A. J. 1968. "Methods of Analysis and Estimation," in Brass, W., Coale, A. J., Demeny, P., Heisel, D. F., Lorimer, F., Romaniuk, A. and Van de Walle, E. (eds). *The Demography of Tropical Africa*. Princeton NJ: Princeton University Press, pp. 88-139.

Brass, W. and Macrae, S. 1984. "Childhood mortality estimated from reports on previous births given by mothers at the time of a maternity: I. Preceding-births technique", *Asian and Pacific Census Forum* **11**(2):5-8.

Coale, A. J. and Demeny, P. 1966. *Regional model life tables and stable populations*. Princeton, New Jersey: Princeton University Press.

Collumbien, M. and Slogget, A. 2001. "Adjustment methods for bias in the indirect childhood mortality estimates.," in Zaba, B. and Blacker, J. (eds). *Brass Tacks. Essays in Medical Demography*. Athlone Press, pp. 21-42.

CoMMiC. 2012. *Childhood morbidity and mortality report*. South African National AIDS Council (SANAC). www.sanac.org.za/resources/cat_view/7-publications/9-reports. Accessed: 31 January 2014.

Darikwa, T. B. 2009. "Estimating the level and trends of child mortality in South Africa, 1996-2006." Unpublished Master's thesis, Cape Town: University of Cape Town.

Darikwa, T. B. and Dorrington, R. E. 2011. "The level and trends of child mortality in South Africa, 1996-2006", *African Population Studies* **25**(1):158-172.

Department of Health (South Africa) and Macro International Inc (USA). 2002. *South African Demographic and Health Survey 1998 Report*. Pretoria: Department of Health. http://www.measuredhs.com/publications/publication-fr131-dhs-final-reports.cfm. Accessed: 8 March 2013.

Department of Health (South Africa) and Macro International Inc (USA). 2007. *South African Demographic and Health Survey 2003 Report*. Pretoria: Department of

Health. http://www.measuredhs.com/publications/publication-FR206-DHS-Final-Reports.cfm. Accessed: 8 March 2013.

Dorrington, R. E., Moultrie, T. A. and Timæus, I. M. 2004. *Estimation of Mortality Using the South African Census 2001 data*. Cape Town: University of Cape Town. http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono11.pdf. Accessed: 20 December 2012.

Dorrington, R. E., Timæus, I. M., Moultrie, T. A. and Nannan, N. 2004. "Estimates of provincial fertility and mortality in South Africa, 1986-1996", *Southern African Journal of Demography* **9**(2):25-57.

Dorrington, R.E., Bradshaw, D. and Laubscher, R. 2014. *Rapid mortality surveillance report 2012*. Cape Town: South African Medical Research Council. Accessed: 17 February 2014.

Garenne, M. and Gakusi, E. 2005. *Under-Five Mortality Trends in Africa: Reconstruction from Demographic Sample Surveys*. Calverton, Maryland: ORC Macro. http://measuredhs.com/publications/publication-wp26-working-papers.cfm. Accessed: 18 January 2013.

Garenne, M. and Gakusi, E. 2010. *Reconstruction of trends in annual Under-Five Mortality rates in sub-Saharan Africa: September 2010 update*. France: Foundation for Studies and Research on International Development. www.ferdi.fr. Accessed: 01 February 2013.

Guillot, M., Gerland, P., Pelletier, F. and Saabneh, A. 2012. "Child Mortality Estimation: A Global Overview of Infant and Child Mortality Age Patterns in Light of New Empirical Data", *PLoS Med* **9**(8):e1001299.

Hallett, T. B., Gregson, S., Kurwa, F., Garnett, G. P., Dube, S., Chawira, G. *et al.* 2010. "Measuring and correcting biased child mortality statistics in countries with generalized epidemics of HIV infection", *Bulletin of the World Health Organization* **88**:761-768.

Hill, A.G. and Aguirre, A. 1990. "Childhood mortality estimates using the preceding birth technique - some applications and extensions", *Population Studies* **44**(2):317-340.

Hill, K. H. 2012a. "Direct estimation of child mortality from birth histories," in Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K., Timæus, I. M. and Zaba, B. (eds). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population (IUSSP). http://demographicestimation.iussp.org/content/direct-estimation-child-mortality-birth-histories. Accessed: 01 February 2013.

Hill, K. H. 2012b. "Indirect estimation of child mortality," in Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K. H., Timæus, I. M. and Zaba, B. (eds). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population (IUSSP). http://demographicestimation.iussp.org/content/indirect-estimation-child-mortality. Accessed: 01 February 2013.

