Measuring child mortality in resource limited settings using alternative approaches: South African case study

Nadine Nannan
University of Cape Town

Thesis submitted for the degree of PhD in Demography to the Faculty of Commerce

July 2017
The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.
ABSTRACT

Post the Millennium Development Goal project a significant number of countries are still faced with the challenge of monitoring child mortality. Despite numerous enquiries since 1996 to provide this basic health indicator, South Africa has experienced prolonged periods of uncertainty regarding the level and trend of infant and under-5 mortality.

The thesis develops an analytical framework to review all available data sources and methods of analysis and presents the results of the four approaches adopted to measure trends in child mortality. Reviewing the demographic indicators produced from seven census and survey enquiries, the overall performance and the strengths and limitations of each approach is evaluated. Poor and extremely poor quality of data for child mortality emerges as a pervasive challenge to census and survey data. The thesis presents the remarkable improvement in the completeness of birth and death registration through South Africa’s CRVS system, particularly since 2000, illustrating the possibility of using CRVS data to monitor provincial child mortality in the future and highlighting statistical challenges arising from the movement of children.

In conclusion, South Africa should focus on improving CRVS for purposes of monitoring childhood mortality provincially and the comprehensive evaluation of available data is a useful lesson for other upper-middle-income countries.
# TABLE OF CONTENTS

ABSTRACT .................................................................................................................. 2
TABLE OF CONTENTS ................................................................................................. 3
LIST OF TABLES ............................................................................................................. 7
LIST OF FIGURES ........................................................................................................ 8
GLOSSARY .................................................................................................................... 10
ACRONYMS AND ABBREVIATIONS ........................................................................... 12
ACKNOWLEDGEMENTS ............................................................................................... 15

## 1 INTRODUCTION ................................................................................................. 16
  1.1 Measurement of childhood mortality ................................................................. 16
  1.2 Childhood mortality in South Africa ................................................................. 18
  1.3 Challenges in the measurement of childhood mortality .................................... 18
  1.4 A framework for evaluation ............................................................................. 19
  1.5 Aims and Objectives ....................................................................................... 22
  1.6 Chapter outline ............................................................................................... 23

## 2 MEASUREMENT OF CHILDHOOD MORTALITY:
A REVIEW OF THE LITERATURE ............................................................................. 24
  2.1 Introduction ....................................................................................................... 24
  2.2 Trends in child mortality .................................................................................. 24
  2.3 Data sources for measuring childhood mortality ............................................. 27
  2.4 Methodological challenges ............................................................................. 38
  2.5 Trends in child mortality in South Africa ....................................................... 43
  2.6 Evaluation of civil registration and vital statistics ........................................... 60
  2.7 The age-pattern of mortality within childhood .............................................. 62
  2.8 Conclusion ....................................................................................................... 66

## 3 METHODS ......................................................................................................... 68
  3.1 Overview of the methodologies ....................................................................... 68
  3.2 Measures of child mortality: definitions, terms and calculation ................. 68
  3.3 The life table .................................................................................................... 71
  3.4 Full birth history .............................................................................................. 73
  3.5 Summary birth history ..................................................................................... 75
  3.6 Adjustment for mother’s survivor bias ........................................................... 76
  3.7 The previous birth technique .......................................................................... 78
  3.8 Deaths in the household ................................................................................... 80
  3.9 Episode data: Individuals, membership, residency and migration
      surveillance .......................................................................................................... 80
  3.10 Regression analysis of count data ................................................................. 82
  3.11 Assessment of CRVS as a tool for measuring child mortality .................... 83
| Table 2-1: Estimates of infant mortality using data collected prior to 1994 | 44 |
| Table 2-2: Estimates of infant mortality using data collected post 1994 | 48 |
| Table 3-1: Direct methods for calculating child mortality rates | 69 |
| Table 3-2: Data sources, type of data and count and analytical method | 70 |
| Table 3-3: Data sources and references used to calculate the number of births based on age-specific fertility rates and the number of women by age | 84 |
| Table 3-4: Data sources and researchers of mortality estimates used to assess completeness of death registration by age | 87 |
| Table 4-1: Comparison of births and deaths by calendar year for the 1998 and 2003 SADHS's | 92 |
| Table 4-2: Comparison of the estimated bias produced by different methods for each source of birth history data, under-five mortality | 101 |
| Table 4-3: Adapted assessment framework for civil registration and vital statistics systems Mahapatra, Shibuya, Lopez, Coullare et al. (2007) | 105 |
| Table 5-1: Lifetime fertility and the survival of children by women’s age, Census 2011 | 118 |
| Table 6-1: Recorded births by age of mother at time of the survey | 126 |
| Table 6-2: Recorded survival status by sex | 126 |
| Table 6-4: Comparison of child mortality indices based on estimates derived from birth histories, area surveillance, and their relative proportions, ACDIS 2000 | 138 |
| Table 6-5: Comparison of demographic details of three groups of infant deaths reported in birth histories and routine area surveillance in 2000 | 140 |
| Table 6-6: Comparison of the diagnostic statistics for assessment of the age, sex, year Models 1-6 | 145 |
| Table 7-2: Estimated completeness of registered births within the year of birth plus the following year compared against various other estimates, 2001-2012 | 157 |
| Table 7-4: Implied completeness of registered births registered in the year of birth and in the year of birth ±1 compared with births estimated from the adjusted DHIS data and the 2011 Census count, 2000-2012 | 163 |
| Table 7-5: Comparison of the implied completeness of death registration in 2011 estimated from the 2011 Census and the implied completeness estimated from the 2007 Community Survey | 173 |
| Table 8-1: Adjustments to the provincial number of births derived from reverse survival (RS), Census 2011 | 185 |
| Table 8-2: Implied completeness of registered births registered in the year of birth ±1 compared with births derived from the 2011 Census and adjusted DHIS births, 2001-2011 | 187 |
Figure 2-1: Model estimates of under-five mortality trends in South Africa, 2005 .... 56
Figure 2-2: Model estimates of under-five mortality, South Africa 2015 .................. 59
Figure 2-3: Cause proportions and the ratio of the neonatal mortality rate to the under-5 mortality rate at different levels of under-5 mortality.......................... 66
Figure 3-1: Lexis diagram representing component death probabilities ............. 74
Figure 3-2: Correlation of proportions rural recorded in the 1996 Census and proportion of home-births recorded in the 1998 DHS.......................... 86
Figure 4-1: Distribution of births seven years before the survey, SADHS 1998 and 2003.......................................................... 93
Figure 4-2: Reported number of deaths in the years before the 1998 and 2003 SADHS’s .................................................................................... 95
Figure 4-3: Age-specific period mortality, single year estimates, SADHS 1998.... 97
Figure 4-4: Age-specific period mortality averaged over two years, SADHS 2003 .. 98
Figure 4-5: Probability of dying before age five averaged over two years, and adjusted for mothers’ survivor bias using IGME approach, SADHS 2003.... 100
Figure 4-6: ACDIS infant and under-5 mortality trend adjusted for mother’s survivor bias .................................................................................. 100
Figure 5-1: Wording of the fertility questions in the 2001 Census...................... 109
Figure 5-2: Comparison of 2011 Census age distribution and the official 2013 mid-year estimates, Stats SA................................................................. 111
Figure 5-3: Trend in the number of births derived from the enumerated population under fifteen years of age compared with various census and survey estimates.............................................................. 112
Figure 5-4: Distribution of deaths in the household under the age of five by month of death, 2011 ........................................................................ 114
Figure 5-5: Comparison of childhood mortality trends (adjustment1 1q0, 4q1, 5q0) using household deaths with the Rapid Mortality Surveillance, 2013 115
Figure 5-6: Comparison of provincial mortality indices produced from the reported household deaths at approximately February 2011........................ 116
Figure 5-7: Comparison of under-5 mortality trend with Stats SA, derived from the summary birth history and the Rapid Mortality Surveillance 2013.... 119
Figure 6-1: ACDIS age-specific mortality trend, 1981-2000 .............................. 127
Figure 6-2: Comparison of ACDIS under-5 mortality trend with other estimates... 128
Figure 6-3: Number of live births by residence status, 2000-2007.................. 133
Figure 6-4: Sex ratio at birth, 2000-2007...................................................... 134
Figure 6-5: Number of neonatal episodes for boys and girls, 2000-2007 .......... 135
Figure 6-6: Age components of under-five deaths, 2000-2007...................... 135
Figure 6-8: Age-specific mortality of all children in the surveillance area, 2000-2007 .................................................................................. 137
Figure 6-9: Infant deaths recorded in the retrospective data in 2000 and found in the prospective surveillance records ........................................ 139
Figure 6-10: Mother and infant residence in 2000 according to data source...... 140
Figure 6-11: Comparative measures of under-two mortality from the ACDIS 2000-2006 data set ........................................................................ 143
Figure 6-12: Graphic presentation of age, sex, and year modelled effects over time, Models 2-6 .............................................................................. 147
Figure 7-2: Comparison of births registered by year of occurrence with model based projections of annual births, 2000-2012................................. 157
Figure 7-3: Comparison of empirical estimates of births, 1996-2012.............. 161
Figure 7-4: Comparison of births registered in the calendar year of birth and the following year of birth (2000-2012) with births occurring in health facilities and births derived from Census 2011 .......................................................... 162
Figure 7-5: Trends in the completeness of birth registration by age measured against an estimate of births derived from 2011 census, 1995-2012............. 164
Figure 7-6: Trend in the reported number of deaths by age, Stats SA 1990-2011. 168
Figure 7-7: Comparison of the number of deaths registered by year of death with model based projections of annual deaths, 1990-2013....................... 169
Figure 7-8: Empirical estimates of the probability of dying before age five compared with mortality estimates derived from vital registration and empirical estimates of births ................................................................. 171
Figure 7-9: Comparison of mortality derived from adjusted vital registration applying alternative assumptions for estimating completeness....................... 172
Figure 7-10: Observed $1q_0$ and $4q_1$ relative to the level of $5q_0$ relationships from the 1998 DHS, 2007 Community Survey and the 2011 Census compared with the Princeton model life tables .......................................... 176
Figure 7-11: Comparison of the relationship between $1q_0$ and $4q_1$ from international agencies at a level of $5q_0$ from the Thembisa model for each year 1986-2011 against the Princeton families of model life tables............................. 177
Figure 7-12: Comparison of the standardised ratios of $4q_1$ relative to $1q_0$ produced by IHME(2013), IGME(2015) and WPP(2015), against the values represented in Thembisa and the West and North MLT, 1986-2011 178
Figure 7-13: Comparison of the standardised ratios of $4q_1to1q_0$ produced by the CRVS, RMS, NBD against the values represented in Thembisa and the West and North MLT, 2000-2011.......................................................... 179
Figure 7-14: Trend in infant and child mortality rates by broad cause group, 1997-2010 .................................................................................................................. 180
Figure 8-1: Provincial births derived by back projecting the recorded province of birth recorded in the 2011 Census, 1996-2010.............................. 183
Figure 8-2: Provincial comparison of trends in the number of births from various sources ........................................................................................................ 186
Figure 8-3: Implied provincial completeness of registered deaths under age five measured against estimates derived from the 2007 Community Survey and the 2011 Census ......................................................... 192
Figure 8-4: Comparison of reported and adjusted $1q_0$, $4q_1$ and $5q_0$ against the NBD study, by province ........................................................................ 193
Figure 9-1: Historical trend in under-5 mortality in South Africa 1975-2011, compared with IGME, IHME, WPP and Thembisa 2015 modelled trends and the RMS 2014 ......................................................... 207
Glossary

<table>
<thead>
<tr>
<th>Mortality Index</th>
<th>Definition</th>
<th>Age range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Neonatal rate</td>
<td>The probability of dying between birth and the 6th day of life</td>
<td>0-6 days</td>
</tr>
<tr>
<td>Late Neonatal rate</td>
<td>The probability of dying between the 6th day and 27th day of life</td>
<td>7-27 days</td>
</tr>
<tr>
<td>Neonatal rate NNMR</td>
<td>The probability of dying between birth and the 27th day of life</td>
<td>0-27 days</td>
</tr>
<tr>
<td>Post-neonatal rate PNMR</td>
<td>The probability of dying between the 1st and 11th month of birth</td>
<td>1-11 months</td>
</tr>
<tr>
<td>Infant mortality rate IMR</td>
<td>The probability of dying between birth and age one and is denoted by (q_0)</td>
<td>0-11 months</td>
</tr>
<tr>
<td>Child mortality rate CMR</td>
<td>The probability of dying between exact age one and exact age five and is denoted by (q_1)</td>
<td>1-4 years</td>
</tr>
<tr>
<td>Under-five mortality rate U5MR</td>
<td>The probability of dying between birth and exact age five and is denoted by (q_0)</td>
<td>0-4 years</td>
</tr>
<tr>
<td>Childhood mortality rate</td>
<td>Refers to the mortality of the, infant, child and overall under-five age groups together</td>
<td>0-1; 1-4 and 0-4</td>
</tr>
</tbody>
</table>

In this thesis two adjustments have been made to the mortality rates:

Adjustment 1

An adjustment to the mortality rates (2006-2011) assuming the completeness of VR deaths based on the estimated mortality from the 2007 Community Survey and the reported household deaths in the 2011 Census under-1 and 1-4 years old.

Adjustment 2

A second adjustment made to estimates of \(q_0\) and \(q_1\) in 2011 based on the assumption of non-changing completeness of the VR deaths based on the estimated mortality from the 2007 Community Survey and the reported household deaths in the 2011 Census under-1 and 1-4 years old.
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACDIS</td>
<td>Africa Centre Demographic Information System</td>
</tr>
<tr>
<td>ACME</td>
<td>Automated Classification of Medical Entities</td>
</tr>
<tr>
<td>AIDS</td>
<td>Acquired immune deficiency syndrome</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral treatment</td>
</tr>
<tr>
<td>ASSA</td>
<td>Actuarial Society of South Africa</td>
</tr>
<tr>
<td>CEB/CS</td>
<td>Children ever borne/children surviving</td>
</tr>
<tr>
<td>Child PIP</td>
<td>Child Healthcare Problem Identification Programme</td>
</tr>
<tr>
<td>CRVS</td>
<td>Civil Registration and Vital Statistics</td>
</tr>
<tr>
<td>CS</td>
<td>Community Survey</td>
</tr>
<tr>
<td>DHA</td>
<td>Department of Home Affairs</td>
</tr>
<tr>
<td>DHIS</td>
<td>District Health Information System</td>
</tr>
<tr>
<td>DNF</td>
<td>Death notification form</td>
</tr>
<tr>
<td>DSA</td>
<td>Demographic surveillance area</td>
</tr>
<tr>
<td>DSS</td>
<td>Demographic surveillance site</td>
</tr>
<tr>
<td>EC</td>
<td>Eastern Cape</td>
</tr>
<tr>
<td>FBH</td>
<td>Full birth history</td>
</tr>
<tr>
<td>FS</td>
<td>Free State</td>
</tr>
<tr>
<td>GOBBI-FFF</td>
<td>growth monitoring (G), oral rehydration (O), breast feeding (B), immunization</td>
</tr>
<tr>
<td></td>
<td>(I), birth spacing/family planning (F), food supplementation (F) promotion</td>
</tr>
<tr>
<td></td>
<td>of female education (F)</td>
</tr>
<tr>
<td>GT</td>
<td>Gauteng</td>
</tr>
<tr>
<td>HDSS</td>
<td>Health and Demographic Surveillance Site</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HMD</td>
<td>Human mortality database</td>
</tr>
<tr>
<td>HSRC</td>
<td>Human Sciences Research Council</td>
</tr>
<tr>
<td>ICD</td>
<td>International Statistical Classification of Diseases and Related Health Problems</td>
</tr>
<tr>
<td>IGME</td>
<td>Inter-agency Group for the Estimation of child Mortality</td>
</tr>
<tr>
<td>IHME</td>
<td>Institute for Health Metrics and Evaluation</td>
</tr>
<tr>
<td>IMR</td>
<td>Infant Mortality Rate</td>
</tr>
<tr>
<td>INDEPTH</td>
<td>International Network for the continuous Demographic</td>
</tr>
</tbody>
</table>
Evaluation of Populations and their health in developing countries

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LP</td>
<td>Limpopo</td>
</tr>
<tr>
<td>MP</td>
<td>Mpumalanga</td>
</tr>
<tr>
<td>KZN</td>
<td>KwaZulu-Natal</td>
</tr>
<tr>
<td>MDGs</td>
<td>Millennium Development Goals</td>
</tr>
<tr>
<td>MLT</td>
<td>Model life tables</td>
</tr>
<tr>
<td>NBD</td>
<td>National burden of disease</td>
</tr>
<tr>
<td>NC</td>
<td>Northern Cape</td>
</tr>
<tr>
<td>NIDS</td>
<td>National Income Dynamics Survey</td>
</tr>
<tr>
<td>NIMSS</td>
<td>National Injury Mortality Surveillance System</td>
</tr>
<tr>
<td>NMR</td>
<td>Neonatal mortality rate</td>
</tr>
<tr>
<td>NPR</td>
<td>National Population Register</td>
</tr>
<tr>
<td>NW</td>
<td>North West</td>
</tr>
<tr>
<td>OHS</td>
<td>October Household Survey</td>
</tr>
<tr>
<td>PBT</td>
<td>Preceding birth technique</td>
</tr>
<tr>
<td>PNMR</td>
<td>Post-neonatal mortality rate</td>
</tr>
<tr>
<td>PPIP</td>
<td>Perinatal Problem Identification Programme</td>
</tr>
<tr>
<td>PMTCT</td>
<td>Prevention-of-mother-to-child-transmission</td>
</tr>
<tr>
<td>RMS</td>
<td>Rapid mortality surveillance</td>
</tr>
<tr>
<td>RS</td>
<td>Reverse survival</td>
</tr>
<tr>
<td>SADHS</td>
<td>South African Demographic and Health Survey</td>
</tr>
<tr>
<td>SALDRU</td>
<td>South Africa Labour Development Research Unit</td>
</tr>
<tr>
<td>SBH</td>
<td>Summary birth history</td>
</tr>
<tr>
<td>SDGs</td>
<td>Sustainable Development Goals</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>TBVC</td>
<td>Transkei, Bophuthatswana, Venda and Ciskei</td>
</tr>
<tr>
<td>U5MR</td>
<td>Under-5 mortality rate</td>
</tr>
<tr>
<td>UN</td>
<td>United Nations</td>
</tr>
<tr>
<td>UNAIDS</td>
<td>Joint United Nations Programme on HIV/AIDS</td>
</tr>
<tr>
<td>UNDP</td>
<td>United Nations Development Programme</td>
</tr>
<tr>
<td>UNICEF</td>
<td>United Nations Children’s Economic Fund</td>
</tr>
<tr>
<td>UNPD</td>
<td>United Nations Population Division</td>
</tr>
<tr>
<td>VR</td>
<td>Vital Registration</td>
</tr>
<tr>
<td>Acronym</td>
<td>Full Form</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>WC</td>
<td>Western Cape</td>
</tr>
<tr>
<td>WFS</td>
<td>World Fertility Survey</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>WPP</td>
<td>World Population Prospects</td>
</tr>
</tbody>
</table>
I would like to thank my supervisors Prof. Rob Dorrington and Prof. Debbie Bradshaw, without whose input and guidance this thesis would not have been completed. During the years under their supervision they have taught me to be more rigorous in my approach to understanding and developing the research question. Their clarity of thought and insights were invaluable. I thank them both for their dedication, diligence and at times much patience for seeing the thesis through to successful completion.

I would like to thank the South Africa Medical Research Council for financial assistance in the first three years of this thesis. I would also like to thank the Medical Research Council for allowing me the necessary time for my thesis and my colleagues in the Burden of Disease Research Unit for allowing me the time away from the collaborative projects within the unit.

Nadine Nannan
Cape Town, July 2017
1.1 Measurement of childhood mortality

The measurement of childhood mortality is prominent on the health agenda. It was identified as a key target of the United Nations Millennium Development Goals (MDGs) project (United Nations 2000) to monitor country progress towards the reduction of under-five mortality. A major conclusion at the end of that project was that low- and middle-income countries face severe challenges measuring and monitoring childhood mortality rates due to inadequate civil registration and vital statistics (CRVS) (AbouZahr, de Savigny, Mikkelsen et al. 2015).

In settings where CVRS are not complete, alternative measures that can provide key indicators of population health are useful substitutes (Hill, Lopez, Shibuya et al. 2007). Census and surveys have, for example, long been used to provide periodic estimates of the level of childhood mortality and more recently demographic surveillance data have become increasingly important sources of child mortality data in many regions of the world. However, due to the cost implications of such methods, many countries have been reliant on international agencies to provide estimates derived from statistical models of infant and under-five mortality (UN IGME 2011).

Various demographic indicators are used as key criteria to assess population health and socioeconomic development globally. Infant mortality is a sensitive indicator of socioeconomic change and has been widely used by international agencies to compare health and socioeconomic progress between regions and individual countries of the world; hence, it informs a broad policy arena. Infant and under-five mortality also play an essential role in the assessment of the health and development of children.

In this regard, child mortality indicators serve as useful indicators for policy formulation and planning because they not only reflect on the health of children but also on the health status of the entire community. This is because a child’s growth and development are greatly dependent on the living conditions of the family, the level of infrastructure (safe water, hygiene and sanitation) provided to the household and services accessible (education and health facility) to the broader community (Mosley and Chen 1984). All of these elements generate the biological factors that impact directly on the child’s health through the incidence and progress of disease or death as a possible outcome. Hence, changes in the social conditions of households and the effect of the broader environment on child mortality can be detected through empirical measurement
in the population. In contrast, another demographic indicator, the life expectancy also identifies similar social dynamics but is fundamentally different in that it is a synthetic indicator, which is not only practically difficult to measure in a population, but is also a function of many variables.

Child mortality indicators have for these reasons also been widely used to illustrate mortality variation by geographic areas, highlighting mortality gradients where higher mortality is usually found among the poorer households (Wagstaff, Bustreo, Bryce et al. 2004). Child mortality is therefore extensively used as an indicator to identify poor health status and social inequalities in health and socioeconomic circumstance between countries, within countries and even between areas within the same country, province or district.

It is not surprising then, that child mortality was identified as one of the eight MDGs (MDG 4) for measuring progress towards improving population health and reducing health inequalities in the world, and tracking progress towards the target of a two-thirds reduction in child mortality by 2015 has taken prominence among international agencies. Through this programme, it has become evident that monitoring child mortality is proving challenging for low- and middle-income countries. The “Who Counts?” series (Setel, Macfarlane, Szreter et al. 2007) highlights the fact that in such countries vital registration is often incomplete, lacking the recording of many deaths particularly of the poor. Through the process of evaluating global progress towards MGD 4, it has been necessary to estimate levels of child mortality on a country basis, which has led to disagreement between the institutions who measure indicators of health and mortality globally. In particular disagreement between, the various sub-agencies of the United Nations such as the Inter-agency Group for the Estimation of child Mortality (IGME) and other research groups such the Institution for Health Metrics Evaluation (IHME) has come to light (PLoS Medicine editorial 2010). Attention was drawn to, in some instances, substantive empirical differences, lack of transparency regarding methodology and difficulties associated with accessing data used by different research groups, and the unfortunate consequence of uncertainty for the countries reliant on estimation of international agencies (Graham and Adjei 2010).

Infant mortality also has a prominent place within the discipline of demography. This is because the probability of dying between birth and age one is critical to the calculation of the life expectancy from birth apart from being an important indicator in its own right.
1.2 Childhood mortality in South Africa

In South Africa, the 1996 Census and the 1998 Demographic and Health Survey were the first enquiries since the implementation of democracy in 1994 to report estimates of childhood mortality for the nine newly demarcated provinces. Evident from these enquiries was that since 1992 South Africa experienced a rapid increase in under-five mortality presumed to be mostly due to the effects of a severe HIV epidemic (Nannan, Timaeus, Laubscher et al. 2007). The eventual policy response was the introduction of the prevention of mother to child transmission programme (PMTCT) in 2002 and as the programme gained momentum, it was also important to monitor the impact of the intervention on mortality.

Although it was to be expected that AIDS would result in an increase in child deaths, empirical data were lacking to assess the extent to which the improved death registration had contributed to an apparent increase in the numbers of reported deaths (Darikwa and Dorrington 2011). Assessing the trend in rates was further complicated by two other health interventions influencing child mortality, namely, the pneumococcal and rotavirus vaccines introduced in April and August of 2009 respectively. Against this backdrop there was a dire need to monitor child mortality trends over this period, although in fact this period was marked by a prolonged absence of reliable empirical information.

The subsequent failure of the 2001 Census and the 2003 DHS to yield any reliable estimates of child mortality prompted a plea to government for accurate data to monitor progress towards the MDG 4 and that the model-based approach informing South African policy was subject to much uncertainty (Bradshaw and Dorrington 2007).

1.3 Challenges in the measurement of childhood mortality

A review of under-five mortality statistics in South Africa (Nannan, Dorrington, Laubscher et al. 2012) concluded that although changes in legislation had led to significant progress in the registration of child deaths since 2000, the extent of under-reporting still rendered the quality of the statistics inadequate for calculating mortality rates directly. Quite aside from the under-reporting of deaths, no evaluation of the completeness of birth registration, equally important to the measurement of childhood mortality, had been conducted.

The need to monitor progress toward meeting the health targets of the MDGs has focussed international attention on the capacity of individual countries to count vital events. Between 1995 and 2000 it is estimated that only 30% of births and 26% of deaths
were registered worldwide (Mahapatra, Shibuya, Lopez et al. 2007). The overwhelming failure to capture these data occurs in developing nations which for various reasons do not have adequately developed administrative systems able to capture all births and deaths exhaustively and have not sufficiently informed communities of the benefits of civil registration nor the legal requirement and probably have not provided suitable enough incentives to encourage registration.

Infant and child mortality rates are ideally estimated from data on births and deaths because a comprehensive civil registration system has the potential to provide a continuous longitudinal record of annual births and death trends facilitating the provision of population counts by age and sex. Like most low- and middle- income countries, South Africa does not have a sufficiently developed CRVS to allow the direct estimation of child mortality rates, necessitating the use of complementary data sources and enhanced methods of analysis to assess mortality rates. Despite significant improvement in the registration of births and child deaths since the new millennium (Darikwa and Dorrington 2011), at the end of the MDG project South Africa’s CRVS is not yet complete.

An analytical framework by Graham and colleagues for measuring maternal mortality in developing countries (Graham, Ahmed, Stanton et al. 2008) usefully outlines the data sources and analytical methods. This can be adapted for the measurement of childhood mortality to highlight the different approaches to measuring child mortality. The framework shown in Figure 1-1 illustrates that in contrast to other developing countries, South Africa has a variety of population-based data as well as several other data systems in the health sector to provide information on child mortality trends. The framework also provides an overview of the opportunities for empirical measurement, direct and indirect methods of estimation, and alternative estimation methods, demographic modelling and regression techniques for producing child mortality estimates. This thesis therefore aims to investigate what is the best way to monitor child mortality in upper-middle income countries using South Africa as a case study.

1.4 A framework for evaluation
An evaluation framework for improving the quality and use of birth, death and cause-of-death information as the criteria to assess the overall functionality of CRVS and the quality of other data sets will be used to help answer this question. An adaptation of the Health Metrics Network framework (WHO 2010) proposed by Joubert and colleagues (Joubert, Rao, Bradshaw et al. 2013) contains four key components and criteria believed to underpin all aspects of civil and vital registration are as follows:
• **Generalisability** informed by criteria on (1) coverage and (2) completeness

• **Reliability** informed by criteria on (3) epidemiological consistency and (4) temporal consistency

• **Validity** informed by (5) content validity, (6) use of ill-defined and non-specific codes and (7) use of age and sex improbable classifications

• **Policy relevance** informed by (8) timeliness and (9) availability of sub-national data.
Figure 1-1: Analytical framework - data sources, methods of analysis and synthesis for measuring child mortality

Adapted from Graham, Ahmed, Stanton et al. (2008)
1.5 **Aims and Objectives**

The overall aim of this research is to assess and document how childhood mortality can be best monitored in upper-middle income countries, using South Africa as a case study. The rapid changes in the components of under-five mortality due to high HIV prevalence rates is particular to this setting, while limited resources is common to most countries. The specific objectives of the research are to:

1. Assess the strengths and weaknesses of full birth history data in terms of estimating the level of and trend in mortality, in particular considering the bias in the data and how this has been affected by South Africa’s HIV epidemic.
2. Assess how census questions (concerning the population count, the summary birth history, deaths in the household and the preceding birth technique) perform in monitoring the level and trend of childhood mortality.
3. Assess the strengths and weaknesses of health and demographic surveillance data for monitoring the level and trend of childhood mortality.
4. Assess the civil registration and vital statistics system as a tool for monitoring childhood mortality by evaluating the following criteria:
   a. the completeness of birth registration
   b. the completeness of death registration
   c. the coherence of the age pattern of childhood deaths over time.
5. Derive own estimates and assess estimates derived from the different data sources against selected independent estimates.
6. Identify areas of improvement to ensure childhood mortality can be measured in future.
7. Assess what lessons countries with limited resources could learn from this analysis.

---

1 In terms of the aims and objectives, the thesis does not intend to describe or evaluate mortality, fertility or any aspect of data quality by racial or ethnic group.
1.6  Chapter outline
This thesis begins with a review of the literature (Chapter 2) followed by a description in Chapter 3 of the methods used. The investigation utilises the data sources from the analytical framework (Figure 1-1). Chapter 4 assesses the strengths and weaknesses of the full birth history for measuring the level and trend of mortality, including the bias for mother’s survival. Chapter 5 assesses how the census questions concerning the population count, the summary birth history, deaths in the household and the preceding birth technique perform in measuring and monitoring the level and trend of childhood mortality. Evaluation of the role of Health and Demographic Surveillance Sites for monitoring the levels of childhood mortality is the focus of Chapter 6. Assessment of the CRVS system to monitor childhood mortality by evaluating the completeness of birth and death registration and the coherence of the age-pattern of the age pattern over time is the subject of Chapter 7. The assessment of provincial CRVS systems to monitor childhood mortality, particularly at provincial level, is explored in Chapter 8. The final chapter discusses the implications of the findings of the previous chapters, integrates their conclusions into an approach for monitoring childhood mortality provincially, and identifies important areas of future research.
2 MEASUREMENT OF CHILDHOOD MORTALITY:
A REVIEW OF THE LITERATURE

2.1 Introduction
This literature review aims to establish the background and theoretical basis for the thesis. It covers seven areas. The first sub-section describes the global context of measuring childhood mortality and its prominence as a universal indicator of health status and social and economic development. A critical review of the available sources of data that inform childhood mortality is followed by the methodological challenges of measurement. The trends in childhood mortality derived from empirical data and statistical modelling for South Africa precedes an overview of evaluations of the civil registration and vital statistics for monitoring childhood mortality. The literature on examination of the age-pattern of childhood mortality is reviewed before the chapter ends with the conclusion, which establishes a motivation for the rest of the thesis.

2.2 Trends in child mortality
Globally child mortality rates from the 1970’s throughout the twentieth century (Ahmad, Lopez and Inoue 2000) and continuing in the new millennia have shown substantial decline (UN IGME 2015; Wang, Liddell, Coates et al. 2014). This decline has been attributed to improvements in a host of factors known to impact on child health such as improvements in nutrition (Gwatkin, Rutstein, Johnson et al. 2007), housing and environmental conditions and the development of health services (WHO 2003). In particular, the child survival approach adopted by UNICEF in 1982, which promoted monitoring of growth (G), oral rehydration (O), breast-feeding (B), immunization (I), birth spacing/family planning (F), food supplementation (F) and the promotion of female education (F) in developing countries known as GOBI-FFF, is credited with having a great impact. Although there is consensus about the decline and factors, which influenced this trend, there is disagreement even between the UN agencies that generate the information – WHO, UNICEF, the World Bank Group, UN Population Division – and the Institute for Health Metrics and Evaluation (IHME) as to the actual number of deaths under the age of five occurring annually. Estimates by the IHME, for example, show that the global average declined from 110 per 1,000 live births in 1980 to 72 in 2005 (Murray, Laakso, Shibuya et al. 2007) but they point out the 1983-2005 UNICEF estimates are consistently higher in comparison. Although the difference in the two sets of global estimates narrow over time, in some years the over-estimation amounted to
more than three million child deaths per year. The authors also note differences between 1999 and 2002 where UNICEF trends show an increasing number of deaths, in contrast to the analysis of IHME that found deaths to be declining. To some extent this is due to uncertainty of the basic data during this era and to some extent differences in the methods used by the two groups as well as differences in access to certain country data sets (Bradshaw, Chopra, Kerber et al. 2008). Lack of coherence between the estimates produced by the UN agencies individually, lead to the establishment of the UN Inter-agency Group for Child Mortality Estimation (IGME) in about 2004.

Methodological revisions for both IHME and UN IGME have occurred throughout the period under review. Prior to 2013, UN IGME used Loess regression to estimate child mortality, however, since 2013, the agency has used a Bayesian B-spline bias-adjusted model (B3 model) to estimate infant and under-5 mortality rates (Alkema and New 2014). IHME on the other hand have used Gaussian process regression modelling in their global assessment of child mortality estimation (Rajaratnam, Marcus, Flaxman et al. 2010) and up until their more recent estimation process (Wang, Liddell, Coates et al. 2014). Post 2012 (approximately) the estimated trends produced by both agencies for South Africa have become very similar in both shape and level (UN IGME 2015; Wang, Liddell, Coates et al. 2014).

Accompanied by the uncertainty in the estimates due to differences in the methodology, variations have also been observed within the general trends, and analysis of DHS data sets show that particular countries in sub-Saharan Africa experienced stagnation and reversals of the downward trend during the 1990’s (Garenne and Gakusi 2006; Rutstein 2000). Although the evidence points to deaths due to HIV/AIDS as being responsible for the observed trends in many countries, particularly in south-eastern Africa, these studies also show that many cases of increasing childhood mortality in African countries were due to severe political or economic crises, and to the disruption of health services associated with these crises (Garenne and Gakusi 2006).

An initial finding by Adetunji (2000), who used survey data from twenty-five countries to investigate the relationship between adult HIV prevalence and under-five mortality was that countries with over five percent prevalence showed increased under-five mortality over time whereas countries with lower adult prevalence showed decreases in mortality. However, the under-five mortality trend pre-HIV in Malawi was described as very high accompanied with high HIV prevalence, yet declining child mortality (Artzrouni and Zaba 2003). Hence, while the association of maternal HIV
status with child mortality is convincing, teasing out the competing risks in the trend is not a trivial task with some studies reporting an impact of vertical transmission, while several other studies have noted that increasing mortality may be attributable to a range of factors other than HIV. Rutstein (2000), identified exogenous influences such as the resurgence of malaria, deterioration of immunisation and health service utilization. In addition, endogenous factors relating to the health status of the mother indirectly through maternal illness (Ng’weshemi, Urassa, Isingo et al. 2003) have also been identified as contributing causes.