Hill, K. H. 2012c. "Introduction to Child Mortality Analysis," in Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K. H., Timæus, I. M. and Zaba, B. (eds). *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population (IUSSP). http://demographicestimation.iussp.org/content/introduction-child-mortality-analysis. Accessed: 01 February 2013.

Human Sciences Research Council. 2012. *South African National HIV, Behaviour and Health Survey 2012 Fact Sheet 1*. South Africa: Human Sciences Research Council. http://www.hsrc.ac.za/Document-4331.phtml. Accessed: 18 January 2013.

Indepth Network. 2014. *Infant and Under-five mortality statistics: Agincourt and Africa Centre*. http://www.indepth-ishare.org/indepthstats/index.php/graphs. Accessed: 30 August 2014.

Joubert, J., Rao, C., Bradshaw, D., Dorrington, R.E., Vos, T. and Lopez, A.D. 2012. "Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa", *Global Health Action* **5**:19263.

Kahn, K., Garenne, M. L., Collison, M. A. and Tollman, S. M. 2007. "Mortality trends in a new South Africa: Hard to make a fresh start", *Scandinavian Journal of Public Health* **35**(Suppl 69):26-34.

Kerber, K., Tuaone-Nkhasi, M., Dorrington, R.E., Nannan, N., Bradshaw, D., Jackson, D. *et al.* 2012. "Progress Towards Millenium Development Goal 4. ", *The Lancet* **379**:1193.

Lozano, R., Wang, H., Foreman, K. J, Rajaratnam, J. K., Naghavi, M., Marcus, J. R *et al.* 2011. "Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: an updated systematic analysis.", *The Lancet* **378**:1139-1165.

Mahy, M. 2003. "Measuring Child Mortality in AIDS Affected Countries," Paper presented at The Workshop on HIV/AIDS and Adult Mortality in Developing Countries. New York, 8-13 September 2003. United Nations Population Division. 1-9.

Mayosi, B.M., Lawn, J.E., van Niekerk, A., Bradshaw, D., Abdool Karim, S.S. and Coovadia, H.M. 2012. "Health in South Africa: changes and challenges since 2009", *The Lancet* **380**:2029–2043.

Mazur, R. E. 1995. *Population structure, fertility and childhood mortality in South Africa: Lessons to be learnt from the analysis of the poverty survey*. South Africa: Centre for Epidemiological Research in Southern Africa, Medical Research Council. Accessed: 13 March 2013.

McKerrow, N. and Mulaudzi, M. 2010. "Child Mortality in South Africa: Using Existing Data," in Fonn, S. and Padarath, A. (eds). *South African Health Review*. Health Systems Trust, pp. 59-72.

Moultrie, T. A., Dorrington, R. E., Hill, A. G., Hill, K. H., Timæus, I. M. and Zaba, B. (eds). 2012. *Tools for Demographic Estimation*. Paris: International Union for the Scientific Study of Population (IUSSP). http://demographicestimation.iussp.org/. Accessed: 01 February 2013.

Moultrie, T. A., Sayi, T. S. and Timæus, I. M. 2012. "Birth intervals, postponement, and fertility decline in Africa: A new type of transition?", *Population Studies* **66**(3):241-258.

Moultrie, T. A. and Timæus, I. M. 2003. "The South African fertility decline: Evidence from two censuses and a Demographic and Health Survey", *Population Studies* **57**(3):265-283.

Nannan, N., Dorrington, R. E., Laubscher, R., Zinyakatira, N., Prinsloo, M., Darikwa, T. *et al.* 2012. *Under-5 Mortality Statistics in South Africa: Shedding Some Light on the Trends and Causes 1997-2007*. Cape Town: South African Medical Research Council. http://www.mrc.ac.za/bod/MortalityStatisticsSA.pdf. Accessed: 18 January 2013.

Nannan, N., Timæus, I. M., Laubscher, R. and Bradshaw, D. 2007. " Levels and differentials in childhood mortality in South Africa, 1977-1998", *BioSocial Science* **39**(4):613-632.

Preston, S. H., Heuveline, P. and Guillot, M. 2001. *Demography: Measuring and modelling population processes*. Oxford: Blackwell Publishing

Statistics South Africa. 2013. *Mortality and causes of death in South Africa, 2010: Findings from death notification*. P0309.3. Pretoria: Statistics South Africa.

http://www.statssa.gov.za/publications/p03093/p030932010.pdf. Accessed: 19 August 2013.

Statistics South Africa. 2014a. *An assessment of household deaths collected during Census 2011: A discussion document*. D0301.5. Pretoria: Statistics Africa. www.statsa.gov.za. Accessed: 24 April 2014.

Statistics South Africa. 2014b. *Census 2011 10% Sample Metadata*. Pretoria, South Africa: Statistics South Africa. www.statssa.gov.za. Accessed: 24 April 2014.

UN-IGME. 2013. *Child Mortality Estimates*. http://www.childmortality.org. Accessed: 23 June 2014.

United Nations. 2000. *United Nations Millennium Declaration*. New York: United Nations. http://www.un.org/millennium/declaration/ares552e.pdf. Accessed: 18 January 2013.

Walker, N., Hill, K. H. and Zhao, F. 2012. "Child Mortality Estimation: Methods Used to Adjust for Bias due to AIDS in Estimation Trends in Under-Five Mortality", *PLoS Med* **9**(8):e1001298.

Wang, H., Dwyer-Lindgren, L., Lofgren, K. T., Rajaratnam, J. K., Marcus, J. R., Levin-Rector, A. *et al.* 2012. "Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010", *The Lancet* **380**:2071-2094.

Wang, H., Liddell, C. A., Coates, M. M., Mooney, M. D., Levitz, C. E., Schumacher, A. E. *et al.* 2014. "Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013", *The Lancet* http://dx.doi.org/10.1016/S0140-6736(14)60497-9

Ward, P. and Zaba, B. 2008. "The effect of HIV on the estimation of child mortality using CS/CEB technique", *Southern African Journal of Demography* **11**(1):39-73.

APPENDIX A DOCUMENTATION FOR THE 2012 HSRC SURVEY DATA

The table below documents the inconsistencies that were observed in the 2012 HSRC survey data and how these were corrected using logical imputations on missing cases and contradicting cases.

Table A 1 Inconsistencies and irregularities in the 2012 HSRC survey and the logical imputations applied to these inconsistencies

Inconsistencies	Logical Imputations
Pregnancies were not reported in the correct order in which they occurred (base: variable q7_17)	Pregnancies were rearranged based on the reported date of birth.
Six (6) cases in which the month of birth was outside the range (1-12) (base: q7_17)	Cases with inconsistent month of birth were set to missing.
Fifteen (15) pregnancies with the duration of pregnancy stated but survival status at birth (born alive/born dead/lost before full term) not stated. Of these pregnancies, fourteen (14) had the duration stated as less than nine months.	Variable for showing sign of life at birth was recoded from missing to 'yes' if the child was reported as born alive. The variable for born alive/dead was recoded from missing to 'born alive' if the child was reported to have showed any sign of life at birth. The same was done for born dead and not showing any sign of life. Cases with the child born alive were retained for analysis.
Three (3) pregnancies where the reported date of birth did not equal the date at which the pregnancy ended. Also there were 770 pregnancies in which the date at which the pregnancy ended was stated and the date of birth was missing (base: q7_17 and q7_22).	Where date of birth of child was missing but the date at which that pregnancy ended was stated, the date of birth was replaced by the date at which the pregnancy ended if the child was reported as born alive and to have shown any sign of life at birth.
Twenty six (26) women reporting the vital status at birth (born alive/born dead/lost before full term) of at least one pregnancy inconsistently (base: q7_14 and q7_15). In these cases, either the child was reported as born alive (q7_14) but also reported as never showing sign of life at birth (q7_15) or vice versa.	The variable for born dead/alive was recoded from 'born dead', 'lost before full term' or 'missing' if date of birth and age at death were not missing. Also, the child was recoded to have shown signs of life at birth in these cases.
Thirty two (32) women reporting at least one child's age at death (B6) as one year.	Nothing was done to correct this.
Two (2) women reported never having sex, but has given birth.	The variable for never had sex was recoded to "yes" if pregnancy history were reported and "no" if the pregnancy history was not reported.
There were 23 cases in which males reported to have given birth.	Men giving birth were recoded to females after it was found that these cases reported on full birth history data