South Africa is one of the countries in southern Africa that since the early 1990s has witnessed a reversal in the gains made in child mortality. It has a “generalised” HIV epidemic characterised by an aggressive increase in HIV prevalence among antenatal clinic attendees from 0.7 per cent in 1990 to a plateau of around 30 per cent since 2006 (Department of Health 2015). The impact on child mortality trends appeared at an aggregate level as early as 1992 from the 1996 Census (Nannan, Timaeus, Laubscher et al. 2007) and the 1998 DHS (Department of Health 2002). Confirmation of this reversal in trend by other data sources representative of select rural populations in geographically diverse corners of the country is abundant (Garrib, Jaffar, Knight et al. 2006; Kahn, Garenne, Collinson et al. 2007; Muhwava, Nyirenda, Mutevedzi et al. 2007). Since these initial findings, research in other settings has shown that much of the increase in childhood mortality can be attributed to mother to child transmission of HIV (Rollins, Little, Mzolo et al. 2007).

In addition, the bias caused by the high correlation of deaths between mothers and their children in high prevalence countries with generalised HIV epidemics has complicated the accurate measurement from retrospective data from census and surveys. Research development in this area has, unravelled the complex “set of interdependent relationships between fertility, mother’s age, stage of HIV infection, chance of mother-to-child-transmission and survival of infected children” (Gregson, Hallett, Kurwa et al. 2009: 1), however, while some researchers argue that an adjustment is necessary (Alkema and You 2012; Hallett, Gregson, Kurwa et al. 2010; Walker, Hill and Zhao 2012), IHME did not correct for this bias (Rajaratnam, Marcus, Flaxman et al. 2010).

Apart from the direct effects of intrapartum, mother-to-child transmission and breast-feeding on child mortality there are other effects that impact through indirect mechanisms. One finding illustrates the relative mortality difference for children born to
HIV positive women as opposed to HIV negative women. A prospective follow-up of Malawian children found significantly higher under-five mortality of children born to HIV infected women, 460 per 1,000 compared to 16 per 1,000 among children born to HIV negative mothers, confirming a very strong correlation between mother and child mortality in high HIV prevalence settings (Crampin, Floyd, Glynn et al. 2003).

This strength of the association between mother and child makes their mortality highly correlated. It is this dynamic that introduces the most serious bias to child mortality estimation for both the summary and full birth history methods. The reason for this is due to the fact that children whose mothers have died from AIDS have a higher mortality (Nakiyingi, Bracher, Whitworth et al. 2003), and this relationship results in an under estimation of child mortality because those mothers are not alive at the time of the survey to report on their children. The bias affects the summary and full birth history methods differently, although adjustments required to correct for the bias in both forms of data have been proposed (Artzrouni and Zaba 2003; Hallett, Gregson, Kurwa et al. 2010; UN IGME 2010a; Ward and Zaba 2008).

2.3 Data sources for measuring childhood mortality

2.3.1 Vital registration

Civil registration and the resulting vital registration constitute a continuous source of monitoring births and deaths over time and provides critical demographic and health information which have the potential to form the backbone of the public health surveillance system. Apart from generating infant and maternal mortality rates and cause of death statistics, a well-developed registration system can be used to monitor teenage fertility rates, low birth-weight rates and the variations between geographical regions (Bradshaw, Kielkowski and Sitas 1998). These data are superlative because annual estimates provide a good source of recent levels and trend information and hence permit the monitoring of short and long term demographic trends (Setel, Macfarlane, Szreter et al. 2007). The usefulness of vital registration data, however, depends on the quality of the information (Mahapatra, Shibuya, Lopez et al. 2007). Notwithstanding the importance of this information for the purposes of demographic analysis and for key functions in population health, these data often suffer from coverage errors where a birth, death, or both can simply go unrecorded; or in the case of death, errors in reporting the age at death may occur. Timeliness of the information, from the processing-stage to publication is also challenging for many developing countries,
however, despite the data imperfections, their utility lies in the fact that even where data are in some manner defective, the possibility exists to adjust and correct the data using alternative data sources such as censuses (Hill, Lopez, Shibuya et al. 2007) and surveys.

Apart from the necessity for governments to count individuals for planning purposes, civil registration brings other valuable benefits to the citizenry such as those pertaining to person-recognition found to influence overall development and social cohesion within societies. Writing about the British mortality decline the historian Szreter (2010) describes beneficial by-products brought about by the development of the civil registration system, such as an historical grounding of communities that played a role in democratizing society and laying down strong traditions of local government in the nineteenth century. He concludes that a comprehensive civil registration system, which establishes the legal identity of individuals, is one of the three institutions necessary for rapid economic growth.

South Africa has quite an unusual past in the area of civil registration. Disjointed legislation affected administrative processes for capturing data, and undoubtedly compromised the quality of vital statistics (Botha and Bradshaw 1985) for many decades, and at various historical stages attempts were made to accommodate societal changes that only served to confuse and complicate the administrative notification process further. Only a brief account of the legislative landmarks leading up to the present day civil and vital registration systems, their requirements, and some reflection on the historical development is given here.

A variety of administrative records of population enumeration in forms such as colonial registers, diaries and journals since the mid-1600s have been used in the past (Khalfani, Zuberi, Bah et al. 2005). Before the formation of the South African Union in 1910, the legislation relating to birth and death registration was different for each of the four colonies, and consequently each geographic territory produced, compiled and published separate reports (Bah 1999).

The establishment of a national statistics office in 1914, however, resulted in a more centralised approach to the collection of vital statistics, with the Births, Deaths and Marriages Registration Act of 1923 going further to ensure a more uniform approach to the registration of vital events. Although the act insisted on compulsory registration for all races in the urban areas and for Africans living in rural areas (where the majority of Africans resided at the time) it was left up to the individual to decide whether or not to ‘volunteer’ this information (Khalfani, Zuberi, Bah et al. 2005). This
and other legislation fragmented the collection of vital statistics along the lines of race, and urban and rural residence, and contributed to what Bah refers to as the “stunting of the development of vital statistics among Africans” (Bah 1999: 46). In addition, the creation of the four independent Homelands and six self-governing territories during the 1970’s further handicapped the way demographic information was collected.

A computerised population register established in 1972 recorded the details of Whites, Coloureds and Indians and maintained a continuous source of records in the database. Attempts to improvement the coverage and content of the population register occurred in three ways. First, in 1986 the registration of Africans through the issuing of uniform identity documents, and subsequently the “Population Registration Act” of 1991 and the “Births and Deaths Registration Act” of 1992 were amended to facilitate more accurate registration of all deaths with the Department of Home Affairs. The repeal of the Population Registration Act in 1991 effectively meant that the Department was no longer required to capture vital events classified by race; however, the act did require the registration of all deaths using the death notification form (B-1663) including medical certification of the cause of death. In settings where no medical practitioner was available for certification purposes, a headman could complete a death report form (BI-1680).

In an ideal world, the calculation of a conventional infant mortality rate from vital registration data is simply a ratio of the number of deaths of children under age one year during a specific period divided by the number of births in the same period. Assuming there is no under-reporting in either the numerator or denominator the same input data can be utilised to derive an accurate estimate of the probability of dying before the first birthday, an essential input to the life table. The derivation of mortality indices for ages between zero and one are more complicated than at the older ages because of the steeply falling force of mortality in the first year of life. This necessitates the probability of dying in the first year of life to be calculated using additional data on deaths under age one to accommodate the various values of $nA_x$, the average length of the interval lived by those who die before reaching age $x+n$. The data required include the calendar month of death, the number of months lived by the infants who died before their first birthday, the number of days lived by infants who died in the first month and the number of hours lived by those who died in the first day.

There is no standard to using these detailed data; rather various formulae calculate the probability of dying before the first birthday. For example, different approaches are
used to produce life tables for the populations of England and Wales on the one hand and the US population on the other (Marszalek 2010). Where vital registration data are defective or incomplete, the approach to assessing the levels of completeness is to compare registered deaths with expected levels from survey or census information (Hill, Lopez, Shibuya et al. 2007; Mathers and Boerma 2010).

2.3.2 Health service statistics
Health service statistics represent another source of continuously collected information collated from a variety of tiers of the health sector. The core limitation related to these types of data is that they are not representative of the total population. Sick children admitted to health facilities are an undefined group not typical of the population at large and therefore introduce bias, although the significance of the bias would depend on the proportion of all deaths that occur in facilities. Hence, a mortality rate reflective of children dying in facilities is not a population-based estimate representative of the general population and not reflective of the ‘true’ mortality rate. However, even though health facility-based data are not suitable for the estimation of mortality of the population as a whole, they can be useful for identifying and monitoring trends and South Africa has several data sources from the health sector. Four sources of facility-based data: The District Health Information System (DHIS) the Perinatal Problem Identification Program (PPIP) and Child Health Identification Program (CHIP), which are voluntary hospital audits, and the National Injury Mortality Surveillance System (NIMSS), are described.

The DHIS was established in 1996 as a routine public health system for collecting important elements of health information, essentially to aid management (Department of Health 2011b). Selected health indicators from the DHIS data were made publically available annually (Health Systems Trust 2005) until the advent of the District Health Barometer in 2005, when “a publication that provides functional information for monitoring progress as a means to help improve health systems organisation and delivery, and in so doing, enhance the quality of and access to primary health care” (Massyn, Peer, Padarath et al. 2015: ii). Evaluation of the DHIS data for purposes of monitoring population health over the past decade, such as assessing trends at a health district level, has seen progress, so that presently the DHIS data is processed into forty-four indicators, representing fifty-two districts (Massyn, Peer, Padarath et al. 2015).
The Maternal and Infant Care Strategies Unit of the Medical Research Council established the PPIP to improve the quality of clinical care and hospital management of neonates by identifying modifiable factors associated with common causes of death. The methods are simple but an essential component of the audit process is staff commitment. Each audit cycle involves the following steps to take the data to action (Pattinson, Woods, Greenfield et al. 2005):

1. The death and underlying cause of death are recorded and a discussion around the modifiable factors within and beyond the health system facilitated.
2. The data from all facilities is synthesised and national priorities are made to reduce death.
3. Recommendations for action are made and implemented.
4. The implementation of recommendations is assessed.

The information is collected as part of a voluntary audit of thus far 164 facilities. For the participating facilities, the PIPP has compiled statistics on stillbirths and neonatal deaths in the six “Saving Babies” reports which have impacted on improvements in programmes and the quality of care in health facilities (Stephen, Mulaudzi, Kauchal et al. 2009). The PIPP initiative has also contributed to the information about avoidable causes of death to still births and new-born babies in the international “Every Death Counts” series of reports aimed at informing policymakers of progress towards the MDG target to reduce under-five mortality (Bradshaw, Chopra, Kerber et al. 2008).

An extension of the PIPP since 2004 lead to the Child Healthcare Identification Program (CHIP) that includes children under 18 years old. The focus of reports is on modifiable factors associated with infant and child deaths occurring in facilities, which are compiled from data collected at ward level (Stephen, Mulaudzi, Kauchal et al. 2009). By 2004, increased participation lead to broader representation covering fifteen hospitals at various levels of paediatric health care in six provinces located in urban, peri-urban and rural settings.

Third, the National Injury Mortality Surveillance System (NIMSS) established as a sentinel surveillance system to collect data on injury deaths. It is currently the only source of regular and detailed information on the causes of non-natural deaths nationally. In 2005, NIMSS data drawn from more than 30 state mortuaries in seven provinces represented approximately 40 percent of the estimated national caseload. The NIMSS covers five of the country’s six metropolitan areas along with several other
major towns and cities and are more representative of urban rather than rural injury deaths (Prinsloo 2007). The NIMSS collates data from standard medico-forensic investigative procedures, including post-mortem reports, police dockets and chemical pathology laboratory results.

2.3.3 Census data
Most censuses conducted decennially with the major purpose of providing a count of the population by age and sex. Censuses carried out at regular intervals will generate data with identical time-reference periods suitable for time-series analysis. South Africa is fortunate in this regard, since democracy as it held censuses in 1996, 2001, and 2011. The 2007 Community Survey substituted for the intended 2006 Census, and the 2016 Community Survey used a significantly larger sample than the annual household surveys in an attempt to collect adequately representative information at municipality level.

In less developed countries, where vital registration systems are inadequate for purposes of measuring child mortality rates directly, a common feature of censuses are the indirect questions included in the questionnaires. “The term ‘indirect’ used to qualify some of the techniques used in demographic estimation has its origin in the fact that such techniques produce estimates of a certain parameter on the basis of information that is only indirectly related to its value” (United Nations 1983: 3). The use of demographic models is commonly associated with these methods, in contrast with ‘direct’ methods, which, do not rely on models.

All three censuses included a summary birth history questions (SBH) including questions on the total number of children ever born to women and the number of children who were alive at the census, often referred to as Children Ever Born/Children Surviving (CEB/CS) data. The appeal of these questions for less developed countries is that it does not rely on accurate recall of dates or ages by the respondent from which to calculate mortality rates. Instead, this indirect technique uses the proportions of dead children reported by the age of the mother to estimate mortality assuming constant mortality and fertility schedules, which reflect conditions of a stable population and a standard life table to derive a set of child mortality estimates at certain reference dates in the past. The overall quality of the 1996 Census data is considered to be better than that of the 2001 Census data for this purpose (Moultrie 2006). Chapter 5 of the thesis examines the usefulness of the responses to the 2011 Census CEB/CS questions to produce internally consistent child mortality estimates in the recent past. The under-five
mortality trend obtained from the 1996 Census provided valuable information about level and trend as well as provincial and racial estimates that were consistent with other data sources (Department of Health 2002; Dorrington, Bradshaw, Johnson et al. 2006; Nannan, Timaeus, Laubscher et al. 2007).

The 2001 Census also asked respondents to report on deaths of any members of the household in the last 12 months. This question has traditionally been used to produce estimates of adult mortality after correcting for completeness of reporting of deaths. Respondents were also asked to report the age and sex of the deceased, whether the death was associated with violence and whether the death was related to pregnancy in the case of women of childbearing age. These data have been found to produce reasonable estimates of adult mortality, but their utility for estimating child mortality is unclear (Dorrington, Moultrie and Timaeus 2004).

In contrast, analysis of the responses to the SBH questions in the 2001 Census yielded implausible estimates (Dorrington, Moultrie and Timaeus 2004) mainly due to serious under-reporting of children ever born by their mothers. The inadequate training of field workers is thought to have been responsible for many of the problems associated with poor data quality.

The most recent survey to have utilised the SBH information is the 2007 Community Survey (Darikwa and Dorrington 2011). Critical assessment of these data found it necessary to correct for errors in the sex ratio of deaths, apply an appropriate adjustment to correct for mother’s survivor bias (Ward and Zaba 2008), and for increasing maternal HIV prevalence over time. Although the full birth history data are preferred (as they allow calculation of period age-specific estimates and are not dependent on fulfilling various methodological assumptions of the indirect method), after careful treatment, the indirect data provided retrospective child mortality trends from 2006. However, direct data are favoured and for example the UN IGME used not to draw on data points derived by the SBH when compiling their country estimates (UN IGME 2010a). However, inputs for the UN IGME latest B3 statistical model (Alkema and New 2014) do include the SBH if only these data are available. For countries with high HIV/AIDS epidemics, only the deaths of children of mothers aged 25-29 in the period most affected by the epidemic are included because this is the closest approximation to \( q_0 \) requiring the least amount of information about the age pattern of mortality.
Survey data

During the 1970’s the World Fertility Survey (WFS) collected data on fertility-related events (Sprehe 1974) using birth histories in an effort to address the dearth of demographic information in the developing world. The International Statistical Institute established the programme in 1972 with funding from UNFPA, USAID and the UK Overseas Development Administration. Conducted from mid-1974 to mid-1977, the programme collected full birth histories permitting regional comparison of fertility dynamics. In some countries women were also administered the summary birth history (SBH) questionnaire providing data for indirect estimation. As the name implies the fertility survey was never intended as a mortality survey, although for the period the results contributed valuable knowledge of child mortality in areas of the world with limited or no information before the survey (Hobcraft 1984).

The series of Demographic and Health Survey (DHS’s) beginning in 1985 succeeded the WFS and broadened the demographic focus by including maternal and child health information. The major areas of interest for the DHS’s were antenatal and maternity care, child morbidity and treatment, child anthropometry, immunization and causes of death. By and large the DHS’s used a live birth history, also called a maternity history (Becker and Sosa 1992), although surveys in some countries such as the first South African DHS in 1998, used the pregnancy history questionnaire. The basic difference being that the latter requires information about every pregnancy outcome; (abortion, miscarriage, stillbirth, or a live birth) whereas the birth history only asks information pertaining to live births. The record of each woman’s pregnancy or birth history aims to provide reliable information about the date of each pregnancy outcome and in the case of a live birth beginning with the first birth, when the birth occurred, the sex, whether the child is still alive, and in the case where the child died, the date of death. Such data permit the study of stillbirth trends, early neonatal rates and the relationship with other components of the infant mortality (Pattinson, Woods, Greenfield et al. 2005). This body of survey information provided an immense resource of evidence related to socioeconomic differentials associated with mother and child biodemographic characteristics of childhood mortality (Boerma, Sommerfelt, van Ginneken et al. 1994). Multiple Indicator Cluster Surveys (MICSs) have been used by international agencies such as UNICEF since the mid-1990’s. Originally, these surveys included only

---

2 The Human Sciences Research Council conducted DHS-type surveys, which collected childhood mortality information between 1987 to 1989, however, these surveys were not included in the international program of DHS’s.
summary birth histories, but more recently, they have included full birth histories, making them more comparable to DHS data.

South Africa is fortunate in the number of surveys, albeit with differing objectives, which have served to provide useful data on child mortality, from full pregnancy and birth histories as well as from the simpler summary birth history questions.

The most important maternal and child health information originates from the international series of DHS’s. Since about 1984, more than 240 surveys in over 85 less developed countries have been undertaken. The DHS’s have earned reputation for collecting accurate, nationally representative information on fertility, family planning and maternal and child health and more recently on gender and HIV prevalence. Typically conducted every five years, these household surveys allow comparisons over time. They collect full birth history information, which allows a direct exposure to risk calculation as described by Rutstein (1984).

Several distortions typical of retrospective data can affect birth history data (Potter 1977). In particular, the inability of the respondent to recall events and their corresponding reference periods accurately and the omission of child deaths, particularly deaths that occur at younger ages, are among the common errors. However, it has been reported (Vallin, Palloni and D’Souza 1990) that such methodological problems have been overcome, to some extent, because DHS’s have evolved to be more probing in design. The well-trained field worker, who is most often a female, seeks to minimise proxy reporting and carefully guide the mother through her reproductive history using a well-structured questionnaire. Hence, two comprehensive assessments of the DHS fertility and child mortality data have reported the data to be of high quality overall (Curtis 1995; Vallin, Palloni and D’Souza 1990) showing this type of survey model well suited to elicit accurate information for estimating fertility and child mortality.

More recently, DHS’s have adapted to accommodate changing data needs of policy makers by including new modules of questions in addition to the traditional core modules. One example is the attempts to capture HIV prevalence information.

The 2008 National Income Dynamics Survey and the 2007 Community Survey did not have the measurement of child mortality as a primary objective. In this sense, they can be viewed as occasional survey opportunities, as opposed to routine data collection. These and other occasional survey data offer the possibility of validating two

---

3 Measure, 2017(https://www.dhsprogram.com/Who-We-Are/About-Us.cfm)
or more sets of estimates in order to learn more about the reliability of the different estimates.

2.3.5 Health and Demographic Surveillance Sites

Another approach to measuring child mortality in settings where vital registration is lacking is the creation of Health and Demographic Surveillance Sites (HDSS), which facilitate longitudinal follow-up of defined populations. Demographic events such as births, deaths and migration are collected intensely by way of interviews conducted by well-trained field workers with members of the area surveillance at regular intervals. Additional health, socioeconomic and programme-specific information for monitoring conditions are also collected. The populations under surveillance are most often rural and homogenous in economic, social and cultural circumstance (INDEPTH, 2002).

The primary purposes of demographic surveillance sites are quite varied. For instance, one of the first demographic surveillance sites established in 1963, Matlab in Bangladesh contributed to the field research program of the International Centre for Diarrhoeal Disease Research and exemplifies the role of a HDSS in programme monitoring. Another example of achievements is the Niakhar study zone located in rural Senegal, which has documented impressive declines in under-five mortality rates over a 37-year period (Delaunay, Etard, Preziosi et al. 2001). Commonly, the identification of causes of death using verbal autopsy are valuable HDSS outputs for the countries concerned. Examples in the recent literature are findings regarding a seasonal association with malaria mortality from seven demographic surveillance sites (Abdullah, Adazu, Masanja et al. 2007) and a ten year cause of death trend from the Manhica site in Mozambique (Sacarlal, Nhacolo, Sigauque et al. 2009) and many others that have identified the measurement of child mortality as a primary output.

South Africa currently has three HDSS in the rural areas of KwaZulu-Natal, Mpumalanga and Limpopo provinces. The HDSS in the Agincourt sub-district of Bushbuckridge situated in the northeast Mpumalanga province, close to the Mozambique border. The purpose of this site is to contribute to the understanding of various health care models. In this instance decentralising health system programmes in maternal health, sexually transmitted infections, child nutrition and mental health to programmes more oriented to a district-based health care model (Kahn, Garenne, Collinson et al. 2007). The area has been under demographic surveillance since 1992, and in 2011 included approximately 16,000 households with a total population of
approximately 90,722 (Shoko, Collinson, Lefakane et al. 2016). It spans 26 villages and is typical of a poor rural setting. There have been 20 census and vital event update rounds, conducted annually since 2000. There have been sixteen census rounds from 1992 up until 2011 mainly collecting data regarding annual births, deaths, migration episodes and antenatal and perinatal health seeking behaviour (Hargreaves, Collinson, Kahn et al. 2004). The cause of death information ascertained using a validated verbal autopsy instrument (Kahn, Tollman, Garenne et al. 2000) has been an important source of information that has allowed investigation of the South African health transition in this population since about 2003.

The Africa Centre Demographic Information System (ACDIS) in northern KwaZulu-Natal is the subject of Chapter 6 of this thesis focusing on the measurement of child mortality utilising ACDIS data, therefore only a brief description is included here.

An important objective of the ACDIS near Hlabisa, monitoring the rapid impact the HIV and AIDS epidemic is having on the demography and health of the population (Tanser, Hosegood, Barnighausen et al. 2008). Subsequent to an initial baseline census of the population conducted in 2000 there have been routine rounds of data collection, which have collected individual and household socioeconomic data, event data capturing births, deaths and migration episodes as well as surveys that have collected information on adult HIV status (Muhwava, Nyirenda, Mutevedzi et al. 2007). Baseline data capturing the socioeconomic conditions of the approximately 80,000 individuals in 11,000 households have highlighted very poor living conditions (Hosegood, Benzler and Solarsh 2005), extensive circular migration (Lurie, Harrison, Wilkinson et al. 1997) and high adult AIDS mortality (Hosegood, McGrath, Herbs et al. 2004).

The Dikgale HDSS in Limpopo was established in 1996 after an initial exercise that mapped eight villages and conducted a baseline census. The estimated population in 2001 was approximately 8,000 and has been stable over the 1996 to 2007 surveillance period (Kanjala, Alberts, Byass et al. 2010). Although child mortality trends from 1996 to 2007 have been estimated, these are presented as adjusted mortality per 1,000 person-years so cannot be included for comparison for two reasons. First, is the difference in the measures of mortality, the probability of dying per 1,000 live births ($n_q$) versus the mortality per 1,000 person-years lived ($\mu_L$). Second, the mortality rate is derived by a

---

1 (http://www.agincourt.co.za/index.php/about)
Accessed 03.03.2017
regression model that adjusts for other covariates. The infant mortality rate in 1998 was 38.9 per 1,000 live births (INDEPTH 2002). The select nature of these rural communities is illustrated by the quite different levels of child mortality observed in the other two HDSS's. Infant mortality in 2000 is much lower in the Agincourt site at 16.9 per 1,000 live births (Kahn, Garenne, Collinson et al. 2007), and much higher in ACDIS at 89.0 per 1,000 live births (Nannan, Bradshaw, Laubscher et al. 2011).

These three South African HDSS's are part of the International Network for the continuous Demographic Evaluation of Populations and their Health (INDEPTH 2002), an amalgamation of nineteen surveillance sites, seventeen of them in Africa. Along with the Manhica site in Mozambique, the three South African sites form the southern African component of the INDEPTH surveillance system.

Overall, the surveillance effort has contributed to a better understanding of socio-demographics in these populations. However, because of the sustained dearth of empirical mortality data for the purposes of constructing African life tables, a valuable contribution has been delivered by the age-specific mortality patterns and the identified clusters provided by INDEPTH (INDEPTH 2004). The extent to which they have been used and accompanying commentary of their suitability and appropriateness, however, remains unclear.

The frequent rounds of data collection characterise event monitoring in these more remote areas and goes some way to ensure recording vital events in the surveillance area that might otherwise be missed by having to report to the provincial Home Affairs offices. Further, the longitudinal nature of the data also means they possess attributes, which allow the study of mortality determinants and differentials over time. Such data are rare. However, the surveillance areas are representative of select settings and therefore the relevance of the findings are limited to those defined populations. Another limitation of this data source is the intensity of the observations makes HDSSs very expensive, which has implications for their sustainability and future.

2.4 Methodological challenges
The methodological challenges associated with the measurement of child mortality are to some extent borne out by the formation of the Inter-agency group, later known as the Inter-agency Group for Child Mortality Estimation (IGME) (Child Mortality Coordination Group 2006). In an effort to improve the consistency of estimates among the UN agencies - WHO, UNICEF, UNAIDS and UNPD - technical collaboration improves transparency of the methods and access to country data was established.
Although methodological challenges exist at most stages of any measurement exercise attempting to capture population trends, the thesis will focus on two methodological issues with potentially sufficient impact to warrant adjustment. These are the effect of mother’s survivor bias and the possible effects of birth transference.

2.4.1 **Effect of mother’s survivor bias on full birth history data**

Vertical transmission results from the close physiological processes between the foetus and mother throughout the gestational period, the consequence of which is a high correlation of mortality between mothers and their children. Other causes of correlation between mother and child mortality include genetic factors affecting both mother and child; other infectious disease transmission from mother to child; the simultaneous death of mother and child involved in the same accident and the death of the mother from any cause could affect the survival of the new-born child. However, these biases are ignored in retrospective demographic surveys, such as the DHSs. The subsequent death of an HIV positive mother indicates a higher risk of seroconversion for the child, thereby increasing the child’s mortality risk. The mortality experience of these children however, will be under-represented in retrospective surveys as the mother would not be alive to report on their behalf. It is precisely this effect, which necessitates an adjustment to the estimates in populations affected by HIV.

The size of the AIDS impact on estimates of child mortality differs according to whether one is using full or summary birth history data to estimate rates and depends on the level of HIV prevalence in the population. This bias is less severe when estimating mortality directly using full birth histories as the method is not reliant on such strong assumptions, is not dependant on the age of the mother and less reliant on births which occurred further back in the past (Mahy 2003).

An investigation to quantify the bias using direct estimation techniques (Artzrouni and Zaba 2003) developed a model to measure the difference in mortality risks between infected and non-infected women in five different scenarios of HIV prevalence and distribution of the disease. It was thus possible to calculate the extent of the bias between mortality risks in the various settings. It’s findings suggested that even in high HIV environments the overall impact of the bias on estimates of child mortality would not exceed five percent in the preceding five years before the survey, although this would increase as the epidemic matured (Mahy 2003). Hence, for instance a model, that estimated the bias 35 years into an epidemic, indicated a seven percent bias in under-five
mortality. This model applied to SADHS 2003 data estimates around five to seven per cent under-estimate of the rate at any stage of the epidemic (Department of Health 2007).

Since this investigation, based on results from simulation models, there have been two other approaches used to estimate the extent of mothers’ survivor bias on child mortality estimates; one using empirical data and the second utilising a simple cohort component projection model in the direct application.

First, empirical data from an open cohort in rural Manicaland in eastern Zimbabwe was used to estimate survivor bias in generalised HIV epidemics (Hallett, Gregson, Kurwa et al. 2010). HIV prevalence in this area decreased from 22 percent to 18 percent between 1998 and 2005, during which three rounds of interviews were conducted. The 2005 birth history data representing the under-5 mortality experience of surviving mothers was adjusted upwards by adding back the mortality experience of the mothers who had died before the survey. The difference between the mortality rates for children born to all mothers and for children born to surviving mothers, as a percentage of children born to surviving mothers, estimates the extent of under-reporting to be 6.7 and 9.8 percent for infant and under-5 mortality, respectively.

Next using additional data generated by a mathematical model (Hallett, Gregson, Kurwa et al. 2010) it was also possible to produce estimates of bias allowing for different epidemic profiles and levels of baseline child mortality. The model provided two important insights into the direct estimates: (a) that the bias increases as the epidemic matures and the time before survey of the estimate and (b) the level of non-HIV child mortality has the effect of reducing the bias.

An alternative approach to estimating the bias used a cohort component projection model (Walker, Hill and Zhao 2012). This involved creating groups consisting of births to HIV negative women, births to HIV positive women not infected and births to HIV positive women who seroconverted. The HIV negative babies were assigned background mortality from model life tables whereas the infected babies were given a probability of dying before age five of 0.62 per 1,000 live births. Child mortality reported by surviving women is then compared to child mortality estimates from all women had they survived to the date of the survey. As with Hallett and colleagues (2010), the magnitude of the bias was found to be dependent on past HIV prevalence and the level of background child mortality before the survey. Also in agreement with the analysis of Hallett and colleagues, these authors (Walker, Hill and Zhao 2012) find
the bias higher than estimated by Artzrouni and Zaba (2003). They estimate the bias to be highest 6-10 years before the survey and predict it will exceed 10 percent in circumstances where HIV prevalence is more that 5 percent.

2.4.2 Effect of mother's survivor bias on summary birth history data

An appropriate method for adjusting summary birth history estimates for mother’s survivor bias has to consider the assumptions of the indirect model. There are three main assumptions underlying the indirect estimation method. The first assumption is that the correlation of the mother and child’s mortality is negligible. Second, that child mortality is independent of maternal age and third, that a specific standard life table, giving an appropriate fit and shape, can describe the mortality of children.

In an HIV environment, violations of each of these main assumptions occur. A breach of the first assumption occurs in settings with high rates of vertical transmission, such as South Africa. The second assumption is also violated, as maternal HIV infection is associated with the age of the mother, with the discrepancy becoming larger as HIV prevalence increases. As far as the third assumption is concerned, up until recently the model-life tables used to describe the mortality of African populations were the Princeton model life-table system (Coale and Demeny 1966); the United Nations model life-table system (United Nations 1983); the African Standard (Newell 1988: 156) and the General Standard (Brass and Coale 1968). A serious limitation of these out-of-date life tables is that they do not depict current mortality patterns of populations affected by AIDS including for young children under-five years of age. The problem with the three systems is that all under-represent child mortality in relation to infant mortality, even before the advent of the AIDS epidemic making it impossible to find a suitable life table to match the empirical data.

Given these violations, initial work by Ward and Zaba (2008) attempted to estimate the bias using a stable population model, which assumed a stationary population and constant HIV incidence over time to estimate the bias at different levels of HIV prevalence. They found that even prevalence levels as low as five percent produced significant downward bias in the estimates. The key factor that determines the size of the bias is the prevalence of HIV, while the underlying mortality determined the importance of the error proportionately (Mahy 2003). In the usual application of the indirect method, only women’s data for the age groups 20-24 to 34-39 are used. However, estimates of HIV prevalence as low as five percent produced significant
errors in child mortality estimates with the two older age groups producing biases over five percent prevalence and in the lower two age groups the bias was apparent at ten percent prevalence (Mahy 2003). Accurate correction factors by age, however, could not be derived, given the unrealistic dynamics of the model, specifically the assumption relating to non-changing HIV incidence over time.

Further development of this original research using a simulation model did provide a means for deriving correction factors by maternal age group utilising information on HIV prevalence and making assumptions about population stability and the epidemiology of a particular epidemic (Ward and Zaba 2008). However, the authors warn that their findings are “subject to a number of qualifications and limitations”. Most problematic is the fact that, the model holds HIV incidence constant over time and second, given the fact that incidence and prevalence are closely linked, implementation of strategies to reduce prevalence will affect incidence and violate this assumption which would result in overestimating the bias and the accompanying correction factors for a given prevalence level.

2.4.3 Birth transference
Birth transference, which is caused by a systematic bias in the reporting of births, can also affect estimates from full birth history data. In countries where HIV prevalence is high, the potential impact of birth transference on the measurement of childhood mortality is of concern. The phenomenon is apparent in DHS’s where full birth questionnaires accompany exhaustive child health questions. Field workers may shift the date of birth of the most recent births and record them as having occurred earlier (before some cut-off date, usually to shift the births in the last five years), thereby avoiding the additional health related questions for children under-five years old. In the case of DHS data, IGME notes that this birth transference tends to be more pronounced for dead children (UN IGME 2010a). If data suffer from birth transference, the resulting under-estimation of mortality may warrant adjustment.

Aside from these two issues, which may produce sizeable impacts, there are also, methodological challenges, which exist at most stages of any exercise attempting to measure population trends. Specifically, countries reliant on survey and census data face challenges related to the quality of the collected data, much of which relates to accurate recording of information by fieldworkers. The training of fieldworkers is therefore central to the quality of the retrospective information about births and deaths of
children. Essentially the accurate recording of date of birth and date of death will lead to accurate mortality estimates. These objectives require strenuous application of the outlined procedural methodology to minimise errors associated with these data as well as management challenges associated with the training and supervision involved with large teams of fieldworkers and supervisors.

Scanning the South African literature concerned with the demographic data required for childhood mortality estimation, dealing with the aftermath of poor data quality emerges as a prominent theme of the Census 2001 (Dorrington, Moultrie and Timaeus 2004), SADHS 2003 (Bradshaw, Laubscher, Dorrington et al. 2004), NIDS 2008 (Moultrie and Dorrington 2009). There is a need for serious review on the part of South Africa’s national implementing agencies to contemplate and investigate the root causes of these failings. Several comprehensive texts, essential to a sound methodological approach such as Shryock and Siegel (1976), the numerous DHS manuals for planning the survey, designing questionnaires, collection and processing the data, give thorough step-by-step instructions.

2.5 **Trends in child mortality in South Africa**

An adequately developed vital registration system serves as a continuous record of statistics on births, deaths, infant deaths and stillbirths. With complete reporting and coverage of the entire population, this information allows for the calculation of birth rates, infant mortality rates and age-specific death rates. As South Africa and many developing countries do not have sufficiently developed civil registration systems, they are reliant on the information from surveys and censuses to estimate levels of childhood mortality.

An historical account of child mortality estimates in South Africa needs to begin with an explanation of how the population was racially segregated under Apartheid as this had direct consequences for how demographic data were collected. Apartheid policies created separate independent ‘Homeland’ states of Transkei, Bophuthatswana, Venda and Ciskei (TBVC), and independent or ‘self-governing’ territories of Lebowa, KwaZulu, QwaQwa, KaNgwane, Gazankulu and KwaNdebele for Africans, and the remainder, representing the more developed areas and most of the arable land of South Africa where the white population resided was designated the Republic of South Africa. These separate geographical areas defined divisions of the population based supposedly

---

5 http://www.measuredhs.com/What-We-Do/Methodology.cfm. Accessed 06.03.2017
on African ethnicity in the case of the Homelands and self-governing territories, and in the case of the RSA based on race. This model dictated that statistics, particularly demographic information be collected and collated by race and geographic area.

This section of the literature review aims to present empirical estimates of infant mortality. The estimates are mostly national, and alternatively black African (children of African descent) where national estimates are not possible, as well as estimates from select populations such as open cohorts under demographic surveillance. This section also includes an account of the derivation of the estimates, the data sources, and some general insight into the problems and inadequacies of the estimates. Estimates of infant mortality from 1979 to 1993 are presented in Table 2-1 where data were collected prior to 1994, before the new democratic system and the movement toward geographical unification and Table 2-2 where data were collected post-1994.

### Table 2-1: Estimates of infant mortality using data collected prior to 1994

<table>
<thead>
<tr>
<th>Author and year of publication</th>
<th>Data source</th>
<th>Reference period</th>
<th>Estimate of IMR per 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimere-Dan (1993)</td>
<td>HSRC 1982 Fertility Survey</td>
<td>around 1979</td>
<td>67.0</td>
</tr>
<tr>
<td>Rossouw and Hofmeyr (1990)</td>
<td>HSRC 1987-89</td>
<td>1978-1982</td>
<td>70.0</td>
</tr>
<tr>
<td>Yach (1988)</td>
<td>see text</td>
<td>1981-1985</td>
<td>94.0-124.0</td>
</tr>
<tr>
<td>Bradshaw, Dorrington and Sitas (1992)</td>
<td>See text</td>
<td>1985</td>
<td>70.5</td>
</tr>
</tbody>
</table>

### 2.5.1 Estimates derived from infant deaths in metropolitan areas

For purposes of estimating infant mortality Yach (1988) accessed information on the total number births, the IMR and total population from Medical Officers of Health from the ten largest metropolitan areas in the RSA (excluding TBVC). This resulted in estimated IMRs between 1981 and 1985 in the range of 94-124. While Yach’s rates based on metropolitan areas seem high, the selective nature of the population the rates are based on could partially explain this. But Dorrington and colleagues (1999) point out an arithmetic error, when correctly calculated his method in fact produces lower rates of 93-107.

### 2.5.2 Estimates derived from HSRC fertility survey

Chimere-Dan (1993) applied the Trussel version of the Brass indirect demographic technique (United Nations 1983) to children ever born and children surviving
information from the HSRC fertility survey of 1982 (Van Tonder 1985). He made an
adjustment for the fact that the data were based on ever-married women because the
method requires information on all women. His estimates of infant mortality declined
from 71 per 1,000 in 1971 to 67 per 1,000 births in 1979. Dorrington et al. (1999) regard
the pattern as highly implausible due to fluctuating rates but also that the rate for girls is
higher than the rate for boys. They discuss a number of possible reasons for this.

2.5.3 The Human Science Research Council Survey 1987-1989
The HSRC conducted a series of Demographic and Health Surveys from June 1987 to
April 1989 (Rossouw and Hofmeyr 1990). The survey design and methodology was very
similar to the international Demographic and Health Surveys run by Macro, which
utilised full birth history information from women of childbearing ages. Infant mortality
rates were calculated directly to be 70 per 1,000 live births for the period 1978-1982. It
appears that these estimates were considered to be reliable as, in 1992, Sadie amended
his estimates derived from a population reconstruction by decreasing his IMR from
1960-1965 on the basis of the HSRC rates as reported by (Dorrington, Bradshaw and
Wegner 1999).

Several of the estimates (those of Chimere-Dan; Bradshaw, Dorrington and Sitas;
and the HSRC) are close to one another. In this regard, Dorrington, Bradshaw and
Wegner (1999) examine the methodologies underlying each of the estimates and assess
their accuracy. They conclude that the best estimates for the period around 1985 are
those from Bradshaw, Dorrington and Sitas (1992) on the grounds that these are
entirely consistent with the mortality underlying Sadie’s re-estimation of the population.

2.5.4 SALDRU Poverty Survey, 1993
The aim of the Poverty Survey was to compile statistics on living standards and
development in South Africa. It was carried out by the South African Labour and
Development Research Unit (Nannan, Bradshaw, Mazur et al. 1998). The survey asked
the simple CEB/CS questions. Mazur (1995) applied the Trussel and Palloni-Heligan
variants of the original coefficients to these data and estimated an IMR of 81 per 1,000
births over the period 1980-1992. These estimates are probably a little on the high side,
due, possibly, to an over-sampling of poor households in the survey.
2.5.5 **Estimates derived from October Household Surveys**

Retrospective data were also collected from the annually conducted October Household Surveys (OHS). Nannan (1998) used the birth history information to apply the CEB/CS technique to the first survey, carried out in 1993, which excluded the former TBVC homelands. The infant mortality rate of 12 per 1,000 births is implausibly low when compared with other estimates presented in Table 2-1.

The data from the 1995 (OHS) was representative of the whole country and was analysed in the same way, but unfortunately also yielded implausibly low mortality estimates when compared with other estimates of infant mortality shown in Table 2-3. Udjo (1997) also utilises these data to calculate child mortality rates and concludes that the quality of the data renders the estimates unreliable, as they are improbably low. He reaches the same conclusion about the 1995 OHS.

The purpose of the October Household Survey was to collect much-needed data on households. The collection of birth history data was an afterthought, not an objective at the planning stages of the survey. Bearing this in mind, the most probable reasons for the failure of the child mortality component was that the detailed questionnaire requires full and accurate recall of births and deaths on the part of the respondent. Given that the birth history was not originally planned, and was not the main purpose of the survey, it is likely the fieldworkers did not receive sufficient training on the questions.

The transition to a new political dispensation included the creation of the government’s post-Apartheid statistical agency, Statistics South Africa, in 1995. Among other changes that influenced demographic information, apart from the removal of official classification by race, there were changes in the collection and statistical compilation of demographic information, the reintegration of the ‘homelands’ into South Africa, and the formation of nine newly demarcated provinces.

The empirical infant mortality estimates shown in Table 2-2 were collected post the 1994 elections. The accompanying text briefly describes the methods used, paying attention to adjustments for known biases in the data and noting innovative procedures employed by the researchers to deal with observed inconsistencies in the data.

2.5.6 **Estimates derived from the 1996 Census**

Together with the 1995 OHS and 1998 SADHS, the 1996 Census was amongst the first nationally representative inquiries conducted under the auspices of the democratically elected government to collect crucial childhood mortality data.
The 1996 census included the summary birth history questions. Using these data, Nannan and colleagues (Nannan, Timaeus, Laubscher *et al.* 2007) found three adjustments to be necessary. The first, relatively minor adjustment, was to correct for non-statement of parity (El-Badry 1961) which isn’t really necessary. The second, was an adjustment needed to correct the data for the excess mortality risk of children born to teenage mothers (Collumbien and Sloggett 2001).

The third, more substantial adjustment was related to the number of reported child deaths and was made based on work done by Moultrie and Timaeus (2003), who argued that the number of child deaths reported in the census was inflated by the misclassification of stillbirths as dead live-births. The question asking about how many children a particular woman had given birth to was asked in a more complex manner than it should have been, with a final phrase in a second set of parentheses that perhaps offset the instruction in the first parenthesis requiring responses to be confined to live births only.

*How many children, if any, has the woman ever given birth to? (live births) (Please include her children who are not living with her and those who have died).*
Table 2-2: Estimates of infant mortality using data collected post 1994

<table>
<thead>
<tr>
<th>National estimates</th>
<th>Data source</th>
<th>Reference period</th>
<th>Estimate of infant mortality per 1,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nannan, Timaeus, Laubscher et al. (2007)</td>
<td>1996 Census</td>
<td>June 1995</td>
<td>56.3</td>
</tr>
<tr>
<td>Dorrington, Moultrie and Timeaus (2004)</td>
<td>2001 Census</td>
<td>no derivation possible</td>
<td></td>
</tr>
<tr>
<td>Udjo (2008)</td>
<td>2001 Census</td>
<td>2001</td>
<td>69.0 for boys /65.0 for girls</td>
</tr>
<tr>
<td>Moultrie and Dorrington (2009)</td>
<td>National income Dynamics Survey</td>
<td>2006</td>
<td>no derivation possible from national data</td>
</tr>
<tr>
<td>Udjo (2008)</td>
<td>2007 Community Survey</td>
<td>2006</td>
<td>56.0 for boys / 53.0 for girls</td>
</tr>
<tr>
<td>Darikwa and Dorrington (2011)</td>
<td>2007 Community Survey</td>
<td>2006</td>
<td>50.9</td>
</tr>
<tr>
<td>- CEB/CS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Previous birth technique</td>
<td>2006</td>
<td>46.0</td>
<td></td>
</tr>
<tr>
<td>- Deaths in the household</td>
<td>2006</td>
<td>53.0</td>
<td></td>
</tr>
<tr>
<td>Statistics South Africa (2015)</td>
<td>Census 2011</td>
<td>2007-2010</td>
<td>44.0, 42.0, 34.0</td>
</tr>
</tbody>
</table>

Estimates from Health and Demographic Surveillance Sites

| Indepth (2002) | Digkale HDSS | Africa Centre HDSS | 1998 | 38.9 |
| Nannan, Bradshaw, Laubscher et al. (2011) | Africa Centre HDSS | 2000 | 89.0 |
| Ndirangu, Newell, Tanser et al. | Africa Centre HDSS | 2000-2007 | 74.1- 24.0 |

On the assumption that stillbirths were reported as dead live births in similar proportions to the 1998 DHS, the analysis concluded that a downward adjustment of approximately 22 percent was appropriate.

Udjo (2005) based his estimates on different assumptions from the method described above. He used parameters for selecting appropriate adjustment factors of $q_x$. 


values from a PhD thesis (Fernandez 1985) making use of an estimate of the singulate mean age at marriage. This adjustment is not employed by other demographers, and is of doubtful accuracy.

2.5.7 **Estimates derived from the 1998 South Africa Demographic and Health Survey**

Due to the close proximity of the date of the 1996 Census and this DHS, valuable insight can be gained from a comparison of the levels attained from these two national surveys. The DHS’s calculate mortality rates using the life table methodology proposed by Rutstein (1984). Investigation of the correlation of rates by province with the census estimates concluded that the direct estimates of childhood mortality were probably only a slight under-estimate of the true rate (Nannan, Timaeus, Laubscher et al. 2007). Appropriate adjustments to both data sets resulted in a convincing trend and minimal difference between the DHS and the census data points.

Considering the value of this type of data, there has been surprisingly little research generated from the 1998 DHS. Although Burgard and Trieman (2006) did investigate the impact in improvement in socioeconomic conditions on child mortality, they used the published five yearly averages (Department of Health 2002) which tend to obscure the increase in mortality due to mother to child transmission of HIV resulting in their main conclusion missing the fact that the IMR, in particular, had increased sharply since 1995.

2.5.8 **Estimates based on the 2001 Census**

There is no consensus among researchers regarding the reliability and usefulness of the 2001 Census summary birth history data. Dorrington with others, (Dorrington, Moultrie and Timaeus 2004) undertook assessment of the data quality for purposes of mortality estimation.

The 2001 Census was the first South African census data was subject to editing and imputation procedures designed to replace all unknown or ‘implausible’ results, and the authors offer useful insight about the impact of editing procedures on the raw data. Apart from the overall finding that, in aggregate the data are of very poor quality, with both the CEB and CS data lacking internal consistency, the main problem with the under-five mortality data was that the number of children born to women of reproductive age was seriously under-reported. The unedited data resulted in “biologically impossible” answers (Dorrington, Moultrie and Timaeus 2004: 43) and
removing those responses made these data inconsistent with the equivalent 1996 Census and 1998 DHS data. The authors therefore concluded that data quality compromised the estimation process to such an extent that no derivation of child mortality was possible.

The only research to have asserted a trend from the 2001 Census is that of Udjo (2008). However, the infant mortality pattern by sex he produced is counter-intuitive. The data points representing the oldest women show identical levels for boys and girls while the most recent points show opposite direction in the divergence, leaving the mortality of boys increasing steeply while that of girls decreases steeply. Apart from identifying this implausible trend, Dorrington and Moultrie also convincingly argue that Udjo’s research is “logically, methodologically and operationally flawed” (Dorrington and Moultrie 2008: 283) raising serious doubt about any of his results and conclusions.

2.5.9 Estimates derived from the 2003 South Africa Demographic and Health Survey
There were numerous problems with this survey. The problems identified (Department of Health 2007) have their root cause in the fact that there was a systematic failure by field workers to capture all the children born in the previous five years, presumably because they were put off by the lengthy questionnaire required for those recent births and concomitant extra work. The clear evidence of “transference” of births, with births in the most recent years being drastically short compared with the period five years earlier.

The infant mortality estimate of 43 per 1,000 live births centres around 2001 and suggests a decrease in infant mortality from the previous survey estimate of 45 per 1,000 centred around 1996, which is improbable given the generalised HIV/AIDS epidemic. However, the 1998 age-specific estimates and the 2003 estimates fifteen years before the survey are inconsistent in level and trend, indicating the 2003 estimates are too low.

2.5.10 Estimates based on National Income Dynamics Survey
The National Income Dynamics Survey is a panel study, which first collected data in 2008. The major findings from the demographic data from the first wave of data collection was reported by Moultrie and Dorrington (2009). The novelty of this survey was that it collected full birth history and summary birth history information. The researchers also attempted to derive child mortality estimates from the deaths in the previous 24 months as reported by households.
Interviews of 9,355 women captured basic details of date of birth and date of death of all their children and identified problems in the data pertaining to dead children including the incorrect recording of their year of birth and their age at death. Unfortunately, the conclusion of this analysis was that it was not possible to derive sensible estimates of the trend or level of child mortality from all these data. Nevertheless, it was possible to assess data from the alternative summarized birth histories for black African children (children of African descent) and this suggested that the under-five mortality was 100 per 1,000 births in 2006 and implied somewhat implausibly that the rate of child mortality had remained stable over the last 10 years. Similarly, the questions about deaths in the household produced problematic data. Unfortunately, the ages of almost twenty percent of deaths were either not recorded at all or recorded as unknown, which resulted in a greater degree of uncertainty regarding the estimation of infant and child mortality than would usually be the case for estimating rates from deaths reported by households.

2.5.11 Estimates based on the 2007 Community Survey
Published to date are two sets of child mortality estimates from the 2007 Community Survey. First set of estimates derived by Darikwa and Dorrington (2011), made extensive use of the child mortality data by using three approaches, the CEB/CS method, the Blacker and Brass (2005) variant of the previous birth technique and child deaths in the last year reported by households to estimate mortality around 2006. The CEB/CS data resulted in higher mortality for girls than for boys. The questions asked whether the child lived at home, the sex of the child and for totals, and the authors examined these data for internal consistency. Several inconsistencies became apparent and thus they decided to use only data with no inconsistencies. This decision had the most important effect, as these data no longer displayed a higher mortality for girls and analysis continued with the reduced data set. Thereafter the data were adjusted for the known bias resulting from the non-independence of the mortality of the mother and her child using an adaptation of the method proposed by Ward and Zaba (2008).

Next, Darikwa and Dorrington used the PBT to estimate an IMR around 2006. Again, the data displayed girl/boy irregularities, even using the reduced data set. Finally, life table measures of infant and child mortality were calculated directly from the deaths reported by households to have occurred in the past 24 months, and the births for a particular year from the ASSA2003 projection model. Central death rates $\frac{1}{100}M_x$ for ages $x$
and \( x + , \) where \( x = 1, 2, 3, 4 \) were derived by dividing the number of deaths of children aged \( x \) last birthday at deaths by the mid-year population aged \( x \) last birthday.

The second set of estimates were calculated by Udjo (2008) who applied the CEB/CS method to these data and used the INDEPTH model life table (INDEPTH 2004) in an effort to allow for the impact of HIV on the estimates. Udjo also found that female mortality was higher, but did not validate the data for internal consistency. Instead, noting that the mortality for girls was consistently higher than that for boys, which was considered atypical of the South African population, he adjusted the Community Survey data based on his analysis of the 2001 Census. First, the mortality trend for boys was derived by extrapolating the 2006 level to 2001. The trend for girls was then estimated as the \( \alpha \) values (level) for boys less the average difference in the boy/girl childhood \( \alpha \) values derived from the reports of older mothers in the 2001 Census.

2.5.12 **Statistics South Africa**

These estimates appear in the South Africa country report on the progress towards fourth MDG - the reduction of child mortality by two-thirds from 1990 to 2015 (Statistics South Africa 2010). Infant and under-five mortality rates were calculated from the 2001 Census and 2007 Community Survey SBH data, although no method is given. The baseline under-five mortality is taken from the 1998 DHS as 59 per 1,000 live births for the year 1998, although this is actually a five-year average representing the period 1994-1998, i.e. centred on the year 1996, and the infant mortality rate is estimated from the 2001 Census at 54 for the year 2001. In 2007, the under-five rate is reported as 104 and the infant rate as 53 per 1,000 live births, estimated from the 2007 Community Survey.

First, it would have been more appropriate to give the infant mortality as 45 per 1,000 and the under-five mortality rate as 59 per 1,000 from the 1998 DHS (Department of Health 2002: 100). This is standard practise to ensure consistency and regularity in the data. A more serious problem, however, is the peculiar relationship between the infant and under-five mortality rates. The estimates imply an increase of 1 death per 1,000 in the IMR, and an increase of 45 deaths per 1,000 in the under-five mortality rate, hence stating that “the infant mortality rate appears to have remained more or less the same despite the upward trend in the under-five mortality rate” (Statistics South Africa 2010: 61) ignores the implausibility of these trends.
2.5.13 **Summary of post 1994 estimates**

Reviewing the estimates in Table 2-2 estimates reveals some important issues. The first point highlighted is that between 1995 and 2007 South Africa undertook numerous data collection exercises designed to provide child mortality estimates.

Central to the issue of differences between estimates is that analysts will rarely adopt identical strategies to manipulate the same data, nor even apply the same method in exactly the same way. Only after the methods are studied, does it becomes apparent that even where estimates derived from the same data source are of similar magnitude, for example, Udjo (56 for boys and 53 for girls) and Darikwa and Dorrington (50.9) both using the SBH method, quite different strategies were adopted in the two instances. The similarity in the numerical value of the estimates can also be misleading as visual inspection of the trend plots reveals striking differences in the estimates. Hence, it becomes clear that procedure is by no means fixed and researchers exercise personal judgement when reaching decisions pertaining to the data and the course of analysis. Nevertheless, there are limitations to the set of options available to the analyst and standard practice regarding an acceptable course of action. It is therefore worrying, for example, that Udjo attempts to produce estimates of mortality rates from the 2001 Census data given that Dorrington and colleagues argue that the data are too poor for such a task.

An intriguing point is Darikwa, Dorrington use more data from the Community Survey than just the normally used SBH data, utilising three techniques to derive three sets of estimates, which reveals potential in data not sufficiently exploited. Other analysts do not report multiple methods but it would provide a mechanism for verifying internal consistency and thereby instilling greater confidence in the estimates. Further, although the three estimates from the Community Survey are numerically fairly close, concern regarding their reliability is raised due to an improbable relationship between the rates for each sex, an impossible relationship between infant and under-five mortality resulting from Census 2001 and the trend implied by the Community Survey data (Statistics South Africa 2010).

2.5.14 **Africa Centre Health and Demographic Surveillance Site**

There are three sets of estimates from the Africa Centre demographic surveillance area presented in Table 2-2 covering about eight years of surveillance. In 2000, a baseline fertility survey was conducted in the area to estimate fertility and child mortality. These retrospective data were analysed to give a twenty-year trend of period child mortality.
rates using the same methodology as the Demographic and Health Surveys (Rutstein 1984). Using this method, the infant mortality rate was 89 per 1,000 live births in 2000.

Next, three years of data collected on quarterly household visits for the period 2000-2002 were used to estimate mortality ratios equivalent to a conventional IMR but not strictly comparable with synthetic cohort probabilities of dying (Garrib, Jaffar, Knight et al. 2006). These longitudinal data are used to estimate a conventional IMR of 59.6 per 1,000 live births in 2000.

The last set of published estimates from ACDIS provide a trend from 2000-2006 (Ndirangu, Newell, Tanser et al. 2010). The authors report under-two mortality by year of birth cohort from prospective data. For purposes of comparison, the neonatal and post neonatal rates as reported in the article were combined to give an overall infant mortality rate using the same method as reported in the paper. Noteworthy from this article are the reported declines in all age groups. The infant mortality by year of birth cohort declined from 74.1 in 2000 to an improbable 27.4 per 1,000 in 2006.

2.5.15  Agincourt Health and Demographic Surveillance Site
The demographic surveillance site in Agincourt for the period 2000-2001 reports an IMR of 36.9, presumably attained by calculating the ratio of deaths under age one to the live births by calendar year (Kahn, Garenne, Collinson et al. 2007). The IMR increased to 39.0 in 2004 and was 39.1 in 2009 (Kahn, Collinson, Gómez-Olivé et al. 2012).

2.5.16  Dikgale Health and Demographic Surveillance Site
For purposes of comparison with the other estimates presented here, the 1998 IMR of 38.9 for the Dikgale demographic surveillance site is used (INDEPTH 2002). The mortality rates were calculated using person-time denominators from episode data.

2.5.17  Inconsistency between Demographic Surveillance estimates
The three estimates from the Africa Centre HDSS are quite varied. Around 2000, the highest estimate of 89 per 1,000 was derived from the full birth history information (cross-sectional); 74 per 1,000 is the estimate from the year of birth cohort analysis and 59 per 1,000 was estimated as a conventional IMR (longitudinal). Although each is an estimate of the IMR, the nature and specific calculation of the data leads to a different estimate depending on which measure is selected. These subtle differences are usually not described although it is important to understand them as the directly calculated
estimates are less prone to measurement error and have the potential to strengthen and complement country estimates (Sankoh 2010). Investigations of child mortality originating from HDSS’ have focussed on comparison between estimates from surveillance areas with national and regional DHS estimates to evaluate the performance of the surveillance systems and assess how representative they are of the population at large. These studies conclude that despite the different nature of the data sources, HDSS’s yield estimates that are similar to measures from cross-sectional surveys and they are an adequate tool to report demographic measures (Nhacolo, Nhalungo, Sacoor et al. 2006) and risk factors of child mortality are also broadly comparable (Hammer, Kouyat’e, Ramroth et al. 2006).

More insight into the actual measurement of child mortality rates is offered by a comparison of longitudinal and cross-sectional approaches described in an Ethiopian HDSS (Byass, Worku, Emmelin et al. 2007). They highlight three points, which can influence or bias the mortality rate. First, lower figures are reported where longer recall is required in a DHS. Second, the DHS approach tends to over-sample younger children, since the younger mothers can only contribute younger children and lastly, the numbers of annual deaths are more stable in the DHS data due to averaging over the whole region or country. These studies have investigated how comparable vital rates are from the two sources, but no one has reported a finding of discrepancies in child mortality rates within the same HDSS area such as those observed in the ACDIS.

2.5.18 Synthesised and model-based estimates
In countries lacking reliable data, model-based estimates have filled the gap in tracking progress towards MDG 4. Essentially the two methods of modelling used to produce infant and under-5 mortality trends are statistical modelling and demographic projection modelling located in the analytical framework (Figure 1-1). Statistical modelling is used by the UN agencies (WHO, UNICEF, UNAIDS) who since 2004 have formed the UN Inter-agency Group for Child Mortality Estimation, the Institute for Health Metrics and Evaluation (IHME) and others for example Garenne and Gakusi (2010). The second method is demographic projection used by the UNPD and ASSA, and more recently the THEMBSISA model, to derive estimates consistent with population dynamics.

In the context of monitoring South Africa’s progress towards MDG 4 the Every Death Counts report (2008) reviewed various models at the peak of the epidemic, shown in Figure 2-1. IGME compiles a set of country estimates derived from a mixture
of country based data sources including complete and partial vital registration systems, censuses, surveys and databases maintained by the UNICEF and WHO (UN IGME 2007). The basic approach IGME used was to derive one set of nationally representative child mortality estimates using Spline curve fitting, and later using Loess regression to fit regression curves of the observed data points against time, by taking the log of the under-five mortality rate against time. These data were then extrapolated to the required reference period and other child mortality indicators such as the IMR are extrapolated using model life tables (UN IGME 2010b). Since 2013, the UN IGME published results produced by a Bayesian B-splines biased-adjusted model, referred to as the “B3 model” (Alkema and New 2014). Compared with the previous Loess regression estimation, improvements include better accounting of data errors such as sampling and non-sampling errors and biases; better capture of short-term fluctuations in the under-5 mortality rate and a better performance regarding short-term projections (UN IGME 2017).

![Figure 2-1: Model estimates of under-five mortality trends in South Africa, 2005](source)

It is important to note that IGME include an adjustment for birth transference and in countries with HIV epidemics, a correction for mother survivor bias (Alkema and You 2012). The magnitude of the bias depends on how much excess under-five mortality due to HIV will go unreported by mothers. Since the information required to estimate the bias is fairly detailed - the distribution of births of HIV+ women by the duration of their infection for instance, which are seldom available, IGME have adopted
a methodology with simple assumptions to adjust HIV related mortality for survey data since 1980 (Alkema and You 2012). The assumptions relate to the distribution of births to HIV+ women and the duration of their infection rates of vertical transmission and survival times for mothers and babies since the birth (see section 2.4.1 methodological challenges).

After correcting for survivor bias, the UNAIDS-WHO estimates of under-five deaths due to HIV/AIDS are subtracted from all the survey data points during the epidemic period. The non-HIV deaths are derived by fitting either Spline or Loess regression curves to the data (individual country decision). Lastly, the HIV/AIDS deaths are added back to the non-HIV regression trend. This approach was used to estimate under-five mortality in 17 countries with population prevalence over five per cent at any time throughout the epidemic in the 2011 round of published estimates (Alkema and You 2012).

An alternative approach proposed by the IHME (Murray, Laakso, Shibuya et al. 2007) seeks to find a statistical model to estimate past under-five mortality and forecast to 2015 for 172 countries. This method uses statistical compilation of survey and vital registration data, which IHME argue has the advantage over other approaches of being reproducible. They merged available datasets and applied Loess regression to fit a polynomial, a method of weighted least squares that gives more weight to points nearer the fitted trend and less weight to points further away. According to these estimates in 2005, South Africa had an under-five mortality rate of 77 per 1,000 (95% CI 73-81). Subsequently the IHME has updated this work (Rajaratnam, Tran, Lopez et al. 2010) with estimates produced using Gaussian process regression. The authors assert, “the technique has better out-of-sample predictive validity than previous methods and captures uncertainty caused by sampling and non-sampling error across data types”. The estimated probability of dying between birth and age five for the 2000-2010 decade was 57.5 (95% CI 37.0-37.6), and for the decade 2010 is 50.9 (95% CI 43.2-60.3), considerably lower than the earlier estimate. The IHME has subsequently conducted several updates of global estimates, for instance the revisions by Lozano and colleagues (Lozano, Wang, Foreman et al. 2011) and by Wang and colleagues (Wang, Liddell, Coates et al. 2014).

Turning to the ASSA2008 model, the Actuarial Society of South Africa has developed this suite of AIDS and demographic projection models since about 1996 in

---

7 http://www.actuariasociety.org.za/Societyactivities/CommitteeActivities/DemographyEpidemiologyCommittee/Models.aspx
order to model the South African HIV/AIDS epidemic. These demographic cohort component projection models essentially generate demographic outputs necessary for population projections – population size, age-specific non-AIDS and AIDS deaths and other key demographic indicators based on numerous data sources including vital registration, survey and census data and data from a behavioural heterosexual transmission model to derive new cases of HIV infection. The Thembisa model (Johnson 2014) developed to replace the ASSA model to include the impact of the PMTCT and ART interventions on demographic outcomes, is an integrated demographic and epidemiological model of the South African HIV/AIDS epidemic.

Figure 2-1 presents the various model estimates compared with South Africa’s MDG 4 under-five mortality target of 20 per 1,000. The projections all show the initial decrease in the under-five mortality rate but by 1995, the estimates ranged from 52 to 68. Thereafter, most of the models show an increasing trend although the trajectories differed. The 2010 IHME estimate stands out as being quite different from the other models, reaching a level of 37.3 per 1,000 live births in the year 2000 while the other estimates range from 69 to 76 per 1,000 live births. This is because the IHME did not adjust the VR data points for incompleteness, although by 2010, the IHME estimates are similar to the ASSA estimates at 51 and 49 per 1,000 respectively. In contrast to the variability displayed by the models around 2005, as under-five mortality peaked, the comparison of model estimates in 2015 (Figure 2-1) show that there is far less difference between the estimates, although the IHME model is much higher than the other three before 2005.

Comparison of the methodologies used by IGME and IHME (Alkema and You 2012) found that country level differences between the two sets of estimates are due to different trend fitting procedures and adjustment procedures. IGME, for example, sought consultation with local experts and incorporated this feedback into their analysis, which the IHME does not. The comparison also found that differences in the data used caused greater difference in the estimates than different trend fitting methods did. However, the major difference between the estimates produced by IGME and IHME in the case of South Africa is because IHME does not adjust the observed under-five mortality for mother’s survivor bias in countries with generalised HIV epidemics, while IGME adjusts for this bias.
Despite the number of empirical enquiries and model-based procedures, there has been in dearth of information for monitoring progress towards MDG 4 during 2000 to 2010, and much of the available information was for long periods contradictory. In many cases, data in question were unusable due to poor quality and in other cases the data produced implausible results because they were inconsistent with previous estimates. A series of articles addressing the problems of the measurement of health indicators provides an evaluation framework to apply to low-to-middle income countries. The articles have highlighted important issues towards the endeavour to produce accurate indicators and have identified key criteria for the robust measurement of health indicators. These include, that providers of estimates should agree on set standards of transparency, scientific rigour and accessibility (Boerma, Mathers and Abou-Zahr 2010). The IHME are of the opinion that five distinct areas contribute to the production of health indicator measurement (Murray and Lopez 2010). These include the development of tools and instrument innovation; capacity to collect, process and archive data and the establishment of norms and standards for health indicators. In terms of the analysis of data the authors highlight the fact the estimates produced from different sources may contradict one another, in that some data may be missing, bias in survey instruments exists, and some groups may be excluded or under-represented by administrative data systems, therefore analysis using the best scientific methods is...
required. These, along with other suggestions will be useful while considering suitable evaluation framework and criteria for a way forward for South Africa.

2.6 **Evaluation of civil registration and vital statistics**
Assessing data quality using completeness, coverage and the quality of information on the causes of death as evaluation criteria, Mathers and colleagues (2005) reported that at the end of 2003, out of 115 countries, death registration data for only 64 countries were considered complete (i.e. at least 90 per cent complete). Combining the three criteria South Africa were ranked with 28 other countries, as having ‘poor’ quality vital registration data with completeness below 70 per cent and/or over 20 per cent of ill-defined codes on death notification forms. This rating, however, is largely determined by the high proportion of ill-defined codes, as estimates of completeness, based on death distribution methods (Dorrington and Bradshaw 2011) estimate adult registration to have been nearly 90 per cent by 2001. Completeness of registration of deaths under age five can only be estimated using child mortality levels gauged from a nationally representative census or survey (Hill, Lopez, Shibuya et al. 2007; Mathers, Ma Fat, Inoue et al. 2005), however comparison with local level with DSS results can also be informative (Garenne, Collinson, Kabudula et al. 2016).

2.6.1 **Assessment of civil registration and vital statistics as a tool for measuring childhood mortality**
Assessment of the quality of South African CRVS (Joubert, Rao, Bradshaw et al. 2013) for the period 1997 to 2007 showed that the proposed method to evaluate the criteria of epidemiological consistency, were not suited for South African data aggregated by age due to the high prevalence of HIV/AIDS deaths. The impact of the HIV epidemic on South Africa’s mortality structure since the early 1990’s, caused a distinctive, temporal feature in the age-pattern of deaths whereby the population experiences two age-peaks, the first, due to an excess of deaths under the age of five and the second peak observed in the economically active population. In addition, because the mortality pattern in children is distinctive compared to the rest of the life course, there is justification for investigating children as a unique group. The framework proposed by Joubert et al (2013) includes the criteria that may contribute to determining the best way of measuring childhood mortality through the CRVS.
2.6.2 The assessment of completeness of births and deaths

The literature documents less research devoted to the measurement of the completeness of birth registration than of the registration of deaths, but the methods that have been used to assess the completeness of births and deaths can be grouped into three categories:

1. Statistical methods based on capture-recapture (Sekar and Deming 1949); survey questions which ask the mother if the birth or the death has been registered and possibly to produce the certificate as proof of registration (UNICEF 2013; Williams 2014) in the case of births, and in the HDSS (Garrib, Jaffar, Knight et al. 2006) where no proof of the death certificate was sought.

2. The 2011 GHS asked questions about the prevalence of birth that had been registered, however, no proof of the birth certificate was sought.

3. Comparison with estimates of the number of births and deaths based on, for example:
   - an independent estimate of deaths from a survey (Darikwa and Dorrington 2011),
   - an independent estimate of births based on age-specific fertility rates and the female population of reproductive age (Nannan, Bradshaw, Mazur et al. 1998),
   - demographic models assuming a stable population (Williams 2014), and
   - population projection models (Bradshaw, Chopra, Kerber et al. 2008; Dobbie, Masebe and Nhlapo 2007; Stats SA 2013c).

In South Africa the National Income Dynamics Survey (SALDRU 2008) asked direct questions about birth registration and found that 11% of children under the age of three did not have a birth certificate, however, no proof was sought in the survey. Another analysis, that of the General Household Survey by the Children’s Institute (Berry and Dawes 2013), suggests that in 2011, 11% of children under the age of three had not had their births registered, although no method is provided.

However, Williams (2014) used survey data to derive empirical estimates of fertility in an urban area of Nigeria by employing demographic models and comparing the implied number of births with a) self-reported registration and b) registration on seeing the birth certificate. This study highlighted wide differences between the two
approaches, with 36% completeness when requesting the birth certificate compared to 91% based on self-reports, clearly indicating the unreliable nature of self-reports.

Lastly, in a recent report “Every Child’s Birth Right” evaluating individual country and regional trends in birth registration, UNICEF relies on Stats SA’s annual birth registration report and claims that in 2011 South Africa achieved 95% completeness of births registered within the first year of life (UNICEF 2013).

2.7 The age-pattern of mortality within childhood

The first large multi-country evaluation of childhood mortality was conducted with comparable national surveys, initially of 41 countries included in the World Fertility Survey (WFS) running from mid-1974 to mid-1977 (Rutstein 1985). The findings highlighted vast differences in the level of childhood mortality between countries and regions of the developing world (Rutstein 1984).

Up until this time, little research had focussed on the attributes of an age-at-death analysis, but the broad consensus was that the risk of mortality declines with age below age five; the risk being highest in the first hours, days, weeks and months since birth. Although exceptions had been noted about a decade before the WFS findings, a detailed description of the age-specific mortality in a demographic surveillance site in rural Senegal between 1962 and 1965 (Cantrelle 1972) reported an extreme exception to this broadly accepted age-pattern of mortality. In this case, the mortality rates in the first and second years of life were the same, and were due to an unusual rise in mortality between the fifth and tenth months of life. This unusual age-pattern was later confirmed with another Senegalese population under demographic surveillance in Ngayokheme, in the west of the country (Garenne 1982).