	as well as summary birth histories. All other variables for women were consistent except the sex variable for these cases. Three (3) cases for ever given birth were recoded from 'yes' to 'no' if sex was male as these did not report anything on summary births or pregnancy histories
There were 25 cases in which women reported on pregnancy histories but reported that they never gave birth.	The variable for ever given birth was recoded from 'no' or 'missing' to 'yes' if the woman reported on pregnancies history but had previously said she never gave birth or if the variable was missing.
Twenty three (23) cases where the respondent reported current use of contraceptive method but reported having given birth in the month of interview.	It is quite difficult to correct this error without going back to the questionnaire to check if it was correctly captured. As such, nothing could be done to correct for this, although it is an error that impacts directly on childhood mortality and thus, compromises the quality of the data.
Eighty (80) cases where a respondent's age at first birth was less than twelve years, as determined by the imputed date of birth of the first child (v212).	Nothing was done to correct this inconsistency. However, all cases with date of first birth missing were dropped.
Forty six (46) cases where a woman reported she had never menstruated in the last five years but had given birth to a child.	Another inconsistency in which nothing could be done to correct it without revisiting the questionnaire to check if it was correctly captured. This compromises the quality of the data.
Fifty five cases with missing date of interview (54 missing month only, 1 missing both month and year).	Five cases (5) with missing month of interview were imputed based on the year of interview and EA number since, in almost every case, the month of interview was the same for each EA.

APPENDIX B HOUSEHOLD DEATHS WITH EITHER AGE, SEX OR BOTH UNSPECIFIED, 2011 SOUTH AFRICAN CENSUS

Table B 1 Deaths with unspecified age expressed as a percentage of total deaths for each sex and population group combination

	Total	African	Coloured	Indian/Asian	White
Total	2.57%	2.71%	1.43%	2.18%	1.69%
Male	2.52%	2.68%	1.42%	2.05%	1.69%
Female	2.43%	2.56%	1.39%	1.97%	1.52%

Table B 2 Deaths with unspecified age expressed as a percentage of total deaths for each sex and province combination

	WC	EC	NC	FS	KZN	NW	GP	MP	LP
Total	0.48%	2.50%	2.81%	2.03%	3.66%	2.75%	2.59%	2.38%	1.83%
Male	0.42%	2.38%	2.70%	1.95%	3.61%	2.85%	2.52%	2.48%	1.88%
Female	0.54%	2.46%	2.69%	2.02%	3.42%	2.45%	2.46%	2.16%	1.63%

Table B 3 Deaths with sex only unspecified and both sex and age unspecified expressed as a percentage of total deaths for each population group

	Total	African	Coloured	Indian/Asian	White
Unspecified both age and sex	0.10%	0.10%	0.04%	0.17%	0.09%
Unspecified Sex only	0.35%	0.35%	0.40%	0.36%	0.32%

Table B 4 Deaths with sex only unspecified and both sex and age unspecified expressed as a percentage of total deaths for each province

	WC	EC	NC	FS	KZN	NW	GP	MP	LP
Unspecified both age and sex	0.01%	0.09%	0.12%	0.05%	0.16%	0.08%	0.11%	0.06%	0.08%
Unspecified Sex only	0.45%	0.33%	0.36%	0.24%	0.47%	0.24%	0.35%	0.30%	0.26%