In further research into the variation in the age-pattern of childhood mortality employing the Princeton families of model life tables (MLTs), Cantrelle confirmed his observations of rural Senegal with small select populations in the Gambia, Burkina Faso, Guatemala and the Punjab in India (Cantrelle, Diop, Garenne et al. 1986). Interestingly he attempted to understand the deviations in the age-patterns of populations in the developing world from those represented in the MLTs and therefore representative of mainly Europe in the eighteen century, and concludes that the differences are due to cause of death structures and environmental factors relating to climate and geography.

Subsequently, the study of differentials in childhood mortality, accompanied by the endeavour to understand the factors determining deaths during infancy as opposed to deaths having a slightly later onset beyond age one to age five, became a focus of
enquiry not only in demographic surveillance sites (Pison and Langaney 1985) but also in the comparative analysis undertaken during this time, formulated important conclusions (Blacker, Hill and Timaeus 1985). First, they showed that high death rates beyond infancy, which had thus far only been observed in sub-national populations, were also observed in the expanded set of WFS’s and were therefore nationally representative. Second, their investigation of African countries showed that although the pattern of high child mortality relative to infant mortality was prevalent in many West African countries, this was not the case for countries in the rest of the continent. Third, although childhood mortality patterns in sub-Saharan Africa demonstrated the most diversity, none of the one-parameter life table systems provided a satisfactory representation of the age-pattern to those found in the empirical data. This conclusion has far-reaching implications for the summary birth history technique used extensively to estimate childhood mortality in the developing world, because the childhood mortality trends derived from the indirect estimates are most sensitive to the age-pattern (Blacker, Hill and Timaeus 1985).

Notwithstanding this apparent weakness in the CEB/CS method, researchers using the WFS data and the ever-growing body of DHS’s undertook further analysis. The next set of childhood mortality statistics came from a review of DHS’s in 28 countries between 1985 and 1989 and corroborated the range of mortality level found by the WFS (Sullivan, Rutstein and Bicego 1994). In addition to describing differences in overall levels, the authors also highlighted and focused analysis on the range in the mortality by age-interval (Sullivan, Rutstein and Bicego 1994). They reported that although sub-Saharan countries displayed the highest mortality compared to the other regions, the ratio of $4q_1$ to $5q_0$ in relatively low mortality countries such as Botswana, Zimbabwe and Kenya fell within the range of values represented in the Princeton MLTs. However, in another group of eight sub-Saharan countries, with relatively high mortality, the ratio was above that of the North Family, representing the highest values. Sri Lanka and Brazil were the only other countries that fell outside of the model values; the ratio for Sri Lanka being well above the North pattern, which is the highest and the ratio for Brazil falling beneath the East pattern, which exhibited the lowest ratios. The investigation of DHS data provided further evidence of relatively high concentrations of child deaths compared with the Princeton models. Furthermore, Sullivan and colleagues reached the same conclusion as Blacker and colleagues (1985) regarding the experience of sub-Saharan countries a decade earlier, pointing out that because countries in this
region exhibited the greatest amount of deviation from the Princeton model-values, and were most affected by data problems, the application of the model age-pattern was challenging.

Little was known about South African demography at this time, primarily due to the campaign of international isolation beginning in the early 1960’s and the UN call for the academic boycott in 1980 (Beaubien 1982). South Africa’s inclusion in the international series of DHS’s occurred in 1998, after the onset of the HIV/AIDS epidemic, which was known to increase the mortality of children under-five years of age through MTCT.

The SADHS data were included in an analysis of the age-pattern of childhood mortality in sub-Saharan Africa by Jasseh (2003). Of South Africa he concludes (Jasseh 2003: 142) that “the 1998 SADHS dataset outlines a childhood mortality pattern that has transformed from one very close to the Princeton ‘West’ to ‘North’ and back to ‘West’ as overall childhood mortality started increasing in the ten year period prior to the survey”. Given the reversal in mortality revealed from the DHS data, these findings signify the detection of a changing age-pattern associated with a change in the level of mortality using this methodology.

The most recent investigation into the behaviour of the age-components of childhood mortality utilised data from the Human Mortality Database (HMD), which consists of high quality VR data from mainly economically more developed countries, and WFS, DHS and HDSS data for countries not represented in the HMD (Guillot, Gerland, Pelletier et al. 2012) investigated the age-pattern in sub-Saharan Africa using the UN and the Princeton MLT’s. Historical evidence from Spain is remarkable, portraying one hundred years of data up until 2009 and clearly illustrates the propensity for countries to change families as mortality declines.

Apart from confirming the high values of $q_1$ relative to $q_0$ in western African countries, their methodology also detected this pattern in eastern and central Africa. Unfortunately analysis of the pattern in South Africa was limited to data from the Agincourt HDSS, and although the estimate of $q_0$ is low compared to many other sub-Saharan populations, it presents values above the range of $q_1$ represented in the North family. Regarding the impact of the HIV/AIDS epidemic, the authors state that although the epidemic has probably contributed to higher ratios than represented in the Princeton models, HIV/AIDS does not seem to have substantially altered pre-existing patterns.
A novel approach taken by Rao and colleagues (2011) sought to predict the distribution of causes of the under-five deaths using statistical prediction models to determine the proportionate mortality of four broad-cause groups from the levels of under-5 mortality using data from the WHO mortality database. Although South African data were not included in the database, the South African 2005 VR data were compared with predicted cause proportions to assess the performance of the model. A reasonable fit was achieved although the observed proportion of other causes, which include the ill-defined, was higher than that predicted by the model. The novelty of this approach lies in the idea of investigating changes in the cause of death structure related to changes in the level of mortality and the age and sex distribution of the deaths. The models were developed to explore the relationships between the proportionate mortality and using the following combinations of covariates; $5q_0$; the ratio of $1q_0$ to $5q_0$; the ratio of $4q_1$ to $5q_0$, the male to female ratio of $5q_0$. The broad-cause of death groups analysed were perinatal, diarrhoea and lower respiratory infection (LRI), congenital and ‘other’ which includes all other causes of death. In addition, the pattern of causes was assessed against the ratio of the neonatal rate (NMR) to $5q_0$.

Figure 2-3 illustrates the trends in the cause of death groups as under-5 mortality declines from 100 per 1,000 to below 20 per 1,000. Rao and colleagues (2011) conclude that the under-5 mortality is associated with changes in the cause composition. There is a strong, inverse relationship between deaths due to infections and deaths due to congenital disorders as overall mortality declines. This is also consistent for the trend in deaths due to perinatal causes as under-5 mortality declines to a level of 20-30 per 1,000 but thereafter, the proportion of perinatal deaths also declines while the proportion of other causes increase. As the overall mortality rate declines, perinatal causes play a more prominent role and then decrease as congenital causes make a larger contribution. In addition, the authors observe that at high levels of $5q_0$, the NMR/ $5q_0$ ratio changes little, however, the ratio increases to almost 60% when $5q_0$ fall into the 30 and 70 per 1,000 range. Rao and colleagues (2011) describe this as the compression of the under-five mortality into the neonatal period, which occurs before the under-5 mortality falls below 30 per 1,000.

---

8 The WHO mortality database comprises vital registration cause-specific mortality data since 1950.
2.8 Conclusion

Measurement of infant and child mortality is ideally achieved through the civil and vital registration systems if there is complete reporting of births and deaths and if coverage extends to the entire population. The lack of adequate registration systems in many less developed countries of the world has necessitated alternative ways for assessing levels of child mortality. Over the past four decades, demographers have developed alternative methods and data sources to provide estimates (Hill, Lopez, Shibuya et al. 2007: 1726).

This review of the literature has provided background and context to the various sources of data available to estimate child mortality in South Africa, and the performance of the various analytical methods utilised to generate empirical estimates. It has also drawn on the trajectories of estimates based on regression models and those based on demographic projection models. It is evident that South Africa has a wealth of demographic and health information, and there is scope for further investigating the best way to monitor childhood mortality.

The review has highlighted that South Africa has faced the same challenges as other low-to-middle income countries in monitoring the level of child mortality and experienced some demoralising failures in collecting good quality census and survey data. On the other hand, South Africa witnessed success in the area of vital registration.
particularly the improvement since 2001, suggesting that South Africa may be on a cusp of being able to use registration data to estimate levels of childhood mortality.

By critically appraising the sources of data and the methods of measurement and by using relevant assessment criteria, this thesis aims to review, in a resource-limited setting such as South Africa, which is the best way in a to monitor childhood mortality.
3.1 Overview of the methodologies
This chapter describes the methods for measuring child mortality in South Africa employed using the various sources of data (cross-sectional surveys and censuses, longitudinal demographic surveillance and the civil registration and vital statistics system) in the following chapters. The sources of child mortality information and assessment of the various approaches, in accordance with the schema depicted in Figure 1-1, are described in later chapters where they are further assessed.

The methods include indirect demographic techniques and direct measurement. A shared feature of most of the data sources is that they often rely on reports from the mother, or guardian, which in most cases means the proxy reporting emanates from a woman probably quite familiar with the child. Although all data sources rely on retrospective reporting, demographic surveillance differs in that the recall period is considerably shorter compared with the recall period required by any other approach. However, with the exception of the derivation of mortality from deaths in the household, all the approaches will be subject to measurement error due to selection bias of the women reporting about their children, and hence methods to adjust for such biases are also covered in the chapter.

The final set of research questions are concerned with assessing the vital registration system as a tool to measure child mortality from a sub-set of information pertaining to registered births and the registered deaths under the age of five in terms of completeness, epidemiological and age coherence.

The chapter begins by defining measures, terms and calculations.

3.2 Measures of child mortality: definitions, terms and calculation
It is useful to define the various measures of child mortality, as they are closely related and often confused. Although methods for calculating direct measures of mortality are covered in many basic demographic texts, for instance Shryock and Siegel (1976), in this thesis the notation will follow that used by Preston, Heuveline and Guillot (2001). To illustrate the relationship of child mortality indices
Table 3-1 presents the definitions of three direct measures.
Table 3-1: Direct methods for calculating child mortality rates

<table>
<thead>
<tr>
<th>Approach</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality ratio</td>
<td>The number of deaths to children under age 12 month during a particular time period is divided by the number of births in the same period for the IMR, and the number of deaths of children under age five years during a particular period divided by the number of births for the same period for the U5MR.*</td>
</tr>
<tr>
<td>True cohort life table</td>
<td>The number of deaths to children under age 12 months (or five years) among a specific cohort of births is divided by the number of births in that cohort.</td>
</tr>
<tr>
<td>Synthetic cohort (or period) life table</td>
<td>Mortality probabilities for single-year age segments based on real cohort mortality experience combined into a common age segments (e.g. infant and under-five mortality).</td>
</tr>
</tbody>
</table>

Adapted from source: UN IGME (2007).
* Although not a standard measure, it is an acceptable approximation if the number of births is not changing rapidly.

An overview of the data sources used in the thesis, the type of data, the count and the method of analysis employed are shown in Table 3-2. Two important caveats that concern the validity of the results obtained from all South African Census data presented in the thesis are:

1) Estimates of population, births and deaths have been adjusted to correct for the census undercount according to the Post Enumeration Survey (PES). This affects comparisons of, for example, estimates of the completeness of birth and death registration. If the undercount is an overestimate, then derived estimates of completeness will be underestimated and if the undercount is an underestimate, the derived completeness will be overestimated. In terms of the model-based comparisons undertaken in the thesis, the models used by, for example, Stats SA, UNPD, ASSA and Thembisa all assume or project population numbers; however, they differ in extent to which these number match the census. The thesis does not investigate this potential uncertainty.

2) Confidence intervals are not provided because the calculation is complicated by the Post Enumeration Survey adjustment and the fact that these adjustments are not specific to the estimates of births and deaths.
Table 3-2: Data sources, type of data and count and analytical method

<table>
<thead>
<tr>
<th>Data source</th>
<th>Type of data</th>
<th>Count</th>
<th>Analytical method</th>
</tr>
</thead>
<tbody>
<tr>
<td>DHS 1998</td>
<td>Pregnancy history</td>
<td>1,808</td>
<td>Abridged life table</td>
</tr>
<tr>
<td>DHS 2003</td>
<td>Birth history</td>
<td>621</td>
<td>Abridged life table</td>
</tr>
<tr>
<td>Community Survey 2007</td>
<td>Number of women (15-49)</td>
<td>13,316,488</td>
<td>Derivation of the number of births</td>
</tr>
<tr>
<td>General Household Survey 2004-2012</td>
<td>Average children &lt;1 year of age with medical aid</td>
<td>128,578</td>
<td></td>
</tr>
<tr>
<td>Census 1996</td>
<td>SBH</td>
<td>CEB 19,190,319</td>
<td>Indirect estimation 1q_0 and 5q_0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS 17,178,885</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Number of women (15-49)</td>
<td>10,938,843</td>
<td>Derivation of the number of births</td>
</tr>
<tr>
<td>Census 2001</td>
<td>Number of women (15-49)</td>
<td>12,642,137</td>
<td>Derivation of the number of births</td>
</tr>
<tr>
<td>Census 2011</td>
<td>Population count (0-10)</td>
<td>11,419,429</td>
<td>Reverse survival</td>
</tr>
<tr>
<td></td>
<td>SBH</td>
<td>CEB 21,023,578</td>
<td>Indirect estimation 1q_0 and 5q_0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS 19,649,929</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Household deaths</td>
<td>54,579</td>
<td>Direct estimation 1q_0 and 5q_0</td>
</tr>
<tr>
<td></td>
<td>Number of women (15-49)</td>
<td>14,423,495</td>
<td>Derivation of the number of births</td>
</tr>
<tr>
<td>Vital registration births 2011</td>
<td>Unit record data</td>
<td>911,353</td>
<td>Completeness of registration</td>
</tr>
<tr>
<td>Vital registration deaths 2011</td>
<td>Unit record data</td>
<td>37,603</td>
<td>Completeness of registration</td>
</tr>
<tr>
<td>District health information system 2011</td>
<td>Provincial births</td>
<td>EC 118,873</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FS 51,676</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GT 212,208</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>KZN 200,637</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LP 129,736</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MP 78,103</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NC 22,375</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW 61,226</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WC 95,030</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SA 969,864</td>
<td></td>
</tr>
<tr>
<td>Demographic surveillance Site 2000-2007</td>
<td>Full birth history</td>
<td>Deaths 2,832</td>
<td>Abridged life table</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exposure 34,150</td>
<td>Poisson regression</td>
</tr>
<tr>
<td></td>
<td>Episode data 2007</td>
<td>Deaths 102</td>
<td>Abridged life table</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exposure 9,866</td>
<td></td>
</tr>
</tbody>
</table>
3.3 The life table

The life-table is a central tool in the measurement of childhood mortality. A life-table derives from a set of age-specific mortality rates that can be used to calculate other measures of mortality such as the under-five mortality rate. The central death rate, denoted by $n m_x$, is the “true” death rate experienced by individuals aged $x$ to $x+n$ years. The age-specific central rate estimated from observed data will be denoted by $n M_x$ when it applies to a specific period. The age-specific death rate of a cohort will be denoted by $n M_x^c$.

Observed period death rates, $n M_x$, are generally calculated as the ratio of the number of deaths occurring to children between ages $x$ and $x+n$, over a given period, to the number of person-years lived by the population in the same age group over the same period and corresponds to the life-table measure $n L_x$. The denominator is often estimated as the mid-period population multiplied by the length of the period in years. In the situations where the population experiences in and out migration, it may be necessary to calculate the person-years exposed using the migration information. Thus

$$n M_x = \frac{n D_x}{PYE},$$

where $n D_x$ is the number of observed deaths aged between $x$ and $x+n$ years and PYE is the exact person-years lived by individuals in this period.

In contrast, the ratio could be a “cohort” ratio based on the number of deaths occurring over the next $n$ years to the group of children who were aged $x$ exactly at the beginning of the period divided by the number of person-years lived by that cohort over the $n$ years. Unless otherwise specified, period measures are presented below.

Most often the central death rates values, $n m_x$, are converted to another life table measure, the conditional probability of someone aged exactly $x$ dying before reaching age $x+n$ which is denoted by $n q_x$. A probability measures the relative frequency (chance) of death over an interval of age among all those alive at the start of the age interval. Thus,

$$n q_x = \frac{\text{the number of deaths in a cohort between age } x \text{ and age } x+n}{\text{the number exposed to risk at exact age } x}$$

and it can be approximated from $n m_x$ in two ways:
(a) \[ nq_x = \frac{n \cdot m_x}{1 + (n - n_{ax}) \cdot m_x} \]

where \( n_{ax} \) is the average number of person-years lived in the age interval by those who die in the interval. The \( n_{ax} \) values differ from population to population according to the level of mortality, but are typically less than \( n/2 \) for very young ages. A suitable set of \( n_{ax} \) values can be chosen or derived from empirical data, see \( (\text{Preston, Heuveline and Guillot 2001}) \).

(b) \[ nq \approx 1 - \exp(-n \cdot m_x) \]

which is based on the assumption that the instantaneous force of mortality is constant over the age interval of \( x \) to \( x+n \) with an average of \( \bar{M}_x \). Using either formula (a) or (b), the conditional probabilities can be estimated using the observed \( \bar{M}_x \).

Although the conventional infant mortality rate (IMR) is the ratio of deaths to children under one year of age in a period divided by the number of live births in the same period, i.e.

\[ \text{IMR} = \frac{\text{deaths under age 1 in the period of investigation}}{\text{the number of live births during the period of investigation}} \]

it closely approximates the probability of dying in the first year of life, \( 1q_0 \).

Mortality in the first year can, however, also be investigated in smaller age intervals, most commonly the neonatal rate (NMR) for ages <28 days and post-neonatal rate (PNMR) mortality rates for ages 1-11. These childhood mortality rates are defined as follows:

\[ \text{NMR} = \frac{\text{deaths < 28 days of age in the period of investigation}}{\text{the number of live births during the period of investigation}} \]

and

\[ \text{PNMR} = \frac{\text{deaths 1 – 11 months of age in the period of investigation}}{\text{the number of live births during the period of investigation}} \]

* The denominator should exclude neonatal deaths, but since these are relatively low, this is considered to be a reasonable approximation.
Chapter 6 deals with longitudinal mortality surveillance using data from a demographic surveillance site. The age-specific probabilities of dying, $nq_x$, were calculated using equation (b), with the exposure days cumulated for each age interval $x$ to $x+n$ and period, and converted into person-years of exposure (PYE) for the age groups: <28 days, 1-11 months, 1-2 years, 2-3 years, 3-4 years and 4-5 years old.

These rates were converted into conditional probabilities of dying, $nq_x$, and then combined to produce childhood mortality indices $q_0$, $q_5$, for each calendar year to represent period mortality as follows:

$$\hat{q}_0 = 1 - (1 - q_0)(1 - q_1)(1 - q_2)(1 - q_3)(1 - q_4).$$

and

$$\hat{q}_0 = 1 - \left(1 - \hat{m}q_0\right)\left(1 - \hat{m}q_1\right)\left(1 - \hat{m}q_2\right)\left(1 - \hat{m}q_3\right)\left(1 - \hat{m}q_4\right),$$

where $m$ is age in months.

3.4 Full birth history

The full birth history (FBH) is a key tool used to measure age-specific child mortality in resource poor settings. The questionnaire is administered to a representative sample of women of reproductive ages, which guides them to recall all of their births, or in the case of a pregnancy history questionnaire, asks about all pregnancies and any subsequent deaths of their children, systematically enquiring about sex, date of birth and date of death of each child. An example of such a questionnaire, the SADHS 1998 pregnancy history questionnaire, is given in Appendix 1. Chapter 4 assesses the strengths and limitations of the full birth history approach utilising national surveys, while the following sub-section of this chapter describes the method used to calculate age-specific mortality rates from such data in detail.

3.4.1 Cohort-period life table calculations

Macro International the agency formally responsible for the Demographic and Health Survey programme, periodically assess the quality of full birth history information collected in DHS's. Such reports offer an explanation (diagrammatically) of the method of how each child’s recorded data is cumulated with that of other children if they experience a common age that spans the same calendar period simultaneously (reported
as exposure). The Figure 3-1 shows the Lexis diagram representation of the age-cohort probability of dying which is accompanied by the following explanation.

The exposure to risk recorded from a full birth history questionnaire translates into a life table measure of mortality, most often, the probability of dying. There are two approaches, the true cohort life table and the synthetic cohort life table. Retrospective birth history information provides the required data for both approaches, which involve an exposure to risk calculation where the exposure can be specified by any two out of three possible variables related to time: the child’s birth cohort; the child’s age when exposed, and the period of exposure. Figure 3-1 illustrates three birth cohorts (A, B or C); aged between \( a_l \) to \( a_u \) in the time period of exposure \( t_l \) to \( t_u \).

The limitation of the true cohort life table approach is that all children are required to experience the full exposure \( a_l \) to \( a_u \), therefore they would have to be born at least 12 months before the survey to meet the required exposure to risk. This inclusion criterion in effect results in exclusion of children even slightly less than 12 months old. Therefore, cohort B represents a completed year of exposure before the survey and is illustrated in Figure 3-1.

![Lexis diagram](image)

**Figure 3-1: Lexis diagram representing component death probabilities**

*Source: Rutstein and Rojas 2003*

“Component death probabilities of three cohorts are defined as children born between dates \( t_l \) to \( a_u \) and \( a_l \) to \( t_u \) (cohort A) \( t_l \) to \( a_l \) (cohort B), and \( t_u \) to \( a_u \) (cohort C). In the component death probability relating to the period between \( a_l \) and \( a_u \) between time \( t_l \) and \( t_u \), cohort B is entirely included whereas cohorts A and C are partially included. Hence, the component death probability for this age interval is calculated as an average weighted according to the exposure experienced by the three cohorts." A STATA programme for doing this is included in Appendix 3.
The other alternative is the synthetic cohort approach - the technique adopted by Demographic and Health Surveys and the method applied to the ACDIS full birth history data. In this instance, the mortality probabilities are based on the same mortality experience as the true cohort but the mortality risks of smaller age segments can be combined into more standard age intervals. This demonstrates partial exposure between the period \( a_i \) and \( a_j \) during time \( t_i \) and \( t_j \), represented by birth cohorts \( A \) and \( C \) and necessitates an assumption regarding the length of partial exposure up until the end of the age interval for which survival status is unknown. Macro International assumes that the mortality of partially exposed cohorts is well represented by calculating one-half of the total exposure and one-half of the deaths they have contributed to a particular period (Curtis 1995; Rutstein and Rojas 2003; Rutstein 1984). Thus, a decision is required regarding the length and number of age intervals, which largely depend on how the information was recorded.

The children born close to the survey date where the time-period ends with the interview date present an exception of this exposure calculation. In this case, the numerator of cohort \( C \) will include all of the deaths between \( a_i \) and \( a_u \) as a fraction of the deaths will have occurred to this cohort between \( a_i \) and \( a_u \).

3.5 **Summary birth history**
Indirect questions minimise the respondents’ reliance on dates and exact ages hence, indirect estimation requires the inclusion of special questions in a questionnaire. Summary Birth History (SBH) surveys have been utilised extensively in countries with inadequate vital registration systems as a fairly simple addition to a survey questionnaire to estimate infant and under-five mortality (United Nations 1983) by means of an indirect method, the so-called children ever born/children surviving (CEB/CS) method first developed by Brass (Brass and Coale 1968).

Although the exact wording used in each census or survey questionnaire may differ slightly, generally the question asks each woman between the ages of fifteen and fifty “how many children have you borne, and how many children are still surviving?” Appendix 2 shows the relevant questions, from the 2011 South African Census.

These data on the proportion of dead children, tabulated according to the age of their mother are used to estimate conventional life table indices of childhood mortality. The thesis applies a variant of the original method utilising coefficients developed by Trussel (United Nations 1983). Proportions of those children born alive who have died before the date of the survey, by the age of the mother are converted to probabilities of
dying since birth by adjusting for the timing of fertility using a set of multipliers

determined by the average parities of young women in the first half of their

reproductive lives. The probabilities of dying over various age ranges yielded by data on

the children of mothers of different ages are converted into a common index, of the

probability of a child dying before their fifth birthday using Brass’s relational logit

system and an appropriate standard life table, which can be selected from a wide range

of options.

There are three main assumptions underlying the indirect estimation method.

First, it is assumed that the correlation of mother and child’s mortality is negligible.

Second, that child mortality is independent of maternal age and third, that a specific

standard life table, giving an appropriate fit and shape, to the mortality of children can

be found. However, in an HIV environment, a violation of each of these main

assumptions occurs and therefore necessitates adjustment to the estimates, explained

next in the chapter.

3.6 Adjustment for mother’s survivor bias

Although all methods for estimating child mortality from reports based on the mothers’

recall about the survival of her offspring are subject to slight selection biases the one

with greatest impact comes from a generalised HIV epidemic (Walker, Hill and Zhao

2012). South Africa’s rapidly increasing HIV epidemic since the early 1990’s has

implications for the treatment of full and summary birth history data sets analysed.

There are two methods of adjustment available for direct estimation. The first

originally proposed by Hill (Walker, Hill and Zhao 2012) and incorporated as part of the

IGME methodology for estimating child mortality and the second approach proposed

by Hallett (Hallett, Gregson, Kurwa et al. 2010), will both be applied and the results

compared. The summary birth history information is corrected using an adjustment

proposed by Ward and Zaba (2008) and the adjustment for changing prevalence

proposed by Darikwa and Dorrington (2011).

3.6.1 Adjustment of the full birth history data for mothers’ survivor bias

Reporting on the bias between 1998 and 2005 in rural Zimbabwe, Hallett, Gregson,

Kurwa et al. (2010: 761) note that “it is clear that by this stage of the epidemic serious

bias had been introduced to child mortality estimates, however, due to the complex set

of inter-dependent relationships between fertility, mother’s age, stage of HIV infection,

the risk of mother-to-child transmission and the survival of infected children, the
magnitude of the bias is not easy to determine”. Simply put, the adjustment hinges on estimating the difference between two mortality estimates: the ‘true’ child mortality rate based on the reports of all mothers, crucially including women who will die during a specified period due to HIV-infection, and the mortality based on the responses of women alive and interviewed in the survey.

The IGME approach requires relatively little input data: distribution by year before the survey of births and HIV prevalence among pregnant women and modelled values of child mortality, HIV-positive child survival and HIV-positive mother’s survival. The component projection derives estimates of the probability of the mother and child dying for the age cohorts 0, 1, 2, 3 and 4 years from the births of HIV- women

HIV- births from HIV+ women and

HIV+ births from HIV+ women

An estimate of the births of mothers alive at the survey and the births of dead mothers, allows for the probability of dying for the two groups of births. The mortality of the ‘true’ group as a proportion of the reported group gives an estimate of the bias introduced by the absence of mothers reporting about the survival of their children in surveys.

The alternative model based on empirical representation of the mortality difference between two groups of women in rural Zimbabwe between 1998 and 2005 (Hallett, Gregson, Kurwa et al. 2010) requires relatively more input data. The following are required by calendar year: the population of women by age group, age-specific incidence, age-specific fertility, women’s life expectancy at birth, median years lived by HIV negative women and the number of female births from 1921.

3.6.2 Adjustment of summary birth history data for mothers’ survivor bias

Using the Brass method, Ward and Zaba estimated the magnitude of the bias resulting from the absence of mothers’ reports by simulating models of stable populations with constant HIV prevalence over time, with the objective of deriving age-specific correction factors. By representing the true probability of dying before reaching age Z as:

\[ P(Z) = \frac{B(Z)}{B(Z) + H(Z)} \]

Hallett’s model is available from: https://www.sugarsync.com/pf/D666097_644631_772754
Last accessed 19.08.2015
\[ q(Z) = q(Z)^e + n(Z) \]

where \( n(Z) \) is the required correction factor based on the level of prevalence among women of reproductive age (15-49). \( q(Z) \) and \( q(Z)^e \) is the unadjusted Brass estimate of child mortality. The following model can be utilised to estimate correction factors:

\[ n(Z) = a \text{PREV} + b (\text{PREV})^2 + c \text{PREV15} \]

where, \( \text{PREV15} \) is HIV prevalence of women aged 15-19.

To allow for increasing HIV prevalence rates over time instead of them being constant Darikwa and Dorrington (2011) propose, somewhat arbitrarily, scaling the adjustment for each age group of mothers in proportion to the prevalence at the time of birth of the children to the time of the survey. Thus, the proposed set of correction factors for correcting the Ward and Zaba adjustments are used in this thesis.

3.7 The previous birth technique
Originally designed for use in health facilities by Brass and Macrae (1984), the question asks pregnant women at the time of an antenatal care visit or delivery, about the survival of their previous birth. The authors show that the proportions dead amongst the most recent births provide a good approximation of the probability of dying before the 2nd birthday or \( q(Z) \), see (Hill 2013) for the theoretical basis of this.

Extension of the question to censuses and surveys asks women of reproductive age the date of their most recent live birth and whether the child is still alive. Appendix 2 (P-38 to P-41) contains the question as it appeared in the 2011 South African Census. It is possible to calculate an IMR from this information assuming the proportions dead represent about two-thirds of the mortality. However, in many instances the results have proven unreliable for two reasons. Typically, the date of last birth in census enquiries is prone to non-response for a number of reasons (Blacker and Brass 2005) and the census question on births in the last 12 months is notoriously under-reported (Hill 2012). For these reasons they propose a model to investigate the relationship of the deaths occurring to births in the last 24 months (Blacker and Brass 2005).

Using the well understood relationship of mortality risk at the onset of life and the rapid change with age the authors demonstrate quite substantial differences in shape and pattern of early age mortality, but nonetheless show that the relationship is actually quite
robust. They apply their model to two populations Bangladesh and Senegal, which display quite different mortality age pattern in the first five years of life. The former where \( q_0 \) is greater than \( 4q_1 \) and the latter, representative of the opposite scenario of \( 4q_1 \) being considerably higher than \( 1q_0 \).

The proportion of births surviving during the previous 24 months is represented by the relationship:

\[
1 - \frac{2L_0}{2}
\]

Using the above equation, the authors seek to describe the relationship that represents the probability of dying from the proportions dead \( (D) \) in the last 24 months. The model is designed to examine the contribution of infant deaths in a 24 months exposure period and the calculation of an adjustment factor \( \frac{q_0}{D} \) needed to convert the proportions dead into a probability of dying because this measure is more useful.

From this, they prove that the necessary adjustment factor can be derived by varying the alpha and beta parameters often used in demographic estimation where beta measures the change in the level and alpha adapts the corresponding age pattern. The authors find some of the resulting combinations of alpha and beta generated by the models “somewhat unrealistic” and therefore fit their model to the mortality experienced in seventeen different African surveillance sites by calculating \( \frac{q_0}{D} \) values. Analysis of these results show a wide mortality range with \( 5q_0 \) ranging from 32 per 1,000 for the Agincourt site compared with the equivalent measure in Bandafassi, Senegal of 255 per 1,000. The authors show that the parameters of the model fitted to survivorship at ages one and five in about seventy percent of the mortality schedules can be represented by conversion factors ranging from 1.087 to 1.097 with a median value of 1.092, which is thought to be an appropriate factor for South African data.

Two points worth mentioning regarding suitability of the method in the South African context. First, the method is advantageous over the SBH as it enquires about deaths in the recent past so the effect of mothers’ survivor bias will be minimal. Second, the concern regarding the exclusion of births in the case where more than one birth occurs in the previous 24 months, raised by Hill (2012), is also thought to have insignificant impact on South African data given the South Africa’s very long median birth intervals of 59 months estimated with the 1998 DHS data (Moultrie and Timæus 2003).
3.8 **Deaths in the household**

The census question that asks about deaths in the household in the last 12 months is typically used in conjunction with other information on the age distribution of the living population to assess levels of adult mortality referred to as growth balance methods (UNPD 2002). See Appendix 4 for the Census 2011 question.

In a report of the 2001 South African Census Dorrington, Moultrie and Timeaus (2004: 27) caution that “questions of this nature are relatively untried and untested in census globally, let alone those in developing countries”. Rather, they suggest the data are useful for producing comparative estimates. The thesis uses these data to estimate $q_0$ and $q_{0.5}$ in mid-2011 to provide comparative mortality against which to estimate the deaths reported to the vital registration system.

The direct calculation of childhood mortality from the reported deaths and the corresponding enumerated census populations were derived by complete life tables for the children aged 0 to 4 years old. The infant mortality rate, assumed to approximate $q_0$, was calculated using the $n a_x$ value of 0.1571 (personal communication William Msemburi). The central mortality rate, $M_x$, for children between ages $x$ and $x+1$ for $x = 1, 2, 3, 4$ were determined by dividing the deaths of children aged $x$ by the population exposed. For the calculation of population it is assumed that those who die between ages $x$ and $x+1$ do so halfway through the year.

The central mortality rate was then converted to the probability of dying between ages $x$ and $x+1$, for $x = 1, 2, 3$ and 4. The under-five mortality rate was then calculated using the following formula:

$$s q_0 = 1 - ((1 - q_0) \times (1 - q_1) \times (1 - q_2) \times (1 - q_3) \times (1 - q_4))$$

3.9 **Episode data: Individuals, membership, residency and migration surveillance**

The core activity of demographic surveillance is to count people, track and record their movement. As a particular focus of the ACDIS is on migration patterns due to the high prevalence of circular migration, the starting point is establishing eligibility of each household member into the ACDIS cohort. This happens during initial household visits, soon after the baseline census. Fieldworkers go through a list of all household

---

This value was used in the 2nd National Burden of Disease Study (Msemburi, Pillay-van Wyk, Dorrington, 2014).
members, deciding whether each of them is a resident or non-resident of that bounded structure. Essentially the fieldworker and individual consider residency based on three pieces of information as set out in the DSA rules. These are (ACDIS 2008: 41):

- The opinion of the person, or if he or she are not there, the opinion of the other household members;
- How much time the person spends at the bounded structure.
- Whether the person has recently moved into, or moved out of the bounded structure.

Application of the rules illustrated in the ACDIS fieldworker manual (ACDIS 2008) give insight into different scenarios and how a decision is reached. A non-resident is typically a migrant labourer who spends most time outside the DSA, however, it is important to note that the fieldworker’s assessment of an individual’s relationship with the household and its members as well as other considerations also determine residency status.

Once residency is established, a careful record of an individuals’ movement from a bounded structure (out-migration) and to another bounded structure is kept by the ending and starting of residency periods. Hence, migration may redefine individual residency, including the distinction between internal migrations (occurring between bounded structures both within the DSA) and external migration (occurring between structures within and outside the DSA). Although more challenging in highly mobile populations it is imperative to capture accurately individual migration so as to minimise double counting the individual exposure and (or) missing the individual and excluding the individuals’ exposure altogether. Hence, despite the fact that residency and migration are very closely related it is essential to specify and separate them for numerator and denominator accuracy in the eventual calculation of rates. In other words, a prerequisite for calculating accurate rates is the correspondence between episodes of exposure in residence and exclusion of exposure outside of the defined DSA.

The structure of the data collected in demographic surveillance sites is quite different from other types of demographic data required for measuring child mortality rates. The data for each individual child is organised into “episodes”. Births and deaths, a change in residence or migration, entrance into school or a change in household socioeconomic circumstance are events that result in closure of the current episode and opening of a new one. Transition into a new episode can also result from surviving into the subsequent age interval. This allows scale of measurement in person-time units and
events allocated to a different set of factors as circumstances for a particular individual change (Emmelin 2009). Hence, accumulation of the exposure presents a fundamental difference from the way other data sources treat the denominator in mortality analysis. The standard practice is to use live births or the population as the denominator, whereas episode data converts the episodes into person-years in order to estimate mortality more accurately. Representation of this mortality index can be as a central death rate in the life table and conversion to age-specific probabilities of dying.

The full birth history produces age-specific synthetic cohort death rates, $\pi n_m x$ values. The Chapter 6 on demographic surveillance utilises episodes of exposure as the unit of measurement, where the mortality is that of an observed birth cohort, and therefore differentiated from the rates in a period denoted with $\pi n M x$.

3.10 Regression analysis of count data
Regression analysis of count data is used in two instances in the thesis. First, to model correlates of the factors associated with child mortality using full birth history data and the negative binomial regression and second, to model the age-sex pattern of child mortality over time using Poisson regression.

3.10.1 Modelling correlates using the ACDIS full birth history
An event history file was constructed for children born in the 20-year period prior to the date of the mother’s interview. Each child’s exposure time, from birth until interview or the point of death, was divided into age groups and then summed across all children sharing a common set of characteristics up until their fifth birthday (Yamaguchi 1991).

An aggregate data file was structured and these data fitted with the negative binomial distribution to allow greater variance than the Poisson distribution resulting in a better goodness of fit statistic (STATA 2005). The following regression model was fitted to the log of the rate calculated as the number of deaths divided by exposure time measured in person-years:

$$\ln \left( \frac{d_{ij}}{PYE_{ij}} \right) = \ln(d_{ij}) - \ln(PYE_{ij}) = \beta_0 + \beta_1X_i + \beta_2X_{1j} + \beta_3X_{2j} + \cdots + \beta_{n+1}X_{nj}$$

where $d_{ij}$ is the count of deaths and $PYE_{ij}$ is the person-years of exposure for a particular age group $i$ and combination of covariates $j$. Exposure times are incorporated in the model as an “offset” term.
Regression models were used to explore mortality by selected socioeconomic covariates at the time of the survey.

3.10.2 **Poisson regression to investigate the age and sex pattern over time**

Poisson regression using STATA 10 (STATA 2007), generalized linear regression models (GLM) was used to investigate the relationship between age, sex and year (time period) and under-five mortality.

3.11 **Assessment of CRVS as a tool for measuring child mortality**

The methods of analysis proposed to evaluate the selected attributes of the CRVS data of children under-five years of age utilised for this thesis follows.

3.11.1 **Completeness of births and deaths**

As pointed out in Chapter 2 assessment of the completeness of birth registration began with a review of the literature on fertility estimates to facilitate comparison with recent, empirical estimates of the number of births derived from 2011 Census and annual live birth data, which forms part of the routine health surveillance system.

3.11.2 **Estimation of the number of births from fertility rates**

3.11.2.1 **Application of age-specific fertility rates**

The fertility estimates required to produce the numbers of births against which to evaluate the completeness of birth registration were calculated from three censuses and two surveys shown in Table 3-3.
Table 3-3: Data sources and references used to calculate the number of births based on age-specific fertility rates and the number of women by age

<table>
<thead>
<tr>
<th>Data source</th>
<th>Derivation of the age-specific fertility rate</th>
<th>Implied number of births</th>
<th>Distribution of women</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011 Census</td>
<td>(Dorrington and Moultrie 2015)</td>
<td>1,124,000</td>
<td><a href="http://interactive2.statssa.gov.za/webapi/jsf/tableView/tableView.xhtml">http://interactive2.statssa.gov.za/webapi/jsf/tableView/tableView.xhtml</a></td>
</tr>
</tbody>
</table>

3.11.2.2 Estimation of the number of births by reverse survival

Reverse survival is a method for estimating fertility and therefore the number of births as required here from census data” (Moultrie, Dorrington, Hill et al. 2013: 82; United Nations 1983). The method follows age cohorts from the reported population backwards to their year of birth and applies an appropriate schedule of age, sex and period-specific probabilities of surviving from birth to age \( x \).

Using life table measures \( \frac{L_x}{l_0} \), where \( L_x \) is the number of person-years lived between age \( x \) and \( x + 1 \) in the cohort life table of which \( l_0 \) is the radix, the numbers aged 0 to 15 counted in the 2011 Census were projected backwards to estimate the number of births for each age cohort. The estimation of the number of births involves the following steps:

1. Cohort survival factors to 2011 applicable to births for the past 15 years were derived from, the ASSA 2008 population projection with the migration assumptions set to zero.
2. These survival factors were derived by dividing the number in the population aged \( x \) in 2011 by the number of births it originated from. For example, the population aged 0 at the middle of 2011 divided by the births in the year from the middle of 2010, followed by the population aged 1 divided by the births in the year starting in the middle of 2009, etc.

\[ ^{12} \text{All Census data used in this thesis have been adjusted with the Post Enumeration Survey} \]
3. Each factor is then used to estimate the number of births (B) by dividing the number counted in the census by the appropriate survival factor. Thus, for example, the number of births in the year starting on 10 October 2010, $B_{2010}$, is derived by dividing the number under age 0 in the 2011 Census ($P_0$) by the survival factor for this from the model $\left( \frac{L_0}{L_0} \right)$:

$$B_{2010} = \frac{P_0}{L_0}.$$ 

4. These data are structured by census year from October to October, however, as calendar year births are required, the census year births were interpolated to produce births by calendar year. Therefore, the calendar year series of births are calculated thus:

$$B_{2010} = 0.7726 \times B_{2011} + 0.2274 \times B_{2010}$$

5. Finally, it is necessary to scale the number of provincial births to ensure that the sum of the births by province is equal to the national total for each year.

3.11.2.3 Estimation of the proportion of births occurring outside public facilities

The objective of the next method is to develop an appropriate adjustment for the births recorded in the District Health Information System (DHIS) to account for the proportion of births absent from that system, namely births occurring in private health facilities and those occurring at home. An important consideration in the development of the method is that it should also be applicable to the provinces.

Exploration of three national household surveys to provide provincial proportions of private facility-births and home-births revealed an association between the proportion of the population designated rural in the provinces and the proportion of home-births in each province. The correlation between the proportion of the population recorded as living in rural areas in the 1996 Census and the provincial proportion of home births from the 1998 SADHS result in an $R^2 = 0.74$ (Figure 3-2) proportions over time from the changes in the proportions rural indicating a strong enough correlation, to predict reasonably the relationship of these.

As the only other reliable measure of home births is that for the country as a whole for approximately 2011 (Shisana, Rehle, Simbayi et al. 2014) the coefficients

---

15 On the assumption that survival from birth in the year starting on 10th October is the same as that for births in the year starting on 1st July of the same calendar year.
were estimated from the following model:

\[ h_{bi} = HB \left( a \frac{r_i}{R} + b \right) \]

where \( h_{bi} \) = % home-births in province \( i \), and \( HB = % \) home-births nationally, and \( r_i = % \) rural in province \( i \), and \( R = % \) rural nationally.

Figure 3-2: Correlation of proportions rural recorded in the 1996 Census and proportion of home-births recorded in the 1998 DHS

Using these coefficients it is possible to estimates the \( h_{bi} \) for 2011 using the \( \frac{r_i}{R} \) from the 2011 Census and \( HB \) from the South African National HIV Prevalence, Incidence and Behaviour Survey, 2012 (Shisana, Rehle, Simbayi et al. 2014).

Finally, the provincial fertility, the number of births produced from reverse survival, the adjusted DHIS estimates and the cumulative births registered in the 26 months ending 12 months after the end of February in the year after the calendar year of birth referred to as ‘VR+1’ in the thesis are assessed against the known patterns of migration by province (Stats SA 2012b).

3.11.3 Calculation of the number of deaths

Beginning with an overview of the level and trend, Table 3-4 presents national estimates of infant and under-five mortality between 1996 and 2011. As with the births, this assessment of deaths utilises previously produced estimates in addition to self-derived estimates.
Table 3-4: Data sources and researchers of mortality estimates used to assess completeness of death registration by age

<table>
<thead>
<tr>
<th>Data source</th>
<th>Derivation of child mortality</th>
<th>Ref date</th>
<th>Estimate of $\overset{\circ}{q}_0$</th>
<th>Estimate of $\overset{\circ}{q}_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996 Census</td>
<td>(Nannan, Timaeus, Laubscher et al. 2007)</td>
<td>1995.49</td>
<td>51</td>
<td>69</td>
</tr>
<tr>
<td>1998 SADHS</td>
<td>(Nannan, Timaeus, Laubscher et al. 2007)</td>
<td>1996.75</td>
<td>50</td>
<td>67</td>
</tr>
<tr>
<td>2001 Census</td>
<td>(Dorrington, Moultrie and Timaeus 2004)</td>
<td>2001.27</td>
<td>48</td>
<td>72</td>
</tr>
<tr>
<td>2007 Community Survey</td>
<td>(Darikwa 2009)</td>
<td>2006.6</td>
<td>49</td>
<td>75</td>
</tr>
<tr>
<td>NBD Study 1997-2010</td>
<td>(Pillay-Van Wyk, Laubscher, Msemburi et al. 2014)</td>
<td>2000.5</td>
<td>44</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2005.5</td>
<td>51</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2010.5</td>
<td>34</td>
<td>52</td>
</tr>
<tr>
<td>2011 Census (Own calculations)</td>
<td></td>
<td>2011.27</td>
<td>34</td>
<td>46</td>
</tr>
</tbody>
</table>

3.11.3.1 Estimation of the trend in completeness of registered deaths by age

Underpinning the derivation of the expected number of deaths annually, is the utilisation of the estimated births the production of which was described in the previous sub-section as the initial population at risk, or from using the mortality derived from a life-table approach for calculating the completeness of reported deaths. Derivation of the mortality between the 2007 Community Survey and the 2011 Census was achieved by interpolating linearly between estimates at the survey dates, and recasting the corresponding calendar year reference date for the reported vital registration deaths. This involved the following steps for deaths under the age of one year nationally and provincially:

1a. The expected number of deaths under-1 in 2006 = births$_{2006} \times \overset{\circ}{q}_0_{2006}$.

The number of births in the calendar year 2011 were extrapolated from those in the Census year 2010 as follows:

$B_{2011} = 1.227 \times B_{2011} - 0.227 \times B_{2010}$.

1b. Estimated Completeness (C) as follows:

$C_{2006} = (0.9 \times VR <1 \text{ deaths}_{2005} + 0.1 \times VR <1 \text{ deaths}_{2006})/ \text{expected } <1 \text{ deaths}_{2006}$.

2a. The expected deaths under-1 in 2011 were derived from reported deaths in the previous 12 months before the Census (at 2011.773). The calculation of childhood mortality from the reported deaths and corresponding enumerated census populations were derived by complete life tables (see section 3.8.).
2b. The completeness in 2011 was calculated as follows:
\[ C_{2011} = \frac{0.773 \times VR\ deaths_{2011} + 0.227 \times VR\ deaths_{2010}}{expected\ deaths_{2011}}. \]

3. Linear interpolation was used to calculate the annual change in completeness between 2006 and 2011. This was applied to the VR deaths in the intervening years to produce a series of adjusted <1 and <5 deaths by calendar year.

4. The corrected \( q_0 \) and \( q_5 \) values were obtained as ratios using the estimated births derived by reverse survival and the value of \( q_1 \) derived from these.

6. Lastly, the provincial adjusted deaths were rescaled to produce the sum of the adjusted total deaths nationally.

3.11.4 **Assessment of epidemiological and age at death coherence**

The assessment of epidemiological coherence by Joubert *et al* (2013), employed models designed to predict changes in cause-specific mortality over time in accordance with the theory of epidemiological transition (Salomon and Murray 2002). Specifically, the cause of death model predicts the behaviour of the broad cause of death groups: communicable, non-communicable and injuries and single causes of death by age and sex. The authors conclude that assessment of epidemiological consistency was not conclusive due to the models’ “lack of discriminatory power” when applied to the South African data.

The framework proposes four features of the measure of quality, which encompass nine assessment criteria. The feature of ‘reliability’ to gauge the consistency of mortality data with regard to established epidemiological expectation. Reliability is assessed using the criteria of epidemiological and temporal consistency. Hence, the thesis seeks to develop an alternative analysis to assess epidemiological and age coherence in children under five years old by using model life tables.

3.12 **Demographic analysis of the age-pattern of deaths in children under-5 years of age**

Systems of model life tables are demographic tools used in circumstances in which the available information about age-specific death rates is lacking in some way. The thesis draws on the recommendation that the Coale and Demeny model life tables (MLT’s) also known as the Princeton MLT, best capture the variation of infant-to-child age
distributions in Sub-Saharan Africa (Jasseh 2003: 68). The Princeton MLT’s were ultimately based on of 192 life tables (out of 376 originally identified) for each sex, spanning 1900 up until 1945. The majority of tables; 92% are representative of the experience in western countries, the remaining 8% of the life tables were from Israel, Japan, Taiwan province and four from the white population of South Africa.

The age patterns of the four families of life-table have the following relationship, the North family has the highest child mortality values relative to infant mortality for \( q_0 \) values of less than 0.15; followed by the South family, then the East family. The West family represents an average of the families and lies in between the other representations.

Based on the 1998 DHS Jasseh (2003: 142) concludes that “the 1998 SADHS dataset outlines a childhood mortality pattern that has transformed from one very close to the Princeton ‘West’ to ‘North’ and back to ‘West’ as overall childhood mortality started increasing in the ten year period prior to the survey”. Given the reversal in mortality revealed from the DHS data, these findings signify the detection of a changing age-pattern associated with a change in the level of mortality using this methodology. The establishment of what has occurred since that time would be informative.
This chapter explores the cross-sectional survey instrument for the collection of full birth history data to estimate national levels of child mortality. It utilises data from two South African Demographic and Health Surveys (SADHS’s) to assess its overall performance, strengths and limitations. Section 4.1 describes the data sources and section 4.2 assesses data quality, discusses potential sources of error and bias typical of retrospective data and assesses national age-specific mortality trends. Section 4.3 assesses the impact of HIV and subsection 4.4 discusses a birth history versus a pregnancy history. The chapter ends with a discussion using an adapted vital registration framework to assess the performance of the full birth history, and draws conclusions and recommendations about the instrument.

4.1 Data sources

South Africa Demographic and Health Survey 1998 and 2003

The absence of basic population health information at the time of political transition to South Africa’s democracy necessitated a nationally representative survey to collect baseline data as part of the National Health Information System of South Africa. A range of demographic and health indicators were collected to assist policy makers in designing and evaluating programmes and strategies for improving health services in the country (Department of Health 2002). The design and implementation of the first survey in 1998 closely followed the Macro International blueprint tried and tested in numerous countries.

DHS surveys consist of five modules, the birth or pregnancy history is located within the reproductive health module. Its design is three tiers of data collection and some standard procedures of quality control. First, the questionnaires are highly standardised and interviewers trained to probe where no information is forthcoming from the respondent. Apart from checks at the end of the birth history to ensure all dates of birth, dates of death, current ages of children, etc., are accurately captured, there are additional questions (in some surveys) for instance to probe where a death is reported at twelve months, as such a death could be misreported and contributes to age heaping.

In terms of field procedures, training of fieldworkers and supervisors is paramount for good quality data. Training usually lasts for three to four weeks where
Particular emphasis is placed on the birth history module because it is the source of information for the direct calculation of fertility and mortality rates and identifies those children who are under-five years of age on the survey date about whom require additional health questions are asked later in the interview.

Field editing is the second level of quality assurance and is focused on the accuracy of birth history data. Field editors are responsible for validation of questionnaires for completeness and consistency. This is facilitated by editing questionnaires whilst the interview team is at a sampling point allowing re-visititation of households if necessary. In addition, field-check tables to monitor data quality during data entry are run on edited data so any problems detected can be communicated back to the team supervisors while fieldworkers are still in the field.

The third mechanism involves the data processing, which transforms questionnaires into data files using rigid rules for data entry, editing and imputation. At this stage, checks are carried out for inconsistencies and decisions made about their correction. For example, inconsistencies with the date of death can be validated with other information such as, with the date of interview, duration of breastfeeding, date of first breastfeeding, age at supplementation and with the date of last vaccination.

The first SADHS in 1998 drew a probability sample of approximately 12,000 women aged 15-49 from the sampling frame used in the 1996 Census and the country stratified into nine provinces and further stratified into urban and rural areas. At the time of implementation, there was the intention to repeat the survey every five years to enable the department of health to monitor trends in health services (Department of Health 2007), however the second survey was only conducted between 2003-2004.

Births and deaths by calendar year from both surveys are shown in Table 4.1, which indicates noticeable differences between the surveys. Striking is the difference in the number of births in the most recent five years with the 2003 survey capturing only 50% of births of women aged 15-49 the earlier survey recorded. Unsurprisingly the equivalent number of deaths is also considerably less; the 2003 survey recorded 38% of the deaths reported in the 1998 survey.
Table 4-1: Comparison of births and deaths by calendar year for the 1998 and 2003 SADHS’s

<table>
<thead>
<tr>
<th>Year</th>
<th>1998 DHS (11735 women)</th>
<th>2003 DHS (7 041 women)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total births</td>
<td>Deaths</td>
</tr>
<tr>
<td>1998</td>
<td>276</td>
<td>5</td>
</tr>
<tr>
<td>1997</td>
<td>1 075</td>
<td>61</td>
</tr>
<tr>
<td>1996</td>
<td>1 065</td>
<td>53</td>
</tr>
<tr>
<td>1995</td>
<td>935</td>
<td>57</td>
</tr>
<tr>
<td>1994</td>
<td>974</td>
<td>46</td>
</tr>
<tr>
<td>1993</td>
<td>957</td>
<td>50</td>
</tr>
<tr>
<td>1992</td>
<td>1 192</td>
<td>80</td>
</tr>
<tr>
<td>1991</td>
<td>1 179</td>
<td>45</td>
</tr>
<tr>
<td>1990</td>
<td>1 058</td>
<td>48</td>
</tr>
<tr>
<td>1989</td>
<td>1 017</td>
<td>63</td>
</tr>
<tr>
<td>94-98</td>
<td>4 325</td>
<td>221</td>
</tr>
<tr>
<td>89-93</td>
<td>5 403</td>
<td>285</td>
</tr>
<tr>
<td>84-88</td>
<td>4 804</td>
<td>355</td>
</tr>
<tr>
<td>79-83</td>
<td>3 999</td>
<td>358</td>
</tr>
<tr>
<td>&lt;1979</td>
<td>4 189</td>
<td>589</td>
</tr>
<tr>
<td>All</td>
<td>22 756</td>
<td>1 808</td>
</tr>
</tbody>
</table>

4.2 Data quality

An evaluation of the quality of the demographic data is an essential step in the research process and therefore a starting point for analysis, as deficiencies could lead to biases and misinterpretation of the results. Common errors encountered with the full birth history data are the omissions of births and deaths, the misreporting of dates of birth and ages at death and the misreporting of the mothers’ birth dates. The most serious error is the omission of vital events (Curtis 1995; Rutstein 1984; Sullivan, Bicego and Rutstein 1990), which is thought to be a particular problem with children who have died, particularly if the death occurred many years before the survey, and for children living away from their mothers. Deaths in the neonatal period, particularly early neonates, are also more under-reported than deaths occurring at older infant ages. Sometimes still-births are included as livebirths and often there is abnormally high proportion of births to very young women more than 15 years prior to the survey. Both would lead to an overestimation of the level and trend in childhood mortality.

Misreporting of birth dates has the potential to affect mortality trends, especially if there are differences according to whether the child was alive or not at the time of the survey. The two types of birth date misreporting are digit preference (age heaping caused by the respondent approximating age) and the more serious problem of birth transference (the mis-recording of the date of a birth by the fieldworker placing it
further back in time to avoid extra health-related questions in a survey). Both the SADHS and ACDIS data are assessed for these deficiencies below.

4.2.1 **Birth transference**
Birth transference is a potential bias and a consequence of the lengthy health related questions DHS’s require for births that have occurred in the five years preceding the survey. The additional questions increase the workload of the field workers considerably for each woman who has experienced a birth after the specified cut-off date. Some fieldworkers shift some of these births to a period before the cut-off, thereby avoiding the time-consuming questionnaire, and resulting in the transference of births to an earlier period.

Plots of the distribution of births by calendar year can reveal evidence of this type of bias and by constructing a simple index of birth transference: the ratio of the number of births in the year prior to the cut-off date to the number in the year after the cut-off. Birth transference has become more common with more recently conducted surveys, possibly due to increases in the length of the health questionnaire.

**Figure 4-1: Distribution of births seven years before the survey, SADHS 1998 and 2003**

The distribution of births by calendar year for the 1998 and 2003 DHSs are shown in Figure 4-1. The 1998 survey shows transference of births from 1994 into 1993 indicated by an index of 1.24, whereas the 2003 survey shows a proportionally more substantial birth transference from 1997 into 1998 with an index of 1.45. Although there is transference in the earlier survey, the displacement occurs within the five year cut-off, and therefore the 1998 survey does not suffer the severely distorted fertility levels in the preceding five years as does the 2003 survey (Department of Health 2007).

Sullivan (2008) explains that the impact of birth transference on mortality estimates actually depends on two factors. First, the extent of the transference – the
relative magnitude of transfers of births of dead or surviving children and, second, in the case of dead children, whether or not both the birth and the death of the child were transferred in or out of the defined period of estimation. The second situation is why the effect of birth transference impacts more on the neonatal and infant mortality rates than on the mortality of children under-five years. Birth transference in early infancy shifts both the birth and the death out of the most recent period of estimation, thus it impacts on the mortality estimate more at younger ages. In contrast, if children die at ages older than one year the death remains in the 1-4 age group minimizing the impact on the under-5 mortality rate.

4.2.2 Underreporting of deceased children

The under-reporting of deceased children also has the potential to effect retrospective survey data. This could occur due to the respondents’ reluctance to report basic details of children that have died particularly children who have died recently. On the other hand, it could also occur due to the interviewer’s intentional failure to record births of recently deceased children to avoid asking the lengthy follow-up questions.

An assessment of the extent of underreporting of births and deaths in a survey would be straightforward if reliable estimates were available from vital records; however, it is precisely this information that is lacking for countries requiring evaluation of child mortality rates through surveys. A way to assess under-reporting of events reported in surveys is to compare the under-five mortality estimates from a pair of surveys for a common reference period (Sullivan 2008). Although, a limitation of this approach is that, it offers a relative evaluation, and the survey yielding higher mortality is considered more accurate, which, may not necessarily be the case.

In the case of the two DHS surveys, the two trends show large differences in the overall level of reported deaths as illustrated by Figure 4-2. The 1998 survey shows an overall increase from 50 to 61 deaths over the five years prior to the survey. The 2003 survey also shows an increase over the five years prior to the survey from 13 to 22, one-third the level of the earlier survey, strongly indicating under-reporting of deaths throughout the five years in the 2003 survey.
4.2.3 Misreporting of age at death
The misreporting of age at death or age heaping is similar to birth transference in that it also has particular consequences for mortality measures in early infancy as mortality risk declines rapidly during the neonatal and post-neonatal age periods. Preference for reporting the age at death as six or twelve months is common (Curtis 1995) and like birth transference, has particular relevance to neonatal and infant mortality. In addition, because of the rapidly changing mortality risk within the first year of life, it is important to report the age at death in months and not just in completed years, as is standard practice for the ages older than one year. Study of the data quality tables in the SADHS reports do not indicate age misreporting errors (Department of Health 2002: p326, 2007: p398).

4.2.4 Recall bias
Recall bias is a potential problem with all retrospective information because recollection of event-dates or episodes is required. It highlights the well-known problem of relying on human memory to recall exact details of the events accurately. The failure to recall accurately may affect the estimates in several ways dependant on whether the event is completely omitted, which would have a different effect from an event being misplaced in time. Events further in the past suffer most errors especially those reported by older women in the sample with higher parities, referring to their older children. Garenne and Gakusi (2005) argue that biases due to recall can be both positive and negative, which they assume generally cancel each other out. The effect of recall bias on child morbidity data is common but there is very little discussion in the literature regarding its impact on the quality of child mortality data. The extent of the bias depends on how far back in time the births occurred. The bias in the recent period before the survey is minimal as...
older women have few births in this period. However, the births 15 to 20 years before the survey carry greater biases as they belong to women who were younger at the time gave birth to a higher proportion of the overall births, hence leading to over-estimation of mortality in the earlier periods before the survey. This truncation of older women’s birth history information is a well-established definitional bias and common to these data.

4.2.5 **Assessment of South Africa Demographic and Health Surveys 1998 and 2003**

The 1998 and 2003 South Africa Demographic and Health Surveys collected full birth history information, which are assessed next. The survey adopts a probability proportional to size methodology and stratifies the country into nine provinces and urban and rural regions. The most important differences between the two surveys were the sample of selected households and the response rates of eligible women in the households by age, resulting in a drop from 11,735 administered reproductive questionnaires in the 1998 survey to 7,041 in the 2003 survey.

DHS’s calculate synthetic cohort probabilities of childhood deaths using the procedure first developed by Somoza (1980) and later modified by Rutstein (1985) and reported in sub-section 3.4.1 of the Methods chapter. These data were extracted based on this methodology and programmed in STATA (see Appendix 3). The age-specific mortality trends from the 1998 survey disaggregated by single year are shown in Figure 4.3.

Regardless of the fluctuations caused by sample size from year to year, the trend clearly shows a reversal of the downward trend from the early 1990’s. The mortality represented by the most recent point measures an incomplete year of exposure; hence, some observations are censored leading to an overstatement of the probability of death. For this reason, it is more accurate to report the most recent mortality as an average of the two years 1997.75 and 1998.75. This results in an estimate of $\text{1q}_0$ of 50.6 and of $\text{5q}_0$ of 67.0 per 1,000 live births.
Figure 4-3: Age-specific period mortality, single year estimates, SADHS 1998

The 1996 Census provided a useful benchmark against which the estimates of this survey can be compared. The provincial estimates from the census and the DHS were reasonably similar excepting the Western Cape in particular, Free State and the North West province due to underreporting of child deaths in the DHS (Nannan, Timaeus, Laubscher et al. 2007). Regression analysis using the relationship between the SADHS and the census suggested that the national childhood mortality found in the DHS needed to be increased by 7.5 percent. This level of under-estimation is consistent with the predicted six percent attributable to omission of deaths under the age of five generally in national surveys (Sullivan, Bicego and Rutstein 1990). Aside from this minor adjustment, good data quality is confirmed by corroboration of the age-specific trend with estimates derived from the 1992 HSRC also using a FBH almost a decade earlier (Department of Health 2002: 101).

In contrast, the 2003 survey data suffer a several odd features indicating problematic data quality. The age-specific rates averaged over two years are shown in Figure 4-4. The estimates indicate a wider margin of statistical error due to a smaller sample than the 1998 survey, as reflected by the SE around the under-5 mortality rate of 3.65 in the 1998 vs. 7.5 in the 2003 survey. Actually, interpretation of the rates is not straightforward due to erratic fluctuation even after grouping over two years. The average rate for 2002 and 2003 yields an estimate of $1q_0$ of 62.9 and of $5q_0$ of 81.9 per 1,000 live births but is severely biased by censored cases in 2003. Figure 4-4 indicates
that deaths under age one were under-reported in 2000, making interpretation of the recent trend in child mortality challenging. Under these circumstances, it may be more sensible to report rates per five-yearly periods. This results in an estimate of $q_0$ of 43.1 and of $q_0$ of 57.0 for 2001, the mid-point of the (1999-2003) interval. Further, assessment of the provinces found the KZN reproductive data particularly wanting as reflected by, for example, implausibly low birth and death rates for that province (Department of Health 2007).

![Graph showing age-specific period mortality averaged over two years, SADHS 2003](image)

**Figure 4-4: Age-specific period mortality averaged over two years, SADHS 2003**

Thus the 1998 survey collected better quality data, as indicated by lower birth transference and standard errors, and illustrated by a clear age-specific trend generated year to year. The extent of the differences between surveys is surprising due to high level of standardisation regarding methodology, fieldwork procedures and data processing between the two surveys and point to poor implementation of the fieldwork of the 2003 survey.

### 4.3 Impact of HIV/AIDS

As well as the two SADHSs, the thesis also includes a full birth history, representative of the Africa Centre Demographic Information System (ACDIS), which is the subject of Chapter 6. The impact of HIV on all three of these retrospective data sets will be included in this section.
Prior to the 1998 DHS HIV prevalence among women attending antenatal clinics was 22.8%, by 2003 HIV prevalence had increased to 27.9% and in KZN prevalence in 2003 was estimated to be 37.5% (95% CI 35.2-39.8) (Department of Health 2005). Throughout this period, no ART and very limited PMTCT were offered by the public sector, meaning that mothers experienced increasing mortality and the mortality risk of children born to HIV positive women increased. The assumption in these circumstances is that 60% of infected children were expected to die before their fifth birthday (Marston, Zaba, Salomon et al. 2005). As the FBH is based on mothers’ reports, and some HIV-positive mothers died and been absent from the survey, the reported child mortality suffers a downward bias necessitating adjustment to the data under review here.

The literature review details two methods for adjusting for this. First, the simple cohort component projection model used by the IGME (Walker, Hill and Zhao 2012) which is available as a customised excel spreadsheet. It requires the number of women by age and prevalence rates of pregnant women for the period in question. The second is a model proposed by Hallett and colleagues (Hallett, Gregson, Kurwa et al. 2010).

Application of this method to the 2003 DHS is shown in Figure 4-5. The adjustment demonstrates the differential bias throughout the period resulting from women’s time since exposure. There is practically no bias in the three years before the survey since most women who give birth in this period are still alive at the time of the survey. The bias peaks averaging 23% in the period 5-9 years before the survey. Although the magnitude of bias appears high, this is consistent with the analysis of six other countries undertaken by Walker and colleagues. Applying this adjustment corrects the reported under-five rate during 1994-1998 from 49.0 to 64.2 per 1,000 live births, much closer to the 1998 survey estimate of 59.4 for the same period.
Figure 4-5: Probability of dying before age five averaged over two years, and adjusted for mothers’ survivor bias using IGME approach, SADHS 2003

The model by Hallett and colleagues requires age-specific HIV incidence, age-specific non-HIV fertility, infant and under-5 mortality and a female birth cohort from 1920 as input data. In the absence of any interventions, both models assumed a 35% MTCT rate.

Application of this method to the ACDIS data produced an estimate of 2.9% bias. The observed and adjusted infant and under-5 mortality implied by the Hallett et al. model are shown in Figure 4-6. This shows that the bias grows as the epidemic matures and that the bias is greater for under-five mortality than for infant mortality as the mothers who gave birth further in the past are more likely to have died.

Figure 4-6: ACDIS infant and under-5 mortality trend adjusted for mother's survivor bias

The extent of bias introduced to the South African estimates warrants comparison of the two models. Table 4-2 compares the estimated bias from both models and an
estimate of the potential bias based on a review by Mahy (2003). The method based on incidence proposed by Hallett produces consistently lower estimates of bias than the method based on prevalence and fall outside of the range predicted for the 2003 national data by Mahy.

**Table 4-2: Comparison of the estimated bias produced by different methods for each source of birth history data, under-five mortality**

<table>
<thead>
<tr>
<th>Source of Data</th>
<th>SADHS 1998 (%)</th>
<th>ACDIS 2000 (%)</th>
<th>SADHS 2003 (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mahy (2003)</td>
<td>-</td>
<td>-</td>
<td>5.0 - 7.0</td>
</tr>
<tr>
<td>Hallett (2010)</td>
<td>0.6</td>
<td>2.5</td>
<td>4.2</td>
</tr>
<tr>
<td>Walker (2012)</td>
<td>1.6</td>
<td>4.7</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Similarly, the models generated the same pattern in the same direction when compared with six other African countries (Walker, Hill and Zhao 2012) and begs the question as to why the method based on incidence produces consistently lower estimates of bias than the method based on prevalence. The only reason cited by (Walker, Hill and Zhao), who noted this discrepancy in their model compared to Hallet’s model demonstrated with Namibian data, is that the overstatement relates to the assumption that all births to HIV-positive women occur four years after infection with corresponding shortened survival times for mothers. Hence, they caution that the method does not adequately account for the inter-relationships early in the epidemic when a large proportion of HIV-positive mothers are at an early infection stage and this probably leads to overestimating the bias in the early years of an epidemic.

Apart from the fact that all three of the methods reviewed in this chapter include simplifying assumptions that may affect the accuracy of the predictions (Walker, Hill and Zhao 2012), further research taking into account the effects of ART on the epidemic in South Africa is required for an understanding of the size of the adjustment in future.

### 4.4 Pregnancy history vs. birth history

A difference between the two SADHS’s already mentioned is that the earlier survey recorded live births using a pregnancy questionnaire whereas the second survey administered a birth history. The pregnancy history asks about the outcome of each pregnancy, systematically recording details beginning with the earliest pregnancy. In this
case, the questionnaire contained probing questions\textsuperscript{14} to determine whether a pregnancy ended in a live birth rather than a stillbirth by asking the mother about any signs of life the new-born may have shown. This proved to be effective, as eighteen percent of women (71) who initially answered the baby was born dead, changed their minds after considering whether the baby showed any signs of life, and recorded them as live births who died shortly after birth. Although the probe serves to distinguish stillbirths from deaths in the first few hours, it is odd these data do not display the expected distribution where most deaths occur in the first day of life, decreasing each day after that with heaping on the seventh day (Frøen, Cacciatore, McClure \textit{et al.} 2011). Instead, there is a peak on the second day (plots not shown).

A recent study from Mali also suggests about twenty percent of 365 neonatal deaths identified in a FBH survey were classified as stillbirths in a verbal/social autopsy survey (Liu, Kalter, Chu \textit{et al.} 2016). Similarly, the 2013 Nigeria DHS found twenty two percent of neonatal deaths recorded in the FBH were classified as stillbirths in a follow-up verbal/social autopsy survey (Adewemimo, Kalter, Perin \textit{et al.} 2017), both proving strong evidence for the case of a pregnancy history.

Nevertheless, the need to eliminate this particular misclassification is important for two reasons. First, to ensure that information related to the IMR and the components of infant mortality are accurately recorded, because, as it is strongly believed, as the overall rate decreases the contribution of neonatal deaths to the IMR increase as a proportion; hence, to distinguish early neonatal deaths from stillbirths becomes critical. Second, these data, to some extent allow measurement and monitoring of a stillbirth rate which has the potential to serve as an important indicator of maternal health, particularly nutritional status of the mother (Hart 1998). In this regard, the potential utilisation of stillbirths as a health indicator has important implications for South Africa for two reasons. Firstly, in light of South Africa’s health targets related to maternal and child health these data appear promising for the research area. Secondly, these data are essential to investigate the relationship between maternal HIV status and stillbirth which requires a better understanding in the context of the impact of the epidemic on maternal health (Department of Health 2013). The 1998 DHS estimated a stillbirth rate of 14.7 per 1,000 births 1995-1998 which is similar to estimates for Brazil,

---

\textsuperscript{14} 208) Women sometimes have pregnancies that do not result in a live born child. That is a pregnancy can end very early, in a miscarriage or an abortion or the child can be born dead. Have you had any such pregnancy that did not result in a live birth? 216) Was the baby born alive, born dead, or lost before full term? 217) Did that baby cry, move, or breathe when it was born?
Jamaica and Mauritius as estimated by the WHO in 1995 (Cousens, Blencowe, Stanton et al. 2011). However, the WHO estimate of the South Africa stillbirth rate is 23 and 20 per 1,000 births in 1995 and 2009 respectively, much higher than the rate derived by the national survey, and this in itself, provides good reason to collect empirical data to monitor such a health indicator as opposed to reliance on modelled estimates of questionable reliability. The 2015 WHO estimate of South Africa’s stillbirth rate indicates a decrease to 17.4\textsuperscript{15} per 1,000 births.