APPENDIX C STATA CODE FOR THE BLACKER AND BRASS (2005) ESTIMATION METHOD

```
****Use birth recode file (ZARBR31FL)****
****Syntax is IMR_BB <input_file_name>
program IMR_BB
    version 12
args File_name
clear all
set more off
quietly {
use `File_name', clear
de,s
replace v007 = v007+1900 if v007<1900 & v007>20
replace v007 = v007+2000 if v007<20
** Keep only the variables needed**
keep v005 b3 b7 v001 v002 v003 v008 v013  v006 v007 b5 v000 b6 v016 bidx
gen wgt=v005/1000000
****generating pseudo random variables***
egen pseudorand1=cut(v002),group(10)
egen pseudorand2=cut(v016),group(10)
replace pseudorand1 = pseudorand1/10+0.05
replace pseudorand2 = pseudorand2/10+0.05
correl pseudorand1 pseudorand2
**Create date of birth variable****
        gen datbrtmo = b3 + pseudorand1
        replace v008=v008 + pseudorand1
**Create age at death in months**
        gen agedthmo = ((b6-100)+pseudorand2+0.5)/30.5 if b6>=100 &
b6<190
        replace agedthmo = (b6-200) + pseudorand2 if b6>=200 & b6<290
        replace agedthmo = 12*(b6-300 + pseudorand2) if b6>=300 & b6<390
        gen datdthmo = datbrtmo+agedthmo
        format agedthmo %5.2f
        format datdthmo %6.2f
***generating cmc between date of birth and interview date***
gen age_cmc = v008-datbrtmo
***births within the past 10 years from survey date in 24 month intervals***
recode age_cmc (0/23.999999=1 "0-24")(24/47.999999=2 "24-
48")(48/71.999999=3 "48-72")(72/93.999999=4 "72-96")(96/119.999999=5 "96-120"),
gen(age_cmc_grp)
replace age_cmc_grp=. if age_cmc >=120
***deaths within the past 10 years from survey date in 24 month intervals***
gen age_death =(v008-datbrtmo-agedthmo)
recode  age_death (0/23.999999=1 "0-24")(24/47.999999=2 "24-
48")(48/71.999999=3 "48-72")(72/93.999999=4 "72-96")(96/119.999999=5 "96-120"),
gen(age_death_grp)
replace age_death=. if age_death >=120
forvalues i=1/5 {
gen b5_`i' =0 if age_cmc_grp==`i'
recode b5_`i' 0=1 if age_death_grp==`i'
```

```
}
*****genereating results variables****
**All births in 24 month period**
forvalues i=1/5 {
gen numimr`i' = 0
gen denomimr`i' = 0
sum b5_`i' [iw=wgt]
gen imr`i' = 1.09*(r(sum)/r(sum_w))
gen propdead`i'=r(sum)/r(sum_w)
gen numdead`i'=r(sum)
gen births`i'=r(sum_w)
replace numimr`i' = numimr`i' + r(sum)/(r(sum_w)*(r(sum_w)-r(sum)))
replace denomimr`i' = denomimr`i' + log((r(sum_w)-r(sum))/r(sum_w))
gen hatsimr`i'  = sqrt(numimr`i'/(denomimr`i'*denomimr`i'))
gen maximr`i' = imr`i'^(exp(hatsimr`i'*1.96*-1))
gen minimr`i' = imr`i'^(exp(hatsimr`i'*1.96))
}
**Most recent births in 24 month period**
*Tagging most recent birth in 24 month period*
egen pid=concat(v001 v002 v003) //gen mother's id//
sort pid age_cmc_grp
egen lastbrt=max(b3) if b3!=., by(pid age_cmc_grp)
replace lastbrt=. if b3!=lastbrt & age_cmc_grp==age_cmc_grp[_n] &
pid==pid[_n]
*list pid brt_grp bidx lastbrt if brt_grp!=.
gen recentbrt=1 if lastbrt!=.
*list pid brt_grp bidx lastbrt recentbrt if brt_grp!=.
forvalues i=1/5 {
gen rec_numimr`i' = 0
gen rec_denomimr`i' = 0
sum b5_`i' if recentbrt==1 [iw=wgt]
gen rec_imr`i'=1.09*(r(sum)/r(sum_w))
gen rec_propdead`i'=r(sum)/r(sum_w)
gen rec_births`i'=r(sum_w)
gen rec_numdead`i'=r(sum)
replace rec_numimr`i' = rec_numimr`i' + r(sum)/(r(sum_w)*(r(sum_w)-r(sum)))
replace rec_denomimr`i' = rec_denomimr`i' + log((r(sum_w)-r(sum))/r(sum_w))
gen rec_hatsimr`i'  = sqrt(rec_numimr`i'/(rec_denomimr`i'*rec_denomimr`i'))
gen rec_maximr`i' = rec_imr`i'^(exp(rec_hatsimr`i'*1.96*-1))
gen rec_minimr`i' = rec_imr`i'^(exp(rec_hatsimr`i'*1.96))
}
**Bias in Proportion dead and imr**
forvalues i=1/5 {
gen bias_propdead`i'=1-(rec_propdead`i'/propdead`i')
gen bias_imr`i'=1-(rec_imr`i'/imr`i')
}
**Calculating the proportion giving birth following the death of the previous birth
sort pid b3 age_death_grp
egen tag1=tag(pid b3 age_death_grp) if b5==0 & b3[_n]<b3[_n-1] & b3!=. &
age_death_grp==age_cmc_grp
forvalues i=1/5 {
```

```
    sum tag1 if age_cmc_grp==`i' [iw=wgt]
    gen brt_dth`i'=r(sum)
    gen N`i'=r(sum_w)
    if _n==1 {
    gen pbd`i'=brt_dth`i'/N`i'
    }
   }
   }
   noisily {
   #delimit ;
   forvalues i=1/5 { ;
   display _newline(2) "     proportion dead among all births for period " 24*(`i'-1)
"-" 24*`i'-1 " months :  " propdead`i' ;
   display _newline(1) "     imr among all births for period " 24*(`i'-1) "-" 24*`i'-1 "
months :          " imr`i'*1000 ;
   display _newline(1) "     95% Confidence Intervals :      " minimr`i'*1000 "-"
maximr`i'*1000
                _newline(1) "   (Assuming simple random sampling)";
   display _newline(2) "     proportion dead among most recent births for period "
24*(`i'-1) "-" 24*`i'-1 " months :          " rec_propdead`i' ;
   display _newline(1) "     imr among most recent births for period " 24*(`i'-1) "-"
24*`i'-1 " months :       " rec_imr`i'*1000 ;
   display _newline(1) "     95% Confidence Intervals :      " rec_minimr`i'*1000 "-"
rec_maximr`i'*1000
                _newline(1) "   (Assuming simple random sampling)";
   display _newline(1) "     Bias in the proportion of children dead for period "
24*(`i'-1) "-" 24*`i'-1 " months :          " bias_propdead`i'*100 "%" ;
   display _newline(1) "     Proportion giving birth after death of previous birth for
period " 24*(`i'-1) "-" 24*`i'-1 " months : " pbd`i'*100 "%" ;
   };
   #delimit cr
   keep rec* prop* imr* min* max* births* numd* pbd*
   keep if _n==1
   save "C:\Users\Boboh\Desktop\dhs_out.dta", replace
   use "C:\Users\Boboh\Desktop\dhs_results.dta", clear
   append using "C:\Users\Boboh\Desktop\dhs_out.dta"
   save "C:\Users\Boboh\Desktop\dhs_results.dta", replace
   }
   end
```

APPENDIX D STATA CODE FOR THE DIRECT ESTIMATION METHOD

```
****Use birth recode file (ZARBR31FL)****
*  Syntax is U5MR <input_file_name> <initial_year> <final_year>
program U5MR
 version 12
args File_name Start_year End_year
tempvar old_file year yrdth nexp e_a e_d fac1 fac5
clear all
set more off
quietly {
use `File_name'
de,s
capture log close
log using `File_name',append text
log off
**Check to see whether data are probably DHS type
confirm numeric variable b3 b5 b6 b7 bidx v001 v002 v003 v005 v007 v008 v013
****keeping only variables needed****
keep b3 b5 b6 b7 v002 v006 v007 v005 v008 v016
**Normalizing the weight variable
gen wgt=v005/1000000
****generating pseudo random variables***
egen pseudorand1=cut(v002),group(10)
egen pseudorand2=cut(v016),group(10)
replace pseudorand1 = pseudorand1/10+0.05
replace pseudorand2 = pseudorand2/10+0.05
correl pseudorand1 pseudorand2
****Create date of birth variable****
gen dob = b3+pseudorand1
****Create age at death (aad) and date of death (dod) in months****
gen aad = ((b6-100)+pseudorand2)/31 if b6>=100 & b6<190 //age at death given in days//
replace aad = (b6-200) + pseudorand2 if b6>=200 & b6<290 //age at death given in months//
replace aad = 12*(b6-300 + pseudorand2) if b6>=300 & b6<390 //age at death given in years//
gen dod = dob+aad
format aad %5.2f
format dod %6.2f
scalar startyear = `Start_year'
scalar endyear = `End_year'
** Check the date ranges
if startyear < 1960 | endyear < 1960 |startyear > 2015 | endyear > 2015 {
      di in red "Beginning or ending period outside range 1960 to 2015"
      error 1
}
if startyear > endyear {
      di in red "End year must be equal to or greater than beginning year"
      error 1
```