4.5 Discussion

This chapter set out to assess the FBH as a source of data for measuring childhood mortality in the South African setting. The chapter began with an acknowledgement of inherent definitional biases in the FBH data because only the birth histories of women alive at the date of the survey are captured, and then illustrated the effects of other biases. Although FBH data has been used frequently in South Africa, this chapter and the literature review make the point that numerous surveys have failed in their primary objective of providing reliable estimates of mortality level. It further highlighted severe data quality problems with birth transference, the under-reporting of deceased children, the misreporting of age at death, recall bias and mother survivor bias.

Against the Health Metrics Network (HMN) criteria shown in Table 4-3, it is clear that central to all the elements of accuracy is the issue of fieldwork. Hence, if emphasis is placed on achieving a high standard of fieldwork the elements of reliability, completeness and comparability will all score highly. It also brings to the fore the design limitations of the FBH such as its inadequacy for estimating levels of childhood mortality in populations smaller than provincial populations.

The data presented shows the FBH can provide accurate age-specific rates and confidence intervals as well as information related to a wide range of factors associated with mortality during childhood. However, the chapter has shown that poor data quality can compromise the measurement of the mortality of the finer age groups within infancy, despite the fact that the FBH is the best instrument to measure these age-specific rates.

Although differences between the standard errors of the estimates the two surveys exist, stark differences in data quality are also apparent and these are probably due to

\footnote{http://apps.who.int/gho/data/view.main.GSWCAH06v accessed April 2017}
procedural differences in the field. While several field procedures exist in order to check the data, it is difficult to guarantee the extent to which they are put into practice (Curtis 1995). It therefore becomes vital to institute quality assurance measures on the three levels of training, data collection and data processing, as outlined in the Macro fieldworker manuals as well as guaranteeing fieldworkers receive strong managerial support from well-trained supervisors during data collection.
Table 4-3: Adapted assessment framework for civil registration and vital statistics systems Mahapatra, Shibuya, Lopez, Coullare et al. (2007)

<table>
<thead>
<tr>
<th><strong>Full Birth History</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy</strong></td>
</tr>
<tr>
<td><strong>Coverage</strong></td>
</tr>
<tr>
<td><strong>Reliability: concerned with the accuracy of the actual measuring instrument</strong></td>
</tr>
<tr>
<td><strong>Completeness: % of events recorded</strong></td>
</tr>
<tr>
<td><strong>Comparability: the extent to which a statistic estimates the same thing in the same way repeatedly</strong></td>
</tr>
<tr>
<td><strong>Missing data</strong></td>
</tr>
<tr>
<td><strong>Policy Relevance</strong></td>
</tr>
<tr>
<td><strong>Cross-tabulation: the degree to which cross-tabulation for priority characteristics are provided by the data source i.e. a demographic and public health perspective.</strong></td>
</tr>
<tr>
<td><strong>Timeliness: the production time from end of survey reference period to time of publication</strong></td>
</tr>
<tr>
<td><strong>Accessibility</strong></td>
</tr>
<tr>
<td><strong>Media: dissemination of information, print, electronic, internet etc.</strong></td>
</tr>
<tr>
<td><strong>Metadata: information about the data, documentation of the data elements, their definition, method of collection, manner of presentation, potential for errors, etc.</strong></td>
</tr>
<tr>
<td><strong>Public use micro-dataset.</strong></td>
</tr>
</tbody>
</table>

Age-specific trends are highly informative and although the FBH can provide the detail of information, high standards of fieldwork are required and this does carry cost implications. To collect and generate this level of information, more thorough and longer training is required, which is far more expensive than the cost of collecting summary birth history information. South Africa experienced a long and detrimental void of FBH estimates since the 1998 DHS, and although the 2007 Community Survey
suggested that under-5 mortality remained quite stable since 1998, South Africa still lacks knowledge of the trends in the neonatal and post-neonatal components of infant mortality by province in 2015. Aside from that, HIV has mostly affected the post-neonatal rates and a finer-age perspective will help us to understand the dynamics of the epidemic on child mortality in an environment of treatment and effective PMTCT.

A primary conclusion of the chapter is that the greatest challenge to the accuracy of the FBH is the underestimation produced by mother’s survivor bias. The chapter compared two available approaches for adjusting the data and showed the underestimation is largely consistent between the two methods with the exception that the cohort component approach overstates the bias in populations affected in the early stages of a country epidemic. To this end, IGME intend to further refine certain assumptions of their model and caution that future analysis will have to take into account the response of interventions and their impact on infection and mortality. The follow-up of populations offered by DSSs to include community/cluster level questions which might help identify missing mothers, and thereby quantify the survivorship bias, as was done by Hallett and colleagues (2010) and will remain important in tracking changes in this bias as the epidemic develops.

The chapter also addressed the issue of a pregnancy or live birth questionnaire and although it is broadly agreed that the pregnancy-related questions ascertain a more accurate early neonatal rate, this finding is inconclusive from these data, as they do not display the expected distribution of deaths in the first week of life. However, recent studies provide strong evidence that a birth history can misclassify around twenty percent of stillbirths as neonatal deaths. In a broader context, this issue plays a role in the improved measurement of the early neonatal rate and therefore the age-components of the IMR over time and the ability to monitor these trends. Not forgetting the function of the pregnancy-related information to derive a nationally representative stillbirth rate, which is undoubtedly of value simply because the WHO do not have reliable estimates for South Africa, and therefore for the sub-Saharan region, we may do better with an empirical estimate rather than an estimate based on model assumptions.
The primary objective of a Census is to count the entire population; thus, the strength of census data lies in the fact that it is not a sample. This feature provides significant power, enabling demographic analysis by sub-groups such as population group and geographically demarcated areas such as province and district. Census data therefore occupy a prominent position in demography due to the utility of these data for measuring the four components of population dynamics; the base population, fertility, mortality and migration in one data resource. Census data from several censuses also allow time-series analysis, cohort analysis and the production of growth rates all of which are central to monitoring development.

South Africa has conducted three censuses post-Apartheid, in 1996, 2001 and 2011. A number of questions from these enquiries can be utilised for the estimation of births and child mortality. The direct estimation of child mortality rates requires the enumeration of the population by age and sex, an estimate of the deaths and ideally an estimate of the births. Censuses also include questions typically asked in developing countries that enquire about mortality indirectly such as the children ever born and children surviving (CEB/CS), often called the summary birth history (SBH) method, or the question asked of woman about their last birth and the survival status of that birth, originally called the previous birth technique (PBT). Censuses also ask (from the 2010 round) households to report directly on deaths of members of the households.

The purpose of the rest of the chapter is to highlight the major challenges identified from the 1996 and 2001 censuses from published material. Second, to assess the performance of the data from the questions required for measuring childhood mortality from the 2011 Census.

5.1 Challenges identified from the 1996 and 2001 censuses

5.1.1 The 1996 Census
Data from the 1996 Census allowed robust estimates of infant and under-5 mortality by population group and by province to be produced. In conjunction with the 1998 DHS these estimates corroborated the reversal in the historical decline in childhood mortality in the early 1990’s and provided important information highlighting racial and provincial inequalities in childhood mortality (Nannan, Timaeus, Laubscher et al. 2007). The census data were also instrumental in informing the life tables and the demography under-
pinning the AIDS and demographic projection models (Dorrington, Bradshaw, Johnson et al. 2006).

An important observation affecting the quality of the parity data relates to the phrasing of the questions. Moultrie and Timaeus (2003) argue that the number of child deaths reported in the SBH data is inflated by the misclassification of stillbirths, reported as live births that succumbed to death. Unlike the DHS pregnancy histories, the SBH questions do not include probes to distinguish between stillbirths and live births. If one assumes that the propensity to report stillbirths as dead births was as strong in the census as in the DHS, the number of dead children reported should be adjusted downward by approximately 22\% to allow for misclassification of the outcome of these pregnancies (Nannan, Timaeus, Laubscher et al. 2007).

This is supported by the fact that (after correcting the census data for coding of childless women as parity not ‘stated’) the numbers of living children reported by the two sources match well. Moultrie and Timaeus (2006: 81) conclude that the reporting of lifetime fertility was not problematic in either source, however, the reporting of dead children was problematic. In addition, the census question was a complex one, with a final phrase in the second set of parentheses that perhaps offset the instruction in the first parenthesis, namely:

“How many children, if any, has the woman ever given birth to? (live-births)
(Please include her children who are not living with her and those who have died).”

5.1.2 The 2001 Census
Stats SA commissioned the Centre for Actuarial Research to evaluate the 2001 Census data on mortality and fertility, subsequently published in two monographs, one addressing the estimation of fertility (Moultrie and Dorrington 2004) and the other the estimation of mortality (Dorrington, Moultrie and Timaeus 2004). Thorough evaluation of the questions in the fertility schedule was undertaken. The wording of the CEB/CS questions changed from the 1996 questionnaire to include the ‘sex’ of the child and to include a prompt to remind the respondents to include all of the mother’s children, including children not living in the household. (The wording of which is shown in Figure 5-1).

Unlike the 1996 Census, the 2001 and 2011 censuses used editing and imputation to replace unknown and inconsistent responses. These data were found to be seriously flawed, “both in terms of the extent of imputation (particularly at the younger ages), and
in terms of the overall level of childbearing implied by the data (which are logically inconsistent with previous census data at the older ages) (Moultrie and Dorrington 2004: 41). The authors highlight the selected impact the editing processes of logical imputation and hotdecking had on the fertility data. For instance that, the fertility questions (Figure 5-1) which consist of five variables, required imputation in over fifty percent of responses.

<table>
<thead>
<tr>
<th>P-20: TOTAL BIRTHS:</th>
<th>How many children, if any, has (the person) ever had, that were born alive?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>If none write 00 and go to P-21 (next question).</td>
</tr>
<tr>
<td></td>
<td>How many of these were boys?</td>
</tr>
<tr>
<td></td>
<td>How many of these were girls?</td>
</tr>
<tr>
<td></td>
<td>Include ALL her children, i.e. those who are still living, whether or not they live in this household, and those who are dead. DO NOT COUNT STILLBIRTHS (children born dead).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P-20a: STILL LIVING:</th>
<th>If the person has ever given live birth:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>If boys: How many boys are still alive?</td>
</tr>
<tr>
<td></td>
<td>If girls: How many girls are still alive?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P-20b: LAST CHILD BORN.</th>
<th>If (the person) has ever given live birth:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>When was (the person's) last child born? (DD/MM/YYYY)</td>
</tr>
<tr>
<td></td>
<td>What is the sex of that child? (M/F)</td>
</tr>
<tr>
<td></td>
<td>Is that child alive or dead? (A/D)</td>
</tr>
<tr>
<td></td>
<td>Write the day, month and year of the last live birth and dot the appropriate box of the sex.</td>
</tr>
<tr>
<td></td>
<td>If multiple birth, indicate only the last child. Dot the appropriate box whether the child is still alive on Census night 9 - 10 October. DO NOT COUNT STILLBIRTHS (children born dead).</td>
</tr>
</tbody>
</table>

**Figure 5-1: Wording of the fertility questions in the 2001 Census**

Source: Moultrie and Dorrington 2004.

Further, the children surviving data were worse than the children ever born data and were logically inconsistent with the results from the 1996 Census despite extensive data editing. While it is possible to edit the birth data for anomalies and contradictions, it is not possible that consistent edits are made to the deaths. They conclude that, “on aggregate, the results presented here point to significant flaws in the data collected on women’s life-time fertility in the 2001 Census. Certainly, women tended to confuse the two questions indicating weakness in the enumerator training. However, this is clearly not the only error present in these data” (Dorrington, Moultrie and Timaeus 2004: 51).

Nonetheless, despite issues of questionable data quality, using the reported deaths in past 12 months, under-five mortality for year 2000 produce a \( q(1) = 45/1,000 \), and a \( q(5) = 69/1,000 \), consistent with most other estimates for that period.

5.2 **The enumerated population**

Essential to the estimation of mortality rates is the denominator, which for childhood mortality are the births and the numbers in the population between ages one and four. Aside from the role of the denominator in estimating mortality centred on the census date, these data are utilised to produce mid-year population estimates in the intercensal
period, making their accuracy important to many applications. It is therefore appropriate to review the performance of the census in counting the population.

The post-enumeration surveys estimated under-counts of the population of 10.7%, 17.6% and 14.6% in the 1996, 2001 and 2011 censuses respectively. It is recognised that censuses, particularly in less developed countries tend to under-count young children. Evaluation of the accuracy of census data, however, is not straightforward and the results are often debated by researchers. For instance, comparison of the 1996 Census with a projection from the estimated 1970 population identified missing children under the age of five, and a possible under-count of the population of between one and two million (Dorrington, Bradshaw and Wegner 1999).

The 2001 Census count was found to be largely consistent with the results of the previous census (Dorrington, Moultrie and Timaeus 2004; Moultrie and Dorrington 2004). Comparing the age distributions of the populations produced from three censuses reveals that the 1996 and 2001 had similar counts by age of the population under age nine. In contrast, the age distribution in 2011 indicates a big increase in the number of children under-five years of age. This observation caused some controversy and was highlighted in the national press, and raised questions about the validity of the census count of the population under ten years of age.

The age distributions suggest that between 1996 and 2001 the population under age one increased by 6.1% and between 2001 and 2011 the number of children under one increased by 27.6%. This was an unexpected result to demographers amongst whom there was a consensus that the number would be lower, based on previous evidence of a declining fertility even before the 1996 Census results (Moultrie and Dorrington 2004; Moultrie and Timæus 2003).

Closely related to the age distribution of the young population is the role census data play in the provision of denominator data required for monitoring childhood mortality during the intercensal period. In South Africa, Statistics South Africa produces mid-year estimates (MYEs) of the population for the current year (and all years back to the year after the previous census) “based on the latest available information” (Stats SA 2011: 2). Stats SA constrains the estimates so that those for 2011 aggregate to the same totals by sex, population group and province, taking into account the growth of the population from the middle of the year to the date of the census. The age distribution of the 2013 official set of MYE has, however, been questioned due to, in some instances, a quite marked deviation from the 2011 Census age distribution (Dorrington 2013).
Although there is no published report for the mid-year population estimates of 2013, both the reports of the 2011 (Stats SA 2011) and the 2015 mid-year estimates (Stats SA 2015b) state that the estimates are based on the latest available information, one can only assume they are calibrated to the 1996, 2001 and 2011 censuses. A comparison of Stats SA’s 2013 MYE and the age distributions from the 2011 Census is shown in Figure 5-2. The most notable differences between the two age distributions occur under the age of 40 years, in that the official MYE are considerably higher than the census below age 9; lower than the census count between ages 9 and 18, and higher than the census count between ages 18 and 40 years of age.

![Figure 5-2: Comparison of 2011 Census age distribution and the official 2013 mid-year estimates, Stats SA](image)

In light of the implications of using the official MYEs when applied to the population under the age of forty, and the lack of justification for deviating from the age distribution of the census count, a set of alternative MYEs have been produced with an age distribution much closer to that of the 2011 Census age distribution (Dorrington 2013).

5.3 **Births**

The method of deriving of births from the enumerated population is outlined in the methods chapter (section 3.11.2.2). The method provides estimates of the number of births over the past 15 years. Enumeration of the population in the province of
residence on census night the ‘census count’ and enumeration by province of birth produce two different series of birth estimates. Figure 5-3 shows the estimate of births from the reverse survival of the population count on the assumption of ASSA2008 survival factors is slightly higher than the estimate from the count of the South African-born population suggesting that the number of non-South African children is small.

To a large extent, reverse surviving the population under fifteen years of age, corroborates the questionable fertility (since census 2001) implied by the census count in that the trend confirms the increase in births necessary to produce the increase in the population under-fifteen.

Further confirmation is provided by the analysis conducted by Dorrington and Moultrie using multiple data sources (2015). In order to resolve the issue of the census count, they carried out a detailed investigation that sought to clarify whether the implied fertility was real or not, using multiple data sources and techniques. The approach included the following: births produced by reverse survival from the 2011, 2001 and 1996 censuses and the 2007 Community Survey; estimates of the TFR from the same sources; comparative trends in the number of births from routine health information systems data; the number of children of school-going age produced from school enrolment data; and as an estimate of births from births registered with civil registration and vital statistics. Although the authors state that further research is still required, they conclude that the age distribution produced by Census 2011 is consistent with all of the relevant fertility and population evidence.

Figure 5-3: Trend in the number of births derived from the enumerated population under fifteen years of age compared with various census and survey estimates
5.4 **Deaths in the household, 2011**

The question asks if there was a death in the household and if so, the month and the year of death, and the sex and age of the deceased. The question then goes on to ask the respondent questions to identify whether the death was natural or non-natural and also to ascertain if the death was a maternal death. The recording of a death in the household in the past 12 months by month of death, sex and age is contained in the mortality module in Appendix 4. The concern with the household death data is that the version of Census 2011 published in 2012 (Stats SA 2012a), is different from the final version of the Census 2011 mortality data, tables of which appear in the discussion document (Stats SA 2014a). Originally, the deaths in the household data included a large proportion of unspecified ‘age at death’, however, subsequently significant scanning errors were identified and the data ‘cleaned’, although exactly how this was achieved is less than clear (Stats SA 2014a).

The errors were attributed to the mortality schedule on the final page of the questionnaire (M-00 to M-08) not being covered by a cover-page and therefore exposed to the general effects of handling and the environment, introducing the risk of markings and consequent misreading during the scanning process (Stats SA 2015a).

The data originally released documented 604,544 household deaths, however, 20 per cent of the deaths (115,221) were recorded with unspecified age and/or unspecified sex. Comparison of a sample of the scanned questionnaire images with the recorded data revealed that during the scanning process, erroneous reading by the scanners occurred in about 21 per cent of the records where the answer to the question of whether a death occurred in the household was actually “No”. This finding suggested that the marks on the questionnaires were the cause of the misreading errors.

Remedial action involved manually recapturing a total 115,221 records (or 468,067 weighted records) with unspecified age and sex, resulting in a 22 per cent reduction in the number of deaths in the final data set (Stats SA 2014a).
Figure 5-4 presents the distribution of deaths under the age of five years occurring by month between 10th October 2010 and 9th October 2011. The peak in June is consistent with a May to June peak consistent over several years described in the 2011 RMS estimates, due to seasonal pneumonia (Bradshaw, Dorrington and Laubscher 2012: 14). However, combining the 2010 and 2011 October deaths result in 11.8% more deaths than in June, suggesting the respondent’s difficulty with accurately recalling the exact date of death. This suggests possible over-reporting of overall under-5 deaths.

Estimating the level of infant and under-5 mortality implied by the reported household deaths and comparing this with another set of reliable estimates is the only option for assessing the performance of deaths in the household question. The mortality estimation is based on the expected number of deaths produced by Darikwa in 2006.6 and the mortality based on a life-table calculation of household deaths occurring in the year preceding the census (mid-point at 2011.27). The probability of dying by calendar year for the next five years is calculated from the VR deaths corrected for completeness. This is achieved by linear interpolation to calculate the annual change in completeness between 2006 and 2011. This was then applied to the VR deaths in the intervening years to produce a series of adjusted under-1 and under-5 deaths by calendar year. The resulting age-specific trends are compared against the age-specific trends produced by the Rapid Mortality Surveillance (Dorrington, Bradshaw, Laubscher et al. 2014) and are shown in Figure 5-5.
Figure 5-5: Comparison of childhood mortality trends (adjustment1 \( q_0 \), \( q_1 \), \( q_0 \)) using household deaths with the Rapid Mortality Surveillance, 2013

The comparison suggests more deaths under the age of five reported in households than estimated by the RMS (44.2 vs 40.0 per 1,000). Child mortality was achieved by subtracting the infant from the under-5 mortality for both data sources. A second error of misreporting relevant to these data is that of heaping of deaths at age one, which has implications for the resulting \( q_1 \).

Turning to the deaths reported by households provincially, life-table measures using the number of reported deaths by age and the enumerated population are compared in Figure 5-6 with mortality indices calculated using the births produced by reverse survival of the population born in each province to produce the number of provincial births as described in the previous section.
Figure 5-6: Comparison of provincial mortality indices produced from the reported household deaths at approximately February 2011

Confidence intervals have not been provided owing to the fact that calculation is complicated by not only for the allowance of PES adjustment to person data, but unknown uncertainty about the estimate of the numbers of deaths reported by households.
The national indices produced from both methods are very consistent. This is not the case, however, for the provincial estimates which is probably due to the lack of correspondence between the estimated births and the reported deaths at a provincial level. Considering the IMR, there is good agreement for the WC, LP, FS, NW and to a lesser extent MP and NC. On the other hand, the EC, GT and particularly KZN do not match well. The error could arise either from the estimated number of provincial births based on reports of these young children by the province of birth or from the deaths, reported by the mother or the head of the household in the province of residence at census date.

The under-five mortality rates produced by both methods have the potential of introducing more error in the estimate of births depending on the stability of the number of births over the last five years. This is because the calculation of the U5MR using the same births for the IMR assumes that the number of births for the previous four years remained the same. Non-correspondence of the births and deaths on a provincial basis is also likely to increase with the age of the child as the probability of inter-provincial movement increases with age.

The life-table calculation limits the error arising from the inter-provincial flow of births. However, the possibility of the death being reported to have occurred in the province of residence (of the adult respondent) and not actually, where the death of the child occurred introduces a possible mismatch of numerator and denominator. For example, according to the vital registration in 2011\textsuperscript{17}, proportions of deceased children under the age of five who did not die in the province in which they lived are as follows: EC 7.6\%, NW 6.0\%, WC 6.1\%, LP 3.4\%, GT 6.3\%, KZN 3.5\%, FS 3.6\%, MP 4.9\%, NC 6.4\%.

The census question about migration found that the following proportions of children under the age of five years born in a particular province were enumerated in a different province: EC 9.4\%, NW 6.9\%, WC 7.5\%, LP 11.7\%, GT 6.5\%, KZN 3.4\%, FS 7.6\%, MP 7.0\%, NC 15.6\%. However, these proportions and some of the interprovincial proportions indicating specific movement from one province into another appear highly suspect and therefore suggest errors, possibly due to the scanning procedure.

\textsuperscript{17} These proportions were derived from the 2011 Vital Registration dataset provided to the SAMRC by Stats SA.
5.5 **Children ever born/children surviving 2011**

At the end of 2015 Stats SA released a monograph of mortality estimation based on the 2011 Census (Stats SA 2015a) which included infant and under-5 mortality derived from the summary birth history SBH data. Early in 2016, Stats SA made a 10% sample of data required for the analysis of CEB/CS module available for research purposes (personal communication, R E Dorrington).

The format of the question differed from that of the previous census (Figure 5-1), in that the answer about the children surviving were separate for children still living in the household and those living elsewhere and by sex. These data revealed illogical answers and in order to improve the estimates it was decided that data for women with inconsistent responses be dropped from the analysis. The inconsistencies were identified as any of the following:

a) the Total Children Ever Born (TCEB) did not equal the total male plus total female CEB

b) the TCEB did not equal to the total surviving children living in the household and total living elsewhere plus the total dead children

c) the total female CEB did not equal the total female surviving children living in the household and total living elsewhere plus the total female children dead

d) the total male CEB did not equal the total male surviving children living in the household and total living elsewhere plus the total male children dead.

Table 5-1 shows that overall, 12.4% of women’s data lacked internal consistency. In comparison, the analysis undertaken by Stats SA found only 9.5 % of inconsistent women’s data (Stats SA 2015a). The 2011 Census questionnaire was identical to the 2007 Community Survey, and the same problem identified where 8% of the data was excluded from the analysis due to these inconsistencies (Darikwa 2009).

<table>
<thead>
<tr>
<th>Age-group</th>
<th>Women</th>
<th>Parity unstated</th>
<th>Children consistent with mother’s age</th>
<th>% inconsistent with mother’s age</th>
<th>Children ever born</th>
<th>Children surviving</th>
<th>Children dead</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-19</td>
<td>2,504,905</td>
<td>584,928</td>
<td>1,900,585</td>
<td>24.1</td>
<td>383,683</td>
<td>369,963</td>
<td>13,146</td>
</tr>
<tr>
<td>20-24</td>
<td>2,679,896</td>
<td>310,137</td>
<td>2,308,839</td>
<td>13.8</td>
<td>1,763,463</td>
<td>1,678,118</td>
<td>58,179</td>
</tr>
<tr>
<td>25-29</td>
<td>2,516,635</td>
<td>166,227</td>
<td>2,267,431</td>
<td>9.9</td>
<td>3,070,249</td>
<td>2,898,937</td>
<td>108,823</td>
</tr>
<tr>
<td>30-34</td>
<td>1,992,804</td>
<td>90,499</td>
<td>1,823,247</td>
<td>8.5</td>
<td>3,510,575</td>
<td>3,298,969</td>
<td>127,434</td>
</tr>
<tr>
<td>35-39</td>
<td>1,758,420</td>
<td>62,119</td>
<td>1,620,420</td>
<td>7.8</td>
<td>3,983,835</td>
<td>3,741,839</td>
<td>147,088</td>
</tr>
<tr>
<td>40-44</td>
<td>1,546,291</td>
<td>48,872</td>
<td>1,420,340</td>
<td>6.1</td>
<td>4,132,207</td>
<td>3,840,475</td>
<td>182,338</td>
</tr>
<tr>
<td>45-49</td>
<td>1,424,543</td>
<td>49,857</td>
<td>1,294,705</td>
<td>9.1</td>
<td>4,179,564</td>
<td>3,821,628</td>
<td>245,849</td>
</tr>
<tr>
<td>Total</td>
<td>14,423,494</td>
<td>1,312,639</td>
<td>12,635,567</td>
<td>12.4</td>
<td>21,023,578</td>
<td>19,649,929</td>
<td>882,658</td>
</tr>
</tbody>
</table>
The method and adjustments used to estimate the under-5 mortality rate in this analysis are explained in section 3.6.2. The resulting trend is compared with the published Stats SA’s estimate (Stats SA 2015a: 27) and the 2013 RMS (Dorrington, Bradshaw, Laubscher et al. 2014) in Figure 5-7.

Figure 5-7: Comparison of under-5 mortality trend with Stats SA, derived from the summary birth history and the Rapid Mortality Surveillance 2013

The CEB/CS estimate of $5q_0$ without adjustment for mother’s survivor bias (diamond), suggests that between 2000 and 2009 the under-5 mortality rate decreased slightly from 38 to 30/1,000 live births and is clearly too low. The $5q_0$ adjustment (W&Z) trend (square) implies the mortality is about 12/1,000 higher in 2009, or about 48/1,000 live births. The $5q_0$ adjustment (D&D) trend (triangle) has the effect of reducing the mortality from adjustment (W&Z), of the children born to the older women between 35 and 49, but has a small effect on the mortality of the younger women’s children, those between 20 and 34 years of age. From 2009 the mortality produced by Stats SA and the RMS are very similar, while in contrast the mortality implied by both adjustments are lower (excepting for estimates around 2000 based on responses from the older women represented by adjustment (W&Z)). Both trends derived by the SBH produce lower mortality rates than the RMS, which does not require an adjustment for mother’s survivor data, although Stats SA does not indicate that they made any adjustments in their estimation procedure.

In the context of the 2011 Census analysis of the SBH, as we have seen, these data require adapting the Ward and Zaba (2008) adjustment for mother’s survivor bias.
to also account for the changing prevalence over time using the adjustment proposed by Darikwa and Dorrington (2011). However, a further adjustment to allow for the increasing proportion of women on ART since about 2001 is also necessary. The Thembisa model provides the proportions of women aged fifteen years and over on ART and the proportion of women in need of and receiving ART. By multiplying the adjustment factors for changing prevalence by the proportions of women requiring ART but not receiving treatment (RE Dorrington personal communication), appears to appropriately correct for the impact of ART intervention on the CEB methodology. The trend shown in Figure 5-8 demonstrates the performance of the proposed third adjustment for ART (D&D ART), showing that the CEB 2009 mortality rate corroborates the mortality produced from the RMS. It is important to note that the application of this adjustment for ART used raw CEB/CS data, not cleaned for inconsistencies, hence uses approximately 12 percent more data than the trend shown in Figure 5-7 (RE Dorrington personal communication). In other words, the proposed adjustment is appropriate, but unfortunately, the quality of the 2011 Census SBH data do not allow for the stringent consistency checks outlined in this section (Table 5-1).

![Figure 5-8: Comparison of under-5 mortality trend derived from the summary birth history including an adjustment for the increasing proportion of women on ART with Stats SA, and the Rapid Mortality Surveillance 2013](image)

Aside from the adjustments made for maternal survivor bias, the other underlying assumption of the method is related to stable fertility, which is contradicted by the two sets of data (the enumerated population and the derivation of births by reverse survival)
which provide evidence that South Africa’s fertility has experienced an unexpected increase since the last census.

Finally, an important point is that the CEB/CS questions do not appropriately reflect changing trends (the upward trend between 1992 and 2005, followed by the downward trend between 2005 and 2015) shown in Figure 5-8, which are also found in the South African DSSs.

5.6 Preceding birth technique 2011
The method of estimation is explained in sub-section 3.7 yielded unrealistically low levels of mortality centred on October 2010 of \( q_0 = 13.1 \), \( q_0 = 15.7 \) and \( q_0 = 18.6 \) per 1,000 live births. The unsuitability of the method relates to very long birth intervals (Hill 2013).

5.7 Discussion
Assessments of the census questions suggest that some data are more suitable than other data for the estimation of childhood mortality. The denominator data (births and the population) and household deaths data offer point estimates centred on the census date. Although the methods presented produce very similar results nationally, some provincial differences are substantial and point to issues related to the non-correspondence of the estimated births and the reported deaths. The proportions dying in a province other than the one they live in highlight extensive inter-provincial movement of young children. Migration may be responsible for the differences in the mortality rates produced by the method involving births; hence, the life-table method restricts the uncertainty introduced by the use of births and is therefore possibly a better alternative for the provinces. The utility of the number of births produced from reverse survival to assess the completeness of birth registration provincially is explored in Chapter 7.

There is evidence that the household death question may produce over-reporting of deaths under the age of five in the 2011 Census. Notwithstanding this observation, the heaping of deaths at age one appears to be a more serious problem with these data. This is investigated further in Chapter 7.

The PBT has the potential to produce a point estimate, but is clearly does not work on South African data. As other demographers have documented, concerns about the data quality of survival of births in the last 12 months and selection bias in the case
of births in the last 24 months means that the method is not recommended (Hill 2012: 143).

The SBH offers the opportunity to derive a mortality trend; however, this analysis demonstrates that the 2011 data require a third adjustment that accounts for the impact of the ART rollout on HIV-related mortality. Importantly, this analysis also demonstrates the extent of the poor data quality.
6 MEASURING CHILDHOOD MORTALITY IN A HEALTH AND DEMOGRAPHIC SURVEILLANCE SITE

The Africa Centre Demographic Information System (ACDIS) was established to describe the demographic, social and health impacts of the HIV epidemic in a population undergoing health transition (Hosegood and Timaeus 2005). Other objectives are to generate demographic, health and related information on the geographic distribution of death and disease, socioeconomic dynamics and migration patterns of a rural population. The ACDIS data are used to investigate measuring and monitoring childhood mortality in a highly mobile population.

Both the episode data collected prospectively from 2000-2007 and the retrospective data collected in 2000 using the full birth history are examined. The chapter begins by describing the data sources, the data collection procedures in the surveillance area, and assessing data quality. The full birth history conducted at baseline census provides an opportunity to examine socioeconomic correlates of child mortality at the beginning of the surveillance in 2000.

Investigation of the longitudinal data reveals three areas of interest that are explored in the remaining sections of the chapter. The first is the difference between the retrospective and prospective mortality estimates derived in 2000. Estimates of the under-two mortality by birth cohorts by Ndirangu, Newall, Tanser et al (2010) are compared with the estimates derived in this chapter using the same data. Regression techniques exploring the relationship between age, sex and period, provide context for interpretation of the observed age-specific trend. The chapter ends by considering the strengths, limitations and lessons for the measurement of child mortality from in a health and demographic surveillance site data.

6.1 Data sources
6.1.1 Baseline Census, 2000
The Umkanyakuda district in KwaZulu-Natal covers about 438 km² and in 2011 the population was approximately 89,000 people living in approximately 12,000 households (Channon, Hosegood and McGrath 2016). It is characterised by great variation in population density of between 20 and 3,000 people per km². While the subject of the previous chapter was measurement of childhood mortality using retrospective history, this chapter contrasts measures of early age mortality using longitudinal data collected
from an open cohort of women and their children between 2000 and 2007 as well as assess the performance of the demographic surveillance.

During this period, considerable infrastructural developments occurred in the area as well as changes in other factors that directly affect levels of childhood mortality. In particular, the area experienced substantial improvements in the determinants of early-age mortality since 2001 such as an increase in the proportion of households with access to free public taps, which increased from 4.5% in 2001 to 42.1% in 2006 and pit latrines increased from 3.5% to 23% during the same period (Muhwava 2008). Apart from household conditions, the uptake of social remittances particularly the child support grant and most influential the introduction of PMTCT programmes in some districts at the end of 2001 as well as the provision of ART to adults, all occurred during the observation period.

The baseline census in 2000 collected full pregnancy history information collected for assessing levels of fertility and child mortality retrospectively very similar to the format used in DHSs except the questionnaire asked about pregnancy outcomes of which a live birth is one possible outcome, unlike the 1998 SADHS, which begins by asking about live births and probes about stillbirths. The women’s data covered all women who were registered either as residents or as non-residents of the DSS at any date between 1 January 2000 and 31 December 2000.

A third data set provided individual and household level socioeconomic information to explore the relationship between socioeconomic status and childhood mortality. The household and individual socio-economic status data were merged with the event history data to explore the factors associated with childhood mortality using regression techniques.

6.1.2 Longitudinal open cohort
The ACDIS cohort comprises household members and distinguishes between membership and residency within the surveillance area, although the individual does not have to be actually resident. “Crucially, this means that ACDIS collects information on resident and non-resident members of households and makes a distinction between membership (self-defined on the basis of links to other household members) and residency (residing at a physical structure within the surveillance area at a particular point in time)” (Tanser, Hosegood, Barnighausen et al. 2008: 2).
Fieldworkers are assigned bounded structures, identified by geographic information system coordinates, and collect different data continuously throughout the year, such as the routine surveillance data collected every four to six months or the socioeconomic data collected biennially. Events data records births and deaths of resident and non-resident children provided by key household informants, which is the focus of this chapter. Between 2000 and 2003, these routine surveillance data were collected three times a year via household interviews, from 2004 it has been collected only twice a year. Pregnancy and birth information are captured using a pregnancy history questionnaire where the outcome is a live birth, stillbirth, miscarriage or an induced abortion. In the past, a pregnancy notification was completed, where a pregnancy is observed during a household interview. The pregnancy notification was dropped with the shift from tri-annual to biannual surveys in 2004, meaning that the follow-up of pregnancy from notification to the outcome of that pregnancy was lost after that.