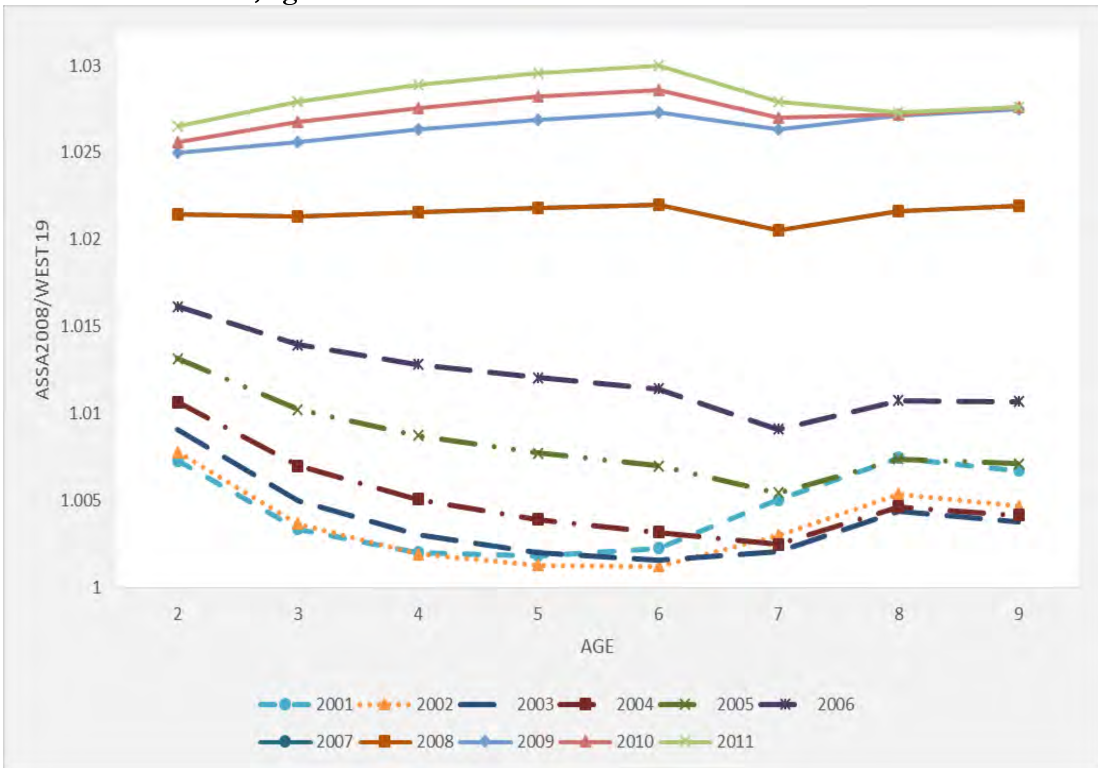
```
}
** Start the calculations
}
 di in yellow _dup(63) "_"
 di _newline(2) "Life Table Mortality Risks in Childhood for Period Jan/"
startyear " - Dec/" endyear ":"
 di _column(10) "(using data from file " "`File_name')"
quietly {
compress
local yearmin = `Start_year' - 1900
local yearmax = `End_year' - 1900
confirm numeric variable wgt dob dod aad
***Tally variables***
gen `e_a'=.
gen `e_d'=.
***Deaths and Exposure variables
forvalues i=0/59 {
  gen d`i'=0
  gen exp_a`i'=0
  gen exp_d`i'=0
  gen n`i'=0
}
gen `yrdth' = int(dod/12)
forvalues year=`yearmin'/`yearmax' {
  capture drop adj
  gen int adj = (`year' - 70)*12
***Identifying deaths aged 0-12 months in each time period (weighted)
forvalues i=0/59 {
  replace d`i'=cond(aad>=`i' & aad<`i'+1 & `yrdth'==`year', d`i'+wgt, d`i')
}
***identifying exposure alive, exposure dead and weighted cases
forvalues i=0/59 {
  replace `e_a'=wgt if 841+adj-dob<`i' & 852+adj-dob>`i'+1 & aad>=`i'+1
  replace `e_a'=wgt*(`i'-840-adj+dob) if 841+adj-dob>=`i' & 841+adj-dob<`i'+1
& 852+adj-dob>=`i'+1 & aad>=`i'+1
  replace `e_a'=wgt*(852+adj-dob-`i') if 841+adj-dob<`i' & 852+adj-dob>=`i' &
852+adj-dob<`i'+1 & aad>=`i'+1
  replace `e_d'=wgt*(aad-`i') if 841+adj-dob<`i' & 852+adj-dob>=`i'+1 &
aad>=`i' & aad<`i'+1
  replace `e_d'=wgt*(aad-841-adj+dob) if 841+adj-dob>=`i' & 841+adj-
dob<`i'+1 & 852+adj-dob>=`i'+1 & aad>=`i' & aad<`i'+1
  replace `e_d'=wgt*(aad-`i') if 841+adj-dob<`i' & 852+adj-dob>=`i' & 852+adj-
dob<`i'+1 & aad>=`i' & aad<`i'+1
  replace exp_a`i'=cond(`e_a'!=., exp_a`i'+`e_a', exp_a`i')
  replace exp_d`i'=cond(`e_d'!=., exp_d`i'+`e_d', exp_d`i')
  replace n`i'=cond(`e_a'!=., n`i'+`e_a', n`i')
  replace n`i'=cond(`e_d'!=., n`i'+`e_d', n`i')
  replace `e_a'=.
  replace `e_d'=.
}
}
```

```
forvalues i=0/59 {
***Total exposure and deaths
  egen event`i'=total(d`i')
  egen expa`i'=total(exp_a`i')
  egen expd`i'=total(exp_d`i')
  egen N`i'=total(n`i')
  gen expo`i'=(expa`i'+expd`i')/12
}
if _n==1 {
***Mortality Calculation
  forvalues i=0/59 {
          gen nmx`i'=cond(expo`i'>0, event`i'/expo`i', 0)
          gen a`i'=cond(`i'==0, expd`i'/(event`i'*12), 1/24)
          gen nqx`i'=(nmx`i'/12)/(1+((1/12)-a`i')*nmx`i')
  }
  gen nnmr=nqx0
  gen `fac1'=1
  forvalues i=0/11 {
          replace `fac1'=`fac1'*(1-nqx`i')
  }
  gen imr=1-`fac1'
  gen pnnmr=imr-nnmr
  gen `fac5'=1
  forvalues i=0/59 {
          replace `fac5'=`fac5'*(1-nqx`i')
  }
  gen q5=1-`fac5'
  gen q41=(q5-imr)/(1-imr)
** Calculate "simple random sample" estimates of confidence intervals for imr
and q5:
  gen numimr = 0
  gen denomimr = 0
  gen numq5 = 0
  gen denomq5 = 0
  forvalues i = 0/11  {
          replace numimr = numimr + event`i'/(N`i'*(N`i'-event`i'))
          replace denomimr = denomimr + log((N`i'-event`i')/N`i')