6.1.3 Episode data
The ACDIS Demography Dataset 2000-2007 includes unit record details of each episode for persons registered in the DSA during the period 2000-2007. These data are organised to record life-course events of every individual as “episodes” that are opened and closed by events such as births, deaths and migration. Episodes are defined by the length of time spent in a particular state before moving into another state for each period between surveillance rounds in each year. These periods define individual exposure. Age group and residency are key criteria that determine status and the corresponding exposure. A change in status marks the end of the current episode and the start of a next episode. Individual episodes are indicated as either resident or non-resident.

6.2 Quality of the retrospective data
These data were provided in a file with women’s pregnancy histories and a file with children’s data. The ‘women’ file included detailed information for each pregnancy outcome of 33 220 women and the data on children contained 39 656 recorded births and 4 413 deaths. Table 6-1Table 6-1 shows the proportion of women and child pairs by age of the women interviewed in 2000 and the recorded births. Table 6-2 shows the survival status of all children recorded by sex.

Table 6-1: Recorded births by age of mother at time of the survey
### Table 6-2: Recorded survival status by sex

<table>
<thead>
<tr>
<th>Sex</th>
<th>Died</th>
<th>Yes</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unknown sex</td>
<td>722</td>
<td>225</td>
<td>947</td>
</tr>
<tr>
<td>Female</td>
<td>17 381</td>
<td>1 954</td>
<td>19 335</td>
</tr>
<tr>
<td>Male</td>
<td>17 121</td>
<td>2 234</td>
<td>19 355</td>
</tr>
<tr>
<td>Missing</td>
<td>19</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>35 243</td>
<td>4 413</td>
<td>39 656</td>
</tr>
</tbody>
</table>

6.2.1 **ACDIS retrospective child mortality trend**

In addition to presenting a detailed analysis of the level and trend of child mortality, the ACDIS data are linked to socioeconomic information to explore factors associated with childhood mortality. A comparative analysis of the ACDIS full birth history trend with other sources of representative data is also undertaken to assess the accuracy and plausibility of each estimate as another way to evaluate the full birth history.

The analysis excludes events that occurred after 31 December 2000 and uses pregnancy history information of women of reproductive ages to calculate the exposure to the risk of death for the period 1981–2000. The period mortality indices are calculated for all children who died before their fifth birthday. There were a total of 39 656 live births and 4 413 deaths during this period. The mortality trend for boys and girls is presented in Figure 6-1.

There were very similar numbers of births for both sexes. Of the 42 609 recorded live births, 48.9% were boys and 48.8% were girls. Sex was either missing or undetermined in 2% of the total live births. The mortality trend shows substantial declines in overall under-five mortality during the 1980’s experienced by children aged 1-4 as well as in the post-neonatal period. Boys had higher numbers of deaths (53%) than the girls (47%). The reversals that began in the early 1990’s were most pronounced in the post-neonates, but also noticeable in children age 1-4 years.
6.2.2 Comparison with other estimates

A comparative assessment of the ACDIS FBH trend with four alternative under-five mortality estimates is undertaken and shown in Figure 6.2. The estimates of the under-5 mortality rate are derived from the HSRC 1992 survey, FBH, the 1996 Census SBH, episode data from the ACDIS open cohort and the ASSA2008 AIDS and demographic projection model.

6.2.2.1 Human Science Research Council Survey, 1992

In 1992 the HSRC conducted an infant and child health survey in the then provinces of South Africa and the former homelands and self-governing territories (Rossouw and Jordaan 1997). Survey methodology very similar to that of the Macro International series of DHS’s including the collection of detailed birth histories, was used. Richards Bay, an area in close proximity to the ACDIS population was one of eight regions included in the KZN sample. Consensus around the high quality of these data sets are confirmed by the congruence in trend provided by this survey with the 1998 SADHS for example (Department of Health 2002).
6.2.2.2 Full birth history

The diagnostics shown in the previous section demonstrate the ACDIS data to be of good quality showing no indication of common deficiencies found in retrospective data. The 1990 to 1995 period is characterised by fluctuations due to the sampling variation caused by annual disaggregation of the data. The probability of dying before age five calculated as described in sub-section 3.4.1, was 116 in 2000, 15% higher than the ASSA model and the ACDIS episode data.

6.2.2.3 1996 Census for black Africans in KwaZulu-Natal province

The 1996 Census collected SBH data to estimate the under-5 mortality of African children in rural KwaZulu-Natal. Technical considerations such as the appropriateness of the choice of model life-table (Dorrington, Bradshaw and Wegner 1999) more especially in an AIDS environment, are central to this estimation procedure. This estimate agrees with the HSRC and ASSA estimates around 1990, although the trend further in the past becomes relatively higher and discordant with the other estimates.

6.2.2.4 Episode data

Subsequent to the initial baseline census, the surveillance area began conducting routine household visits in 2000, collecting event data prospectively. Unlike the summary and full birth histories, this estimate is not based on mothers’ recall, but rather on continuous surveillance records and as such provides a useful comparative for the FBH estimates because the data are from the same sample of women. Section 6.5 describes the mortality calculation necessary for episode data. Interestingly, the surveillance data
in 2000 produced an U5MR of 96 per 1,000 live births, 18 percent lower than the FBH estimate. This discrepancy is investigated later in this chapter.

6.2.2.5 ASSA2008 model
The ASSA2003 AIDS and demographic projection model is described in section 2.5.18. Significant revisions made to the 2008 model include the incorporation of several interventions affecting child mortality including antiretroviral treatment (Actuarial Society of South Africa 2011). Important specifications to note here are the assumptions around the survival of HIV-infected children, which when compared with other models such as the UNPD (UN IGME 2011) estimates and those developed by the Inter-agency group (UN IGME 2011) appear to underestimate the under-5 mortality in the more recent period, 2010. The ACDIS 2000 point derived from episode data is identical to the ASSA2008-estimate of 96 per 1,000 that indicates that ASSA mortality assumptions equate to the mortality actually experienced by the surveillance population and captured by the episode data.

The ASSA trend presented here is modelled on the mortality risks of black African children in the entire province whereas the ACDIS population, although African reside in rural areas, and experience poorer living conditions which are known to generate higher mortality (Tanser, Hosegood, Barnighausen et al. 2008). Although there is a fundamental difference in the calculation of period rates from FBH and rates from episode data, the primary reason(s) for this disparity are methodological and/or procedural factors, and these are the subject of sub-section 6.6.

6.3 Correlates with child mortality
This sub-section explores another utility of full birth history information, when linked to socioeconomic differentials of mothers and household by means of regression analysis. Full birth histories permit individual level data analysis, which have contributed substantially to knowledge of the determinants of child mortality. The aim of this regression analysis is to investigate the relationship of socioeconomic factors on child mortality in a relatively homogenous rural population.

The model building strategy involved an initial model including age, sex and the time trend. To the basic model, single variables were added to explore the strength of relationship of a particular variable on child mortality. This followed fitting the full model comprising all the socioeconomic and health care access variables. Statistical significance of the additional variables was determined using a “post-test”. The variables
that were not significant at the 5% level of significance were dropped from the final model. Lastly, a separate model was fitted in a similar way to data on births in the last six years before the survey.

Assignment of household area-type was based on the population density. The non-urban population was divided into rural (areas of less than 400 residents per square kilometre) and peri-urban (areas with more than 400 residents per square kilometre) (Tanser, Hosegood, Barnighausen et al. 2008). Household wealth was measured using an asset index allocating household assets into low, medium or high socioeconomic status based on this simple distribution (Hosegood, McGrath, Herbst et al. 2004).

The results of the regression are presented as relative risks and p-values of the covariates of the basic model of age, sex and period and the same estimates of the final model in Table 6-3. While the results of regression depict an unclear mortality trend from 1983 to 1994, with peaks and troughs, where the raw data show a decline up until 1992, followed by an increase, the model building process reveals the strength of the associations with and the contribution to child mortality of each variable. The final model captures the additional effects of water source, maternal education and being a recipient of the child support grant. The model-building process shows that even after controlling for age, sex and period, all these factors display significant associations with childhood mortality in 2000.

As well as the final model, a separate model explored the associations of the covariates on the mortality of children born in the six years before the survey to address the loss of cohesion between the current household socioeconomic circumstance and the prevailing conditions of the more distant births. In this model the child grant, instituted in 1998, had a greater protective effect on children living in households that were receiving the grant in 2000. The relative risk is 1.79 (95% CI 1.39-2.29) for non-recipients, but the association with the level of maternal education becomes statistically insignificant, with RR of 1.28, 1.06 and 0.83 for incomplete primary, incomplete secondary and, complete secondary and above, respectively.

These results provide a suitable context in which to discuss another limitation of the FBH that of the associations as reported at the time of the survey and the attenuation of those associations further back in the past. From the socioeconomic variables explored, improved water source, higher levels of maternal education and being a recipient of the child support grant at the time of the survey were significantly positively associated with childhood mortality in the fully adjusted model. The
modelling strategy showed that controlling for urban-rural residence makes sanitation less important than water source, but maternal education remains important and has a greater effect on child mortality, as does being a recipient of the child support grant. Although some of the effect of an incomplete secondary education is due to motherhood at an early age, this is small as there were only 4.2% of mother and child pairs represented in the 15-19 year old age group.

Restricting analysis to births in the last 6 years shows that while the relative risk of being a recipient of the child grant remains statistically significant, the effect of maternal education at the time of the survey on child mortality disappears. The attenuation in the relationship with maternal education status is likely to be due to better access to grants by more educated women (Cleland 1990) given the well-established fact that education acts as a proxy for socioeconomic status of the family (Desai and Alva, 1998).

In the context of modelling FBH data there are limitations to this approach. The instrument delivers a cross-sectional estimate at a single point in time. The retrospective nature of these data limits the investigation of socioeconomic factors to a snap-shot centred on 2000 information as it relates the household conditions at the time of the survey. Notwithstanding this, there are general relationships observed from the historical mortality trend and the socioeconomic characteristics of the households, particularly for variables not likely to have changed rapidly over the time. However, it cannot be assumed that the current (at survey date) socioeconomic status applies to the older births and therefore it is sensible to restrict births to recent births, which to some extent limits this effect, as done in this analysis.

18 Although an association between education and socioeconomic status would not necessarily imply better-educated women are more likely to claim remittances, Cleland’s proposed three components of the “empowerment of women through education pathway”, is quite compelling. These are: 1) instrumentality, the ability to manipulate and feel control over the external world. 2) social identification, concerned with the ability to engage with institutions and bureaucracies and 3) confidence, which is largely indirect, therefore can be understood as, more educated women better utilize for example health services for themselves and their dependents, hence to varying degrees they empower themselves.
Table 6-3: Negative binomial regression estimates of risk ratio for childhood mortality

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Basic Model</th>
<th>Full Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deaths</td>
<td>RR</td>
</tr>
<tr>
<td>Years before survey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998-2000</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1995-1997</td>
<td>0.77</td>
<td>0.67</td>
</tr>
<tr>
<td>1992-1994</td>
<td>0.69</td>
<td>0.60</td>
</tr>
<tr>
<td>1989-1991</td>
<td>0.54</td>
<td>0.46</td>
</tr>
<tr>
<td>1986-1988</td>
<td>0.71</td>
<td>0.60</td>
</tr>
<tr>
<td>1983-1985</td>
<td>0.69</td>
<td>0.59</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Boy</td>
<td>1.18</td>
<td>1.08</td>
</tr>
<tr>
<td>Age group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;28 days</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>1-2 months</td>
<td>0.16</td>
<td>0.01</td>
</tr>
<tr>
<td>3-5 months</td>
<td>0.07</td>
<td>0.14</td>
</tr>
<tr>
<td>6-11 months</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>12-23 months</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>24-35 months</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>36-47 months</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>48-59 months</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>Water source</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Piped private/public</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Borehole/well/rainwater</td>
<td>0.95</td>
<td>0.83</td>
</tr>
<tr>
<td>River/dam</td>
<td>1.77</td>
<td>1.51</td>
</tr>
<tr>
<td>Missing socioeconomic data</td>
<td>No Education</td>
<td>1.00</td>
</tr>
<tr>
<td>Mother’s education</td>
<td>Complete primary</td>
<td>0.94</td>
</tr>
<tr>
<td>Complete secondary</td>
<td>0.78</td>
<td>0.67</td>
</tr>
<tr>
<td>Matric &amp; above</td>
<td>1.28</td>
<td>1.05</td>
</tr>
<tr>
<td>Missing socioeconomic data</td>
<td>Receiving the child grant</td>
<td>0.00</td>
</tr>
<tr>
<td>Child support grant</td>
<td>Non recipients</td>
<td>1.36</td>
</tr>
</tbody>
</table>
6.4 Quality of the episode data
The data of interest are the births and deaths of children under the age of five years registered within ACDIS from January 1st 2000 up until December 31st 2007. The dataset has 1,517,165 episodes of which 258,092 are for children under-five years of age. Of these 239,926 (93%) are resident episodes.

6.4.1 Births and deaths
A cumulative total of 14,994 pregnancies resulting in 15,193 live-births from 2000 to 2007 were recorded in the surveillance area. A steady increase in the number of births over the eight-year period with an increasing proportion of the births occurring outside of the surveillance area is shown in Figure 6-3. The number of births occurring during episode of non-residence increased from under 1% of the 1,652 births in 2000 to almost 13% in 2007.

![Graph showing number of live births by residence status, 2000-2007](image)

Figure 6-3: Number of live births by residence status, 2000-2007

Although the number of live births increased by 15.7% over the eight years, the trend was not quite monotonic with a slight decrease in 2003 followed by an increase of 11.1% between 2003 and 2004, thereafter a decline.
6.4.1.1  Sex ratio at birth

The sex ratio at birth was calculated by utilising a generated date of birth variable for each live birth to generate the number of births by sex as the variable ‘live-birth count’ is not coded by sex.\footnote{This was achieved by using the date variable that accompanies each episode of ObservationStart and ObservationEnd. Thus where Episode = 1, AgeGrp =1 and AgedIn =1; the ObservationStart will be the date of birth.}

The sex ratio fluctuates between 106.6 and 99.1 (Figure 6-4) giving an average of 102.5 over the eight-year period, which, although it is consistent with the sex ratio of 103 reported for African populations, it is higher than would be expected for eastern and southern African populations (Garenne 2004). However, given the wideness of the confidence interval (CI 113.92- 99.30), neither of these results is significant.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ratio.png}
\caption{Sex ratio at birth, 2000-2007}
\end{figure}

6.4.1.2  Child exposure

There were 128,508 girl-episodes and 129,579 boy-episodes for children under-five, resulting in 258,087 with sex specified (only five episodes had sex missing). The number of episodes in the neonatal age group is shown in Figure 6-5. This would be expected to follow a similar trend to the number of live births. There are consistently more boy-episodes than girl-episodes in the first age group. It is not clear why there was a decrease in both boy and girl neonatal episodes of exposure in 2003.
6.4.1.3 Components of under-five deaths

A total of 1,241 deaths (one record had missing sex) under the age of five were recorded throughout the eight year period. The annual number of deaths declined from 174 in 2000 to 102 in 2007. Overall under-five deaths peaked in 2002 then decreased sharply between 2003 and 2004 (Figure 6-6). Of note is the considerable decline in neonatal deaths from 31 in 2003 to 9 in 2004. The highest number of deaths during non-residence occurred in 2004 (17 deaths) followed by 2002 (13 deaths).

The trend in the number of deaths for girls and boys by age group is shown in Figure 6-7. Most deaths occur in the post-neonatal period, and this is where most of the decline has occurred. Both the neonatal and post-neonatal trends are quite different for boys and girls, whereas the numbers of deaths at the older ages are similar. There have been impressive declines in child deaths in all four age groups, over eight years. Noteworthy is a distinctive fall from 2003 to 2004 in all age groups but particularly
dramatic for boy neonates. The neonatal deaths declined by 58% overall, with noticeable differences by sex.

The post-neonatal pattern is similar for each sex starting at 46 and 52 in 2000 for boys and girls respectively. There is a 56% decline in this age group over the period. In 2002, there is a far greater difference in boy/girl deaths in the post-natal period, which is reflected in the number of infant deaths. The overall infant trend shows that infant deaths rose from 2000 to 2003 and dropped in 2004 (Figure 6-7).

![Graph showing number of age-specific deaths for boys (top panel) and girls, 2000-2007](image)

**Figure 6-7: Number of age-specific deaths for boys (top panel) and girls, 2000-2007**

Exploration of the data show an absolute decline in deaths for boys and girls over the period, however, the pattern is not the same for boys and girls. Of note in terms of the data quality is the decrease in both boy and girl neonatal episodes of exposure in 2003 and the sharp decline in neonatal and post-neonatal deaths 2003-2004. These changes coincide with a change in data collection in 2004 from thrice-yearly visits to twice-yearly and the use of pregnancy notification form for the collection of pregnancy details was also dropped.
6.5 **Longitudinal childhood mortality trend, 2000-2007**

The age-specific mortality risks calculated per 1,000 live births shown in Figure 6-8 reflect the mortality of all children in the DSA, regardless of their residence status. As observed with the deaths, all the age-specific mortality rates declined over the eight years.

![Figure 6-8: Age-specific mortality of all children in the surveillance area, 2000-2007](image)

Between 2000 and 2007 the neonatal rate declined from 13 per 1,000 live births to 5 (61% decline) while the post-neonatal rate from 55 to 23 (59% decline). The infant mortality rate declined from 66 to 28 (58% decline), the child mortality rate from 31 to 25 (19% decline) and the overall under-5 mortality declined from 95 per 1,000 to 52 representing (46% decline). The most significant decrease for all age groups occurs between 2003 and 2004, except the 1-4 year old rates, where the decline between 2002 and 2003 was 15% versus 13% between 2003 and 2004.

Childhood mortality in the HDSS for each calendar year from the beginning of 2000 up until the end of December 2007 is set out in Appendix 5.

Conversion of exposure from person-days into person-years for the six age intervals represented in life table requires a weight for the first two age groups where the exposure is not for the entire first year of life. Hence, the denominator for the neonatal period would reflect one month and the post-neonatal period would require eleven of the possible twelve months of denominator exposure.

The remainder of the chapter focuses on three issues of measurement highlighted in the surveillance data. The first of which pertains to the difference between the estimates derived by the full birth history and those derived from the longitudinal data already encountered. Second, a comparison of the child mortality estimates with those
published by Ndirangu, Newell, Tanser et al (2010) and third interrogation of the trend in age and sex pattern of mortality as a counter argument to the conclusions proposed by Ndirangu and colleagues.

6.6 Reconciling the retrospective and prospective estimates of infant mortality in 2000

Comparison of the estimates derived by the two systems shows estimates based on retrospective reporting of experience produced higher mortality than the routine surveillance in all age groups in 2000 (Table 6-4). The ratio of the estimate from routine surveillance to that based on the birth history is 80% for neonates, 74% for post-neonates, 75% for infants and 103% for children age one to four, but it is impossible to know which value better represents the ‘true rate’. Furthermore, the routine surveillance produced consistently higher estimates of mortality for infant boys than infant girls yet the episode data in 2000 had more girl deaths than boy deaths.

<table>
<thead>
<tr>
<th>Mortality Index</th>
<th>Estimates based on birth history (per 1,000)</th>
<th>Estimates based on routine surveillance (per 1,000)</th>
<th>Routine surveillance relative to birth history (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neonatal</td>
<td>15</td>
<td>12</td>
<td>80</td>
</tr>
<tr>
<td>Post-neonatal</td>
<td>75</td>
<td>55</td>
<td>74</td>
</tr>
<tr>
<td>1q0</td>
<td>89</td>
<td>67</td>
<td>75</td>
</tr>
<tr>
<td>4q1</td>
<td>30</td>
<td>31</td>
<td>103</td>
</tr>
<tr>
<td>5q2</td>
<td>116</td>
<td>96</td>
<td>82</td>
</tr>
</tbody>
</table>

There are two possible reasons for these discrepancies. The first is that the disparity results from slightly different methods of calculation, but given the magnitude of the differences, this is unlikely. The second possibility relates to the question of whom exactly (which mothers) reported on which children in both data collection processes. In order to reconcile mother and children in both data sets the respondents and their children were identified in order to understand the exposure contributed in both calculations. This involved establishing a mechanism to cross-reference the child in both data systems and verify their identities. These processes identified all the infants in the retrospective data in the year 2000 who were then sought in the prospective data set using the Child's Individual Internal Identification.

There were 175 infant deaths identified in the retrospective data, of which only 102 (58%) could be located in the surveillance data (Figure 6-9). Of the 73 infants not found in the surveillance data, one case was a duplicate record in the retrospective data, leaving 72 infant deaths (41%) not found in the surveillance data. It was, however,
possible to explore these children’s details using the variable *Mother’s Individual Internal Identification*, which identified 49 children reported in the area surveillance and revealed 43 mothers who did not appear on the surveillance data and seventeen deaths not found at all.

![Diagram: Infant deaths recorded in the retrospective data in 2000 and found in the prospective surveillance records](image)

**Figure 6-9: Infant deaths recorded in the retrospective data in 2000 and found in the prospective surveillance records**

Examination of the characteristics of the 49 children revealed a large proportion of them had quite mixed patterns of residency. Commonly babies had been born outside of the HDSS although the mother was resident the episodes of residency and non-residency varied quite a bit.

The three groups of children identified as either in the birth histories, in the area surveillance or in both systems (Table 6-5) also displayed quite different mortality and further inspection revealed that children living away from their mothers have a relative risk of dying over twice as high (2.1) as children who lived with their mothers in the DSA.
Table 6-5: Comparison of demographic details of three groups of infant deaths reported in birth histories and routine area surveillance in 2000

<table>
<thead>
<tr>
<th></th>
<th>Birth history only</th>
<th>Birth history &amp; routine surveillance</th>
<th>Routine surveillance only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Headcount</td>
<td>926</td>
<td>2,877</td>
<td>753</td>
</tr>
<tr>
<td>Exposure (person years)</td>
<td>435</td>
<td>1,423</td>
<td>338</td>
</tr>
<tr>
<td>Deaths</td>
<td>65</td>
<td>98</td>
<td>20</td>
</tr>
<tr>
<td>Death rate</td>
<td>88</td>
<td>69</td>
<td>68</td>
</tr>
<tr>
<td>$t_{60}$</td>
<td>84</td>
<td>67</td>
<td>66</td>
</tr>
<tr>
<td>RR</td>
<td>2.1</td>
<td>1.0</td>
<td>0.9</td>
</tr>
</tbody>
</table>

The emerging pattern therefore suggested that living arrangements of mother and child are associated with increased mortality risk for children living away from their mothers. A possible scenario of the respondents represented in both data sets and their probable living arrangements based on residency patterns is illustrated in Figure 6-10. The birth history did not collect residency status so it is not possible to verify residency status of mothers and children found exclusively in the birth history data. However, the mortality differences associated with the mothers’ varied residency episodes, gives credence to the hypothesis linking living arrangements to the differential mortality observed in these three groups and is further supported by findings showing an increased probability of child migration caused by the death of parents and other adults within households (Ford and Hosegood 2005).

Figure 6-10: Mother and infant residence in 2000 according to data source
Turning to the comparison of the longitudinal and cross-sectional approaches and experience from other settings, the Agincourt HDSS also found the under-five mortality rate produced from the retrospective reports was about 36% (although from a low base) higher than the equivalent mortality based on surveillance data (Nannan, Dorrington, Laubscher et al. 2012). Conversely, an earlier investigation (Garenne and Van Ginneken 1994) in Niakhar, Senegal found that two independent retrospective surveys underestimated the under-five mortality relative to the mortality derived from surveillance records. The researchers also noted the surveillance data’s ability to detect seasonal differences in births and deaths, that both retrospective reports failed to pick up even in the most recent period.

On the other hand, the experience of the Indepth Butajira site in Ethiopia (Byass, Worku, Emmelin et al. 2007) illustrates that these are not universal traits. Comparison of longitudinal data with DHS estimates found the two independent estimates of the under-five rate to be within 10% of each other and concluded that the two approaches show overall consistency. Although over time the cross-sectional approach underestimates, relative to the longitudinal approach, which in their opinion is probably due to recall error (Byass, Worku, Emmelin et al. 2007).

Evaluation of demographic indices measured by means of surveillance methodology, have focused on what seems to be the central question, that of how representative the surveillance populations are of the broader population (Sankoh and Byass 2012). However, a different perspective by Fottrell and colleagues (Fottrell, Byass and Berhane 2008) was to assess the sensitivity of HDSS data by introducing various permutations of random errors and omissions on a “gold standard data set” over a ten-year period. They simulated the effects in 10% or 20% of cases of randomly assigning values for sex, inflating age by 10% or 20% and in a further 10% or 20% of randomly selected cases and randomly missed 10% of deaths. Although the missing death data in the age range 5 to 59 had little effect on the age specific deaths rates, missing deaths under age one, between ages 1-4 years and ages above 60 years, exerted more influence and, widened the mortality estimates. Surprisingly, the paper concludes that tolerable margins of error can be as high as 20% before any major distortions on the population pyramid manifest, but does not mention the potential impact on mortality. This assessment found a 25% difference in the infant mortality between the methods, which is a significant error.
It is possible that the high mobility of the ACDIS population has resulted in the observed child mortality difference in this setting. There are differential exposures at play for children who for at least some part live away from their mothers and travel out of the defined HDSS, which appear to have generated higher mortality risks. This finding has important implications for South African survey data when analysed for geographic sub-populations with high migration rates. The lighter mortality for babies living in the surveillance area without their mother is probably partially due to under-ascertainment of the mothers’ residency status given the fact the data originate early on in the development of the ACDIS and defining individual resident status of all household individuals was not yet complete.

6.7 **Decline in early life mortality in a high HIV prevalence rural area: lessons from a HDSS site**

The second question pertaining to the measurement of child mortality emerges through a published article utilising the same ACDIS data set, reporting a substantial reduction in the under-2 mortality rate (U2MR) trend despite continuing high HIV prevalence in the area (Ndirangu, Newell, Tanser *et al.* 2010). The paper presents neonatal, post-neonatal and U2MR during 2000 and 2006 and aims to assess the impact of PMTCT and ART on the U2MR by analysing the temporal and spatial trends. Although the central message of the paper is that treatment had more impact on early-age mortality than prevention, the purpose of this discussion is solely to consider the interpretation of the mortality decline. In particular, to investigate the author’s claim that the decline in the neonatal and post-neonatal rates are entirely attributable to either intervention, when, in fact they are more plausibly due to methodological changes.

Significant overall declines in child mortality presented earlier in this chapter corroborate Ndirangu and colleagues reported a 36% drop in the U2MR between 2001 and 2004 compared with a further 20% decline between 2004 and 2006. Noteworthy is the disjuncture in the trend from 2003 to 2004 and the preceding period during which most of the decline occurred, and a more gradual subsequent decline. The authors designate two distinct trends as being 2001 to 2004, the era that saw the introduction of PMTCT, and 2004 to 2006, the era after the implementation of the ART programme at the end of 2004. Comparative analysis of age-specific rates calculated in this thesis revealed discrepancies with the published article, which stem from the fact they are based on slightly different measures of mortality. In fact, although the paper describes an under-2 mortality rate calculated by year of birth cohort it never specifies whether
they use a $2q_0$, or a $2m_0$. In contrast, this chapter presents period mortality unless otherwise stated.

Figure 6-11: Comparative measures of under-two mortality from the ACDIS 2000-2006 data set

Figure 6-11 presents three alternative measures of under-2 mortality relative to each other, and serves to highlight two points. First, it illustrates that period mortality and the mortality based on the birth cohort are different measures and as they include slightly different exposures, the representations of level and trend will differ. In this instance the greatest difference between the period and birth cohort measures are about 23% in 2002 and 2003, suggesting the potential deviance of these two measures warrants more clarity of the definition of the precise index.

The second measurement-related aspect highlighted by the comparison in Figure 6-11 is the differential mortality observed in resident compared to non-resident children and corroborates the earlier finding that the latter experience heavier mortality. It is interesting to note (Figure 6-3) that the surveillance population also shows that increasing proportions of live births have occurred during episodes of non-residency increasing from under 1% in 2000 to 10% in 2007 and by 15.7% in the surveillance population over the eight years. Essentially, the authors included all live births in the HDSS, regardless of whether the mother was resident or non-resident at the time of birth in the number of babies at risk in the calculation of the mortality rates. They do not appear to account for migration of children under-2 years moving out of the area, but they do not state who was included.
As mentioned earlier, an important change in the data collection procedures occurred, which probably affected the estimate of U2MR, but the article neglects to mention this. In 2004, data collection changed from thrice-yearly visits to twice yearly visits and the use of a pregnancy notification form for the collection of pregnancy details was dropped (Muhwava 2008). In practice, field workers were no longer required to administer the pregnancy notification form therefore no follow-up of the pregnancy outcome was sought at the next visit. The decrease in the number of neonatal and post-neonatal deaths coincided with these changes. Since pregnancies resulting in a neonatal death may be more likely to go unreported during the next round of data collection, than pregnancies where the baby was alive at the time of the next data collection, it is possible that the change in method could partially be responsible for the decline in death rates, particularly in the neonatal age group.

The empirical data demonstrate a clear mortality decline but the article does not acknowledge that a change in the method of measurement could have contributed to the drop in deaths between 2003 and 2004. Specifically, the decline in neonatal mortality from 18 in 2003 to an average of four per 1,000 during 2004 to 2007 seems implausible in the context of similar populations. For example, the mortality for African non-urban neonates was 22 in 1998 (Department of Health 2002) and the rate estimated from the FBH presented in Chapter 4 was 17 in 2000. Confronted with these facts, it is difficult to accept such a dramatic decline in U2MR as correct, hence in order to scrutinise the observed mortality trend by age and sex further, statistical modelling was undertaken.

6.8 **Age and sex pattern of under-5 mortality**

Mortality risk by age constitutes an extremely robust relationship for boys and girls in all populations except where female infanticide is practised. Although involving complex interactions of biological factors, there is consensus that boy babies experience slightly higher risk due to an increased susceptibility to infections (Guillot, Gerland, Pelletier et al. 2012). The age and sex pattern of mortality observed in ACDIS 2000-2007 highlighted several atypical features (Figure 6-7), namely, the overall trends for boy and girl neonates are quite different. At the beginning of the period, boy neonates experience an increase followed by a substantial mortality decline between 2003 and 2004 and is greater than the decline experienced by girls. Thereafter, a gradual levelling off up to the end of the period is experienced by both. Also apparent are sex differences in the pattern of post-neonatal deaths. For the boys, there is an increase from 2000 to 2002, then a decrease, whereas the girls display an opposite trend.
Hence, a logical line of further enquiry is to employ regression techniques to investigate the relationship between age, sex and year (time period) and under-five mortality using Poisson regression using STATA 10, generalized linear regression models (GLM).

6.8.1 Modelling the age and sex over time
Exploratory analysis led to the following categories of the variables represented in the regression Models 1 to 6 to investigate the age and sex pattern of child mortality over time: sex; age groups: <28 days, 1-11 months (reference category), 12-59 months, and Exposure Year, in person years, is the “offset” term.

Table 6-6 illustrates how the interaction terms were formed in the model building process from Model 1 to Model 6 and provides a comparison of diagnostics enabling assessment of the overall fit of each model to the data. The graphs in Figure 6-12 show the modelled effects of the covariates in Models 2-6, which are variations of the basic (age, sex, period) Model 1. For instance, Model 2 includes an interaction term of (age*year) in addition to sex.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>age, sex, year</td>
<td>sex &amp; (age*year)</td>
<td>year &amp; (age*sex)</td>
<td>age &amp; (sex *year)</td>
<td>(age<em>sex) &amp; (age</em>year)</td>
<td>(age<em>sex</em>year)</td>
</tr>
<tr>
<td>Maximum Log Likelihood</td>
<td>-384.82</td>
<td>-370.29</td>
<td>-382.74</td>
<td>-381.33</td>
<td>-368.21</td>
</tr>
<tr>
<td>Deviance</td>
<td>394.1</td>
<td>364.98</td>
<td>389.88</td>
<td>387.07</td>
<td>360.83</td>
</tr>
<tr>
<td>AIC</td>
<td>8.246</td>
<td>8.235</td>
<td>8.245</td>
<td>8.319</td>
<td>8.234</td>
</tr>
<tr>
<td>BIC</td>
<td>6.1</td>
<td>40.9</td>
<td>11.04</td>
<td>31.05</td>
<td>45.89</td>
</tr>
<tr>
<td>Residual df</td>
<td>85</td>
<td>71</td>
<td>83</td>
<td>78</td>
<td>69</td>
</tr>
</tbody>
</table>

The models generally show stark differences in mortality between boy and girl neonates and between the years 2000-2003 and 2004-2007. Modelled effects demonstrate the mortality decline between 2003 and 2004 is substantially more pronounced for the neonates than for the other age groups and more so for boy neonates than girl neonates. The effects are minimal for the other age groups, the 1-11 month and 12-59 month old age groups are more or less stable for both boys and girls throughout the period.

The coefficients, relative risks, lower and upper 95% CI for Model 6 are shown in the table in Appendix 6. Assessment of the (age*sex) interaction term was carried out by
calculating 95% CI’s for the coefficients in 2003 and 2004 by age group (Models 3 and 5), however, none of the coefficients fell within the 95% CI (not shown).

The BIC statistic in Table 6-6 indicates additional information between Models 2 and 5 has no significant effect. In addition, the BIC and other diagnostics suggest that Models 5 and 6 provide the best fit to these data. The models examine \((age*sex)\) and \((age*year)\) relationships and demonstrate the exaggerated neonatal decline, suggesting these effects are not homogenous across age groups and point to a peculiarity in neonatal mortality of boys especially in the 2003-2004 period.

Although impossible to conclude that the decline in neonatal deaths between 2004 and 2007 is real or that it is attributable to a change in data collection, the latter explanation is more probable given the fact that the decline coincided with methodological changes and also because the modelling indicates such a decline produces an odd neonatal pattern for boys. Aside from that, the systematic under-reporting of neonatal deaths has resulted in an extremely improbable mortality rate, in the same period that other data sources point to a stable neonatal trend for the country as a whole.
Figure 6-12: Graphic presentation of age, sex, and year modelled effects over time, Models 2-6
6.9 Discussion

The objectives of this chapter were to measure and describe the trend in child mortality derived from longitudinal data and to assess the performance of the demographic surveillance in the measurement of childhood mortality. The analysis uncovered several relevant topics, which warranted further investigation and became focus points of the chapter. These were the comparison of the prospective and retrospective approaches and three issues emanating from the Ndirangu et al. (2010) article, namely the differences produced by different measures of mortality, the differential mortality of resident versus non-resident children and a misinterpretation of the impressive decline in ACDIS early age mortality.

The apparent 25% difference in under-five mortality estimates produced by the retrospective and longitudinal approaches is a major concern and poses the question of what are acceptable margins of error within surveillance populations? The literature is scant on the subject but there seems to be a tendency to classify estimates within 10% of each other as being similar, whereas estimates corresponding to around 20% difference as less representative (Byass, Sankoh, Tollman et al. 2011; Byass, Worku, Emmelin et al. 2007) although no strict criteria are evident. The simulation exercise created up to 20% of random error in some cases, which turns out to have surprisingly little effect on the population structure and highlights a general strength of HDSS data sets as being robust due to large sample sizes. Importantly, the infant mortality rate was most sensitive to the introduction of errors (child mortality less so), illustrating more vulnerability in the measurement of the IMR.

The analysis showed that the disjuncture in mortality level is the result of non-resident infants living away from their mothers and is better represented in the retrospective data, whereas resident infants are better represented in the surveillance data. The cause of discrepancy lies in coverage differences rather than in the alternative methods of measurement, and these manifest in the calculation of rates, and is probably due to inadequate ascertainment of individual residential status at the commencement of the DSA. This type of error is categorized as a non-measurement bias emanating from the respondent (respondent-error) arising out of a misunderstanding (Fottrell, Byass and Berhane 2008).

From a research perspective, the mortality differential between the two groups found infants living away from their mothers experienced twice the risk of dying compared with other children and typifies the appealing feature of the indicator, displaying sensitivity to a spectrum of living conditions that duly reflect variation in
individual risk of child mortality. Perhaps herein lays the essence of why child mortality is not ideally measured by HDSS in the South African setting, especially in areas of the country where circular migration is prevalent.

Data from DSSs can be informative about demographic trends, if not level, of estimates where the experience is that they have been highly consistent with national trends in the past. However, an elementary difference of South African HDSS’s from other sites elsewhere is that they represent select rural populations in geographically diverse parts of the country, known to experience quite different levels of childhood mortality. This analysis of factors related to childhood mortality within ACDIS demonstrates intra-rural variation even within high levels of socioeconomic homogeneity. On the other hand, a principal objective of HDSS elsewhere in the world is how representative the surveillance populations are of the broader population, how generalizable the demographic and epidemiological representations of the surveillance community are of the regional or national populations. In contrast, the primary utility of the surveillance systems in South Africa is to gain demographic information specifically for the rural populations known to be under-reported in the national civil registration system.

The review of the article by Ndirangu and colleagues (2010) was important for two reasons. First, it demonstrated the subtle but important differences between the measures of a \( z \bar{q}_0 \), a \( \bar{m}_0 \) or an U2MR produced from a birth cohort. Although the definition is slightly different, each measure will reflect a different dimension of mortality that produces a slightly different level and as this has implications for the interpretation of results, recommended ‘good practise’ for the analyst is to be clearer about the measure. This comparison also made clear the mortality differential of resident compared with non-resident children and makes the case for distinguishing these groups, as the latter are highly mobile and introduce various biases, which cannot be accounted for.

The primary conclusion of the article, resulting from the empirical evidence, is a clear decline in under-2 mortality but the article does not acknowledge that a change in the method of measurement in addition to the abandonment of the pregnancy notification form could have contributed to the decrease in deaths between 2003 and 2004. We argue that methodological changes are responsible for the consistent under-reporting of neonatal deaths throughout the period, which falsely exaggerate the overall mortality decline in children under-2. Accepting such a dramatic decline in the U2MR as
correct and as the result of a single intervention was not substantiated by the data, and disappointing given the need at the time for information for formulating the PMTCT policy.
Civil registration and vital statistics are the cornerstone of demographic and health sector information. A complete vital registration system serves to record all births and deaths. These records enable the collection of vital statistics, which allows child mortality rates to be calculated directly. However, in South Africa, not all events are registered and consequently both the numerator and denominator suffer from under-reporting. Evaluation of the vital statistics information needed to monitor childhood mortality requires the assessment of completeness of both the birth and the death registration.

This chapter investigates vital registration data to assess the quality of birth, death and cause of death data related to the population under five years of age drawing on an evaluation framework used to assess the quality of South Africa’s death registration for the period 1997 to 2007 (Joubert, Rao, Bradshaw et al. 2013). The overall objective of this chapter is to assess the vital registration system as a tool for monitoring childhood mortality by evaluating the following criteria:

- the completeness of birth registration
- the completeness of death registration
- the coherence of the age pattern of childhood deaths over time.

### 7.1 Evaluation of the completeness of birth and child death registration

Essentially, “completeness is a measure of the extent to which births and deaths that occur in a country in a given year are registered by the civil registration system” (WHO 2010: 34). The literature indicates that the methods for evaluating the completeness of adult deaths are the most developed. Relative to the era when methods for evaluating the completeness of adult deaths were proposed, it is surprising that it is only recently methods of assessment or adjustment to correct defective death registration data pertaining to children have been published. In both the cases of births and of deaths, completeness can be assessed using independent estimates measured in the population using either direct or indirect questions asked in sample surveys and censuses (Hill, Lopez, Shibuya et al. 2007).

The evaluation of completeness of the birth registration data in this thesis adopts two approaches. First, the reported number of births is compared against model-based estimates and second they are compared against an estimate of births derived from
independent data sources. The evaluation of the completeness of death registration utilises published estimates (Pillay-Van Wyk, Laubscher, Msemburi et al. 2014), compared with a trend derived from previously published indirect estimates (Darikwa 2009) and estimates based on the 2011 Census data.

7.2 Civil registration and vital statistics in South Africa
The registration of births and deaths is governed by the South African Birth and Deaths Registration Act of 199220 (Act No. 51 of 1992) and is administered by the Department of Home Affairs (DHA). Although the Act has been amended from time to time, probably the most significant modification was the introduction of new birth and death notification forms in 1998. The revision of the forms aimed to secure data for the public health information needs of the newly established health information system and to some extent ensure future public health information needs. The changes were also intended to bring South African data on a par with international standards, such as recording details of the death on the new death notification form proposed by the International Classification of Diseases and Deaths (Bradshaw, Kielkowski and Sitas 1998). An important initiative to improve the timely registration of births was the highlighting of the registration process at antenatal care visits where mothers were encouraged to obtain identification documents if they did not already have them and by providing the necessary birth notification forms for newly born babies in state facilities at the time of delivery (Bradshaw, Kielkowski and Sitas 1998).

Unlike countries such as India and the USA where federal policies around civil registration may differ from one state to the next, a national legal framework pertains to birth and death registration in South Africa and applies to all nine provinces.

The Statistics Act (1999)21 governs the national statistics office, Statistics South Africa (Stats SA), as the provider of official statistics for the country. Stats SA currently publishes an annual statistical release reporting on live births (Stats SA 2013c) and as annual statistical release reporting on causes of death from the death notifications for example the report for 2013 (Stats SA 2013b).

7.3 Birth registration

The terms and definitions used in this chapter are presented in Table 7-1.

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
<th>Comment</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Late registration as defined on the Birth Notification Form (BNF)</td>
<td>A birth registered after the thirtieth day of birth</td>
<td>Home Affairs definition as stated on the BNF</td>
<td>Population register DHA</td>
</tr>
<tr>
<td>Current birth registration, Stats SA</td>
<td>Refers to births registered in the calendar year of birth or up to the end of February of the year following the calendar year of birth</td>
<td>The definition of a late birth registration used in the annual live birth reports processed by Stats SA and differs from the Home Affairs definition</td>
<td>Stats SA</td>
</tr>
<tr>
<td>Late birth registration, Stats SA</td>
<td>Refers to births registered after February in the year following the calendar year of birth</td>
<td>This definition is used in the Stats SA annual live birth reports processed by Stats SA</td>
<td>Stats SA</td>
</tr>
<tr>
<td>Year of registration for the late registrations, Stats SA</td>
<td>Births are considered to have registered in a calendar year if they were registered between 1 March of that year and the end of February of the following year</td>
<td>A 12-month period</td>
<td>Stats SA</td>
</tr>
<tr>
<td>Registered in the year of birth*</td>
<td>Refers to births which occurred in a calendar year registered in the 14-month period including the calendar year of birth up to the end of February of the following year</td>
<td>This term is used in this thesis</td>
<td></td>
</tr>
<tr>
<td>Registered in the year following the calendar year of birth*</td>
<td>Births registered in the 12-month period from 1 March of the calendar year following the calendar year of birth</td>
<td>This term is used in this thesis</td>
<td></td>
</tr>
<tr>
<td>Registered in the year of birth +1*</td>
<td>Cumulative births registered in the 26 months ending 12 months after the end of February in the year after the calendar year of birth</td>
<td>This term is used in this thesis</td>
<td></td>
</tr>
<tr>
<td>Registered in the year of birth +2, +3 etc. *</td>
<td>Cumulative births registered in the year of birth plus the 2nd, 3rd year etc. following the calendar year of birth (up to the end of February of the following year)</td>
<td>This term is used in this thesis</td>
<td></td>
</tr>
</tbody>
</table>

*Although analysis in this thesis does not extend to 2013, it should be noted that the cut-off ‘registered in year of birth’ changed to the end of calendar year instead of the end of February of the following year.

7.3.1 Reporting of birth statistics

Home Affairs administers three variants of the ‘notice of birth’ form DHA-B24 (Appendix 7), depending on the age of the person being registered. Registration occurring within 30 days of birth requires form B1-24; registration occurring after the
child’s first birthday, but before the fifteenth birthday requires a B1-24/1 form and children and adults registered after the age of fifteen require a B1-24/15 form.

The birth and death registration Act (Act No.51 of 1992) makes it compulsory to register a birth within 30 days. Registrations taking place after 30 days are considered late registrations by DHA. The birth notification form comprises two sections. Part 1 collects demographic details of the child and parents. Part 2 contains several questions related to socio-economic status and educational attainment of the parents, as well as information related to new-born and maternal health such as birth weight, Apgar\(^{22}\) score, any abnormalities as well as risk factors during pregnancy amongst others. Data contained in Part 2 are not captured by either DHA or Stats SA (Dobbie, Masebe and Nhlapo 2007).

DHA captures the information on Part 1 of the notice of birth form onto the National Population Register. Data for these events are transferred electronically into the vital registration database maintained by Stats SA on a monthly basis. This administrative division between Home Affairs and Stats SA marks the separation of civil registration and vital statistics.

Stats SA publishes an annual report (P0305) on live births recorded in the previous year, together with updates of the number of births registered late for the previous years. The major shortcoming of the birth registration data is that a significant proportion of the births are registered after the calendar year of birth, despite the stipulation that the registration should take place within one month of birth. Thus, the database of births is continuously being updated with late registrations from earlier years\(^{23}\). This means the total number of births registered as occurring in any given year becomes asymptotically more complete with each passing year.

The most recent report considered in this research (Stats SA 2013c), for example, reports on the registration of births occurring up until the end of 2012 including all late registrations captured up until the end of February 2013 and was released in November 2013 resulting in an eleven month time lag, however, it must be acknowledged that the lag varies.

---

\(^{22}\) The Apgar score devised in 1952 to assess the health of a newborn child quickly. The five criteria of the Apgar score are Appearance, Pulse rate, Grimace, Activity and Respiration of the newborn.

\(^{23}\) The annual live birth reports published by Stats SA capture registrations up until the end of February of the following years’ births. The reports include the births registered in the calendar year to which the report pertains, excluding any late registrations. For example, the 2011 report does not include any late registrations recorded after February 2012 in the 2011 totals. Late registration in 2011 will only be added to the 2011 births totals in the 2012 annual report.
7.3.2 **Trends in registered births**

Examining the trend in birth registration over the last 22 years\(^a\), Figure 7-1 shows the number of registered births as of February 2013. It also shows the trend in the number of births registered in the year of birth, those in the 12 months after the end of February following the calendar year of birth and the combined number of these registered births (year +1). It can be seen that the number of births registered by year of birth increased from 946,755 in 1993 to peak at around 1,080,000 between 2006 and 2008. The slight but continuous decline thereafter is the result of births missing due to late registrations. Notable is the significant increase in birth registration after 1992, as the new legislation in the South African Birth and Deaths Registration Act of 1992 included the previously so-called independent homelands accompanied by efforts to strengthen vital registration.

Evident also is the marked improvement in the proportion of current registrations since 1998. In fact, the figure illustrates a clear shift in the timeliness of registration coinciding with the introduction of the new birth notification form in 1998. Prior to 1998, more registrations occurred in the year following the year of birth. However, since 1998 this trend has reversed, showing continuous improvement in the proportions of births registered in the year of birth and since 2000 declining proportions of births registered in the year following the year of birth.

7.3.3 Evaluation of the completeness of birth registration based on projection models

The registered births are compared against three model-based estimates, namely: the mid-year population births produced using the Spectrum model25 (version 4.39) adapted to produce the official mid-year population estimates for 2011 (Stats SA 2012c); World Population Prospects (2015 Revision) produced by UN Population Division26 and the ASSA2008 model27. While the estimates from the ASSA model predate the 2011 Census, the other two sets of estimates were produced after the release of the Census. Total registrations up to and including 2012 (captured to the end of February 2013) by year of birth are shown in Figure 7-2 with the modelled estimates from 2000 to 2012.

The modelled estimates of births decline steadily over this period, whereas the number of registered births by year of birth increased from 2003 to a plateau from 2006 to 2008 and then declines. While late registrations might explain the observed difference between the models and the registered births in the most recent years, and the considerable difference for the earlier period could be due, in part at least to low levels of registration, the comparison does call into question the assumption in the models that births declined steadily over the period.

---

25 http://www.avenirhealth.org/software-spectrum
26 https://esa.un.org/unpd/wpp/DataQuery/
27 http://www.actuariesociety.org.za/
Since the VR system updates annually with late registrations and the bulk of the registrations occur with the period up to a year after the year of birth, a useful indicator of the trend in completeness over time is the completeness of births registered up to a year following the year of birth. Estimates of completeness of current registration as measured against the three models are shown in Table 7-2. All three models suggest a similar range of completeness over time, improving from about 37% in 2001 to about 86% in 2012.

**Figure 7-2: Comparison of births registered by year of occurrence with model based projections of annual births, 2000-2012**
All births (dotted line) refers to births registered up to the end of February 2013 by year of birth.

**Table 7-2: Estimated completeness of registered births within the year of birth plus the following year compared against various other estimates, 2001-2012**

<table>
<thead>
<tr>
<th>Year</th>
<th>UN estimates</th>
<th>ASSA estimates</th>
<th>Stats SA Mid-year estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>37</td>
<td>36</td>
<td>41</td>
</tr>
<tr>
<td>2002</td>
<td>43</td>
<td>42</td>
<td>48</td>
</tr>
<tr>
<td>2003</td>
<td>50</td>
<td>49</td>
<td>48</td>
</tr>
<tr>
<td>2004</td>
<td>56</td>
<td>55</td>
<td>53</td>
</tr>
<tr>
<td>2005</td>
<td>66</td>
<td>66</td>
<td>63</td>
</tr>
<tr>
<td>2006</td>
<td>72</td>
<td>73</td>
<td>69</td>
</tr>
<tr>
<td>2007</td>
<td>79</td>
<td>80</td>
<td>76</td>
</tr>
<tr>
<td>2008</td>
<td>79</td>
<td>80</td>
<td>77</td>
</tr>
<tr>
<td>2009</td>
<td>85</td>
<td>86</td>
<td>83</td>
</tr>
<tr>
<td>2010</td>
<td>83</td>
<td>83</td>
<td>81</td>
</tr>
<tr>
<td>2011</td>
<td>84</td>
<td>84</td>
<td>83</td>
</tr>
<tr>
<td>2012</td>
<td>86</td>
<td>86</td>
<td>86</td>
</tr>
</tbody>
</table>
7.3.4 Evaluation of completeness of birth registration based on independent empirical estimates

The availability of an empirical estimate of the number of births against which the registered births can be compared presents challenges due to the paucity of data on recent births. The next sub-section appraises potential sources of data by considering their suitability or otherwise for the purpose of estimating the total number of live births nationally. Hill and colleagues (2007) offer broad guidelines for assessing the appropriateness of various data sources for the estimation of births and child mortality. In addition to two of the sources suggested by Hill, Lopez and Shibuya et al. (2007), census and survey data, a third data source was utilised in this analysis, namely the routinely collected record of the number of births occurring in public sector health facilities captured by the District Health Information System28 (DHIS).

7.3.4.1 Empirical estimates of births

The 2011 Census, offers the possibility of deriving estimates of the numbers of births by reverse survival whereby the locally-born census population in each year of age is projected backwards to the corresponding year of birth (see methods 3.11.2.229).

A second source of census data are indirect data pertaining to current and lifetime fertility used in the P/F ratio method to calculate the age-specific fertility rates from which the number of births is derived. Published estimates of the age-specific-fertility rates from the 2011 Census (Stats SA 2015) forthcoming, and reworked estimates of TFR (Dorrington and Moultrie 2015) from those previously published from the 2001 and 1996 Censuses (Moultrie and Dorrington 2004) and the 2007 Community Survey (Darikwa 2009) were utilised.

The third source is DHS data that directly calculates cohort-period fertility rates to produce ASFR’s from full birth history information. To date, only the 1998 survey, has produced reliable fertility estimates (Department of Health 2007).

The fourth source of data are the routinely collected data on births occurring in public health facilities. As the DHIS was designed for purposes of health management at district level, the system only includes state facilities and therefore excludes the private sector. Hence, the DHIS birth data will underestimate the true number of births. In addition, it is important to note that coverage of the DHIS was incomplete before 2004, as the data collection system was still being phased in before that. The DHIS

28 http://www.hrhrresourcecenter.org/node/1070
29 As mentioned previously census data are after adjustment for undercount as estimated by the post enumeration survey.
births would therefore require adjustment for the proportion of births occurring in the private sector and occurring at home to be representative nationally.

7.3.4.2 Adjustment of DHIS data

Producing an estimate of total births from the births estimated in surveys occurring in private facilities and home-births with public sector births is not straight forward due to the fact that the proportion of births occurring in the public sector has increased since the introduction of the free health care (Gilson and McIntyre 2007), rendering the earlier survey data outdated.

Three household surveys asked questions about the ‘place of delivery’ and hence provide a distribution of home-births and private facility-births. Importantly the national surveys provide provincial distributions, therefore, the 1998, and 2003 SADHS’s and the 2012 HSRC survey were explored to obtain proportions of home-births and private facility-births by province.

Before discussing the implications of these findings, a few caveats relating to the quality of the data are important to note. There is consensus among researchers that the 1998 SADHS is of better quality than the 2003 survey. In particular, Chapter 4 showed that the full birth history in the latter survey was particularly wanting. Specifically the under-reporting of recent births in the last five years resulted in implausibly low fertility for most provinces, particularly so in some provinces, and the country as a whole. The proportion of data missing by province also attests to the poor overall quality of data.

<table>
<thead>
<tr>
<th>Table 7-3: Provincial distribution of “place of delivery”, SADHS 1998, 2003 and HSRC 2012 household surveys</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998 SADHS</td>
</tr>
<tr>
<td>Public</td>
</tr>
<tr>
<td>Private</td>
</tr>
<tr>
<td>Home</td>
</tr>
<tr>
<td>Missing</td>
</tr>
<tr>
<td>2003 SADHS</td>
</tr>
<tr>
<td>Public</td>
</tr>
<tr>
<td>Private</td>
</tr>
<tr>
<td>Home</td>
</tr>
<tr>
<td>Missing</td>
</tr>
<tr>
<td>2012 HSRC survey</td>
</tr>
<tr>
<td>Public</td>
</tr>
<tr>
<td>Private</td>
</tr>
<tr>
<td>Home</td>
</tr>
<tr>
<td>Missing</td>
</tr>
</tbody>
</table>
Aside from these considerations, there are several interesting trends gleaned from the household survey data, which spans twelve years and are presented in Table 7-3. The first point is the proportion of births not in the public sector more than halved between 1998 and 2010. The second point is that the richer, more urban provinces have proportionally more births in private facilities (Gauteng and the Western Cape) than the poorer, more rural provinces, which account for a greater proportion of home-births (the Eastern Cape, Mpumalanga and Limpopo). The third point is that although the decrease in home-births is accompanied by an increase in public facility-births, the proportion of missing data in the 2003 SADHS limits the usefulness of the relative proportions of each by province and necessitates a review of other potential data sources that might be used as proxy measures of home-births and private facility-births.

An alternative is the use of estimates of the proportion of children under-1 year of age covered by medical aid schemes or health insurance as reported by the General Household Survey (GHS), which can be taken to be a proxy for the proportion of births that occurred in the private sector. The average over the nine-year period 2004 to 2012 was 11.9%. To estimate the proportion of home births one can make use of the already mentioned (section 3.11.2.3) association between the proportion of each province living in rural areas and the proportion of home-births. The correlation between provincial proportion rural from the 1996 Census and the provincial proportion of home-births from the SADHS 1998 (Table 7-3) result in an $R^2 = 0.74$ indicating a strong enough correlation to predict reasonably the relationship of these proportions over time from the changes in the proportion rural. The regression equation models a relationship defining the proportion of home-births by province to that of the national proportion by using the ratio of the provincial proportions designated rural to the national proportion rural. The resulting relationship and an estimate of home-births nationally in 2012 can predict the proportions of home-births by province. The method is described in more detail in the methods chapter (section 3.11.2.3).
Validation of a proposed correction (accounting for private-facility and home-births) involves comparing the number of births implied by the adjusted total births with other estimates of national births. The five point estimates of the number of births from the fertility rates as reported by women shown in Figure 7-3 suggest a fairly level trend in number of births ranging from 1,130,571 in 1996 to 1,124,000 in 2011 indicating a difference of only 3.2% over the seventeen-year period. The estimate of the number of births from the reverse survival of the population count (hollow circles) is slightly higher than the reverse survival estimate from the count of the South African-born (solid triangles). The estimates of the number of births produced from reverse survival of population numbers and the adjusted DHIS (increased by 16.3% to allow for births not captured by the system) are slightly higher than both the estimates from the 2007 Community Survey and 2011 Census based on ASFR’s. However, from 2006 the reverse survival of the South African-born population and the adjusted DHIS are practically identical.

The next sub-section turns to the comparison of completeness implied by each alternative to correct for the non-public sector proportion of births.

7.3.4.3 Comparison of empirical estimates of completeness of reporting of VR births
Any estimate of the completeness of birth registration is entirely dependent on the estimate of births it is measured against. Evident from the comparison of the number of births (Figure 7-3) is that the estimates result in a 3% difference between the Census
1996 (the highest) and the Community Survey 2007 (the lowest) number of annual births, which does have bearing on the choice of independent estimate.

The TFR is based on the reports of women of reproductive age and quantifies the fertility experience for those women if fertility remained constant. Features of these data are that the reporting becomes less accurate with the age of the respondent, and that under-reporting of recent births also affects these data. Apart from these two typical errors, the question is asked of all women of reproductive age in the household; however, the data can be disaggregated by place of birth in instances where analysis of South African fertility is desired as in this case.

Reverse survival, on the other hand, does not require any fertility question, but is based on the count of the number of people by age (and place of birth). The accuracy of these data is therefore heavily dependent on the accuracy of the population count by age (i.e. both the count and accuracy of reporting of age) and to a lesser extent on the accuracy of the survival factors used to project the numbers back to the year of birth. The most serious constraint of these data in general is the well-established fact that children and particularly infants tend to be under enumerated in censuses (Moultrie, Dorrington, Hill *et al.* 2013: 82). It is important to note, however, that the Census count is supported by school registration data, DHIS data and population registration on the National Population Register (Dorrington and Moultrie 2015).

![Figure 7-4: Comparison of births registered in the calendar year of birth and the following year of birth (2000-2012) with births occurring in health facilities and births derived from Census 2011](image-url)
Figure 7-4 compares births derived from the 2011 Census, reverse survival (RS), and the corrected DHIS estimate of live births with births registered in the year of birth and the number of births registered in the year of birth plus the following year. It illustrates that after adjustment the DHIS data produce very similar numbers of births as the Census, and the extent of additional registrations in the year following the calendar year of birth.

Next, to assess the completeness of birth registration in the year of birth and in the year of birth +1, these numbers of births are compared with the number of births estimated from the Census population and the adjusted DHIS (Table 7-4) and provides a range of completeness.

<table>
<thead>
<tr>
<th>Year</th>
<th>Registered in the year of birth</th>
<th>Registered in the year of birth +1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Adjusted DHIS</td>
<td>Census</td>
</tr>
<tr>
<td>2001</td>
<td>-</td>
<td>47</td>
</tr>
<tr>
<td>2002</td>
<td>-</td>
<td>57</td>
</tr>
<tr>
<td>2003</td>
<td>-</td>
<td>62</td>
</tr>
<tr>
<td>2004</td>
<td>73</td>
<td>69</td>
</tr>
<tr>
<td>2005</td>
<td>75</td>
<td>70</td>
</tr>
<tr>
<td>2006</td>
<td>76</td>
<td>75</td>
</tr>
<tr>
<td>2007</td>
<td>75</td>
<td>71</td>
</tr>
<tr>
<td>2008</td>
<td>77</td>
<td>76</td>
</tr>
<tr>
<td>2009</td>
<td>75</td>
<td>73</td>
</tr>
<tr>
<td>2010</td>
<td>78</td>
<td>75</td>
</tr>
<tr>
<td>2011</td>
<td>78</td>
<td>77</td>
</tr>
<tr>
<td>2012</td>
<td>79</td>
<td>-</td>
</tr>
</tbody>
</table>

Since the DHIS is incomplete before 2004, assessment of completeness of registrations based on the adjusted DHIS has to be restricted to the period since 2004. Estimates derived from both sources are very close, suggesting that completeness rises from 73% to 79% in the year of birth and from 82% to 87% in the year of birth +1 (Table 7-4).

Extending this comparison to late registrations at higher ages allows us to examine the pattern of completeness by age over time. Figure 7-5 illustrates the improvement in registration from the first to the fifth birthday. Providing the Census estimate of the total number of births is reasonably accurate, this suggests that in 1995, about 25% of registrations took place before the first birthday and 50% took place by the fifth birthday. However, by 2008, these proportions had improved to 76% and 89%, respectively, and in 2011, 77% of children under age one and 83% of children under age
two were registered. It is also evident that most improvement up to 2004 occurred under age one.

![Figure 7-5: Trends in the completeness of birth registration by age measured against an estimate of births derived from 2011 census, 1995-2012](image)

7.3.5 The need for common reporting standards
The evaluation of the completeness of birth registration has shown great improvement since 2000, across all the provinces. The sub-section also unearthed the effects of employing different standards against which completeness is measured. A model-based approach is only valuable if the fertility assumptions reproduce an accurate projection of births. This was demonstrated by reviewing the Stats SA estimate derived using the Spectrum model, the UN World Population Prospects model and the ASSA model, which all show a declining births over time when in fact registered births show a changing trend - one that declined to the early 2000s then rose to peak around 2008 declining since then.

In addition to the degree of uncertainty surrounding these estimates, what is apparent from the literature evaluating the international programme to strengthen and modernise civil registration and vital statistics is an underlying lack of clarity of key terms, such as those defining the proportion of births registered by age including the reporting practices within government departments. This is illustrated by the claim that 90% of births were registered within the year of birth in 2011/2012 (Department of Home Affairs 2012), referring to the financial year. Although the report does not explain how this figure was derived, it also states that 54% were registered in the first thirty days after the birth. The former statistic is subsequently quoted by the department
of Women Children and Persons with Disabilities (2012) in their submission to the UN Convention on the Rights of the Child, and in a review of the law associated with realising children’s rights. Berry and colleagues (2014), who also rely on reports from the 2012 DHA report, state that birth registration improved from 68% to 85% between 2003 and 2008, and of these 87% were registered in the year of birth and the remaining 13% were registered over the period 2009 to 2012. By way of comparison, the analysis based on the Census estimates that completeness in the year of birth improved from 63% to 78% over the same period.

UNICEF’s Every Child’s Birth Right (UNICEF 2013) relies on Stats SA annual birth registration reports and claims that South Africa achieved 95% completeness of births registered within the first year of life. It appears that UNICEF calculated the proportion of births registered up to the mid-year of a particular year, registered in the ‘year of birth’ (or the current births, which includes all births registered up to 28 February of the year following the year of birth). Thus for the births in 2011 the 95% represents the proportion of births registered up to the middle of 2012 which were registered by 28 February 2011. As with the DHA, the problem with referring to this as an estimate of completeness is that the numerator continuously increases due to late registrations. In contrast, the comparison of completeness of registration from this chapter for 2012 up to the end of February 2012 divided by an estimate of the number of births in 2012 is only 78% and as Figure 7-5 shows has remained at about that level since 2006.

7.3.6 Discussion
Valid estimation and monitoring of completeness in future will require information about the non-public sector proportion of births. In this regard, the birth surveillance provided by the DHIS has proved an important source of birth information ideal for monitoring births during the inter-censal period and therefore worthy of investing in the resources necessary for improving the quality of routine statistics. The proportions of private sector and home births required for monitoring provincial births are best estimated from household surveys such as the DHS, but the quality of these data is also a key factor related to the accuracy of this indicator.
7.4 **Death registration**

The latest amendment to the Births and Deaths registration Act of 1992 applicable to the registration of deaths replaced the previous Death Notification Form (DNF) BI-1663 with a longer form, DHA-1663, in 2009 (Stats SA 2013b), which is shown in Appendix 8. The new form is three pages instead of one requires more information, but similar to the previous form, the last page contains demographic details and the causes of death information and is: “Confirmation for Medical and Health use Only (After completion seal to ensure confidentiality)”. In the case of minors, the details of the deceased child and the mother are included in this section.

The DHA capture personal and basic demographic information of the deceased, on the Population Register, including whether the death was natural or non-natural. A medical physician in the case of a natural cause of death carries out medical certification of causes of death included on the DNF. However, in rural areas where often no medical physician is available, a headman” completes a DHA-1680, also referred to as a death report. In the case of a death due to an external cause of death, certification is carried out at a mortuary because an inquest is required.

In cases where an individual dies before the birth is registered, the details of the death are not captured on the register because issuing the identification number is dependent on the notification of the birth. In such instances, both events of birth and death by-pass the population register entirely.

Since 2000, the MRC has developed a database that is updated approximately every four months with these basic details, to provide up-to-date age and sex mortality profiles known as the Rapid Mortality Surveillance System. These data have been valuable for monitoring trends in adult mortality (Dorrington, Bourne, Bradshaw *et al.* 2001), during a period when Stats SA were unable to provide current information due to delays in the processing of cause of death statistics. Since 2009, the RMS data have been used to produce key childhood and adult mortality indicators for monitoring purposes and are used by the Health Data Advisory and Coordination Committee to monitor progress on key mortality indicators (Dorrington, Bradshaw, Laubscher *et al.* 2014). The DHA data therefore represent a precursor subset of the vital registration data which particularly under-represent deaths of children due to the interdependence of birth and death notification process.

---

30 A headman is authorised to act and serve the local population in the role of a community leader in many rural parts of South Africa.
7.4.1 Reporting of death statistics

Although the DHA and Stats SA are governed by the same legislative Act regarding the registration of births and deaths, the functional differences between the civil registration system and that of vital statistics require different reporting practices. In the case of civil registration, the DHA issue a burial order and an abridged death certificate to the deceased’s family on receipt of a death notification form. Entry of the deceased’s details on to the National Population Register then occurs on condition that a record of the individuals’ birth is also on the NPR. A key feature of these data therefore is that most of the bias due to under-reporting affects babies who died in the early hours, days, weeks and months of life (Dorrington, Bradshaw, Laubscher et al. 2014).

In the case of processing vital statistics, late registrations “refer to deaths that were registered later than the year in which they occurred” (Stats SA 2013b: 12). Although Stats SA aims to publish an annual statistical release on mortality and causes of death (PO309.3) registered in the previous calendar year, the time taken to produce the report varies, but up to 2010 has been approximately 24 months. The release of the 2010 annual report was published 1 year and 3 months from the end of 2010; and, the turnaround time to publication for the 2013 data was reduced to 11 months. As with the birth register, the database of deaths is updated with late registrations, but the timeliness differs from that of births. For instance, the 2010 data indicate that 90% of all registered deaths were registered within the week of death.

Part one of the DNF documents captures details of the child death such as age at death, province of death and information related to the person who registered the death. Part 2 provides cause of death information consisting of three lines for the “immediate” cause of death, the “consequential” causes of death leading up to the immediate cause and the “underlying” cause of death. Part 2 also contains information about the mother, documenting her previous pregnancy outcome, antenatal treatment, and further important details of the deceased child such as birth-weight.

Once entered on the Population Register, the DHA sends the forms to be recaptured by Stats SA where nosologists classify the immediate cause of death information after which it is allocated to an underlying cause of death using an Automated Classification of Medical Entities (ACME) programme. The cause of death information is coded according to the International Classification of Diseases revision 10 (ICD10)31 to single causes of death (4 digit codes).

31 http://apps.who.int/classifications/icd10/browse/2015/en
7.4.2 Trends in registered deaths
The nineteen-year trend in reported deaths under age five years (Figure 7-6) show the total number of registered deaths rose from under 23,000 in 1993, peaked in 2006 at 64,380 and declined to 47,435 by 2010. Deaths in the post neonatal age group account for the majority of total number of registered deaths under the age of five throughout the period. The number of post neonatal and child deaths increased from 1993 up until 2006, and declined thereafter, while a smaller increase in the early neonatal trend up to 1998 is observed, followed by a dip in 2001 and increase to 2005. There is little change in the late neonatal age group, which contributes the least proportionally. Unfortunately, since mortality rates are expected to have increased to a peak and then declined over this period due to the impact of HIV on mortality and the later impact of ART on HIV mortality, it is difficult to estimate the completeness of reporting from these trends over time (Figure 7-6).

![Figure 7-6: Trend in the reported number of deaths by age, Stats SA 1990-2011](image)

7.4.3 Evaluation of completeness of death registration based on model projections
Registered deaths under age five years are evaluated against the mortality estimates produced by three model-based estimates shown in Figure 7-7, namely, the Interagency Group for Child Mortality Estimation (IGME), the Actuarial Society of South Africa (ASSA) 2008 Aids and demographic projection model and the Institute for Health

---

32 http://www.childmortality.org/files_v17/download
33 http://www.actuarialsociety.org.za/Societyactivities/CommitteeActivities/DemographyEpidemiologyCommittee/Models.aspx
Metrics and Evaluation (IHME). These are plotted against two empirical mortality trends produced using VR under-5 deaths and the births registered in the calendar year of birth +1 (solid line) and the VR under-5 deaths and births estimated by reverse projecting the census count (dashed black line).

![Mortality Trends Chart](http://www.ghdx.healthdata.org/country_profiles)

Figure 7-7: Comparison of the number of deaths registered by year of death with model based projections of annual deaths, 1990-2013

The estimates from the three models and the registered deaths all show a similar trend, peaking between 2003 and 2005, although there are differences in level. The model-based trends display significant mortality declines since 2005. The two VR trends show a 20 per cent difference in the level of mortality in 2010, which is accounted for by the under-reporting of births. The more correct level (VR/VR+1 births solid line) is because the under-registration of births and deaths are assumed to be of a similar order of magnitude, whereas the lower level (VR/Census births dashed line) is because the births are closer to correct but the deaths are under-reported.

7.4.4 Evaluation of completeness of death registration based on independent empirical estimates

Unlike the births and despite a debilitating lack of independent childhood mortality estimates there have been efforts to estimate the completeness of the registration of deaths in earlier years. The assessment by Darikwa and Dorrington (2011) of the under-reporting of civil registration deaths against the mortality derived from the 2007

---

34 [http://www.ghdx.healthdata.org/country_profiles](http://www.ghdx.healthdata.org/country_profiles)
Community Survey suggests completeness of death registration