  }
  gen hatsimr  = sqrt(numimr/(denomimr*denomimr))
  gen maximr = imr^(exp(hatsimr*1.96*-1))
  gen minimr = imr^(exp(hatsimr*1.96))
  forvalues i = 12/59 {
          replace numq5 = numq5+ event`i'/(N`i'*(N`i'-event`i'))
          replace denomq5 = denomq5+ log((N`i'-event`i')/N`i')
  }
  replace numq5 = numq5+numimr
  replace denomq5 = denomq5 + denomimr
  gen hatsq5 = sqrt(numq5/(denomq5*denomq5))
  gen maxq5 = q5^(exp(hatsq5*1.96*-1))
  gen minq5 = q5^(exp(hatsq5*1.96))
}
```

```
}
noisily {
capture log close
log using `File_name',append text
#delimit ;
disp _newline(2) "  Neonatal Mortality Rate          :  " %5.4f nnmr*1000 ;
disp "  Post-Neonatal Mortality Rate     :  " %5.4f pnnmr*1000 ;
disp "  Infant Mortality Rate      (1q0) :  " %5.4f imr*1000
     _newline
   "         (95% Confidence Limits :  " %5.4f minimr*1000 " to " %5.4f
maximr*1000
  _newline
  "         assuming simple random sampling)";
disp "  Child Mortality Rate       (4q1) :  " %5.4f q41*1000 ;
disp "  Under Five Mortality Rate  (5q0) :  " %5.4f q5*1000
     _newline
   "         (95% Confidence Limits :  " %5.4f minq5*1000 " to " %5.4f
maxq5*1000
  _newline
  "         assuming simple random sampling)";
keep nnmr pnnmr imr min* max* q41 q5;
keep if _n==1 ;
save "C:\Users\Boboh\Desktop\dhs_out.dta", replace ;
use "C:\Users\Boboh\Desktop\dhs_results.dta", clear ;
append using "C:\Users\Boboh\Desktop\dhs_out.dta" ;
save "C:\Users\Boboh\Desktop\dhs_results.dta", replace ;
};
#delimit cr
end
```

APPENDIX E ASSA2008 VS PRINCETON WEST LEVEL 19

Figure E 1 Comparison of the Princeton West level 19 model life tables with the ASSA2008 life tables, age 0-20



Note: ASSA2008 is divided by West level 19 for each of the years between 2001 and 2011 inclusive

The shape of mortality in ASSA2008 model is compared to that of Princeton West Level 19 as shown in Figure E 1 above. Assuming the same shape of mortality in the two models, one would expect an approximately horizontal line. Thus ASSA2008 and Princeton West level 19 have different shapes of mortality for the years prior to 2008. ASSA2008 model assumes heavier mortality than Princeton West level 19 for age two and below which monotonically decreases until age six. For the years 2008 to 2011, the mortality in ASSA2008 is heavier than in Princeton West level 19 but the shape of mortality is the same in the two models as evidenced by near horizontal lines shown in Figure E 1. There is also a significant difference between the curves for the years 2001 to 2007 and the curves for 2008 to 2011, with the curves for the years 2008 to 2011 being higher than those for the years 2001 to 2007. The difference can be attributed to the fact that ASSA2008 model projected the rates of non-AIDS mortality based on the mortality experiences prior to the year 2008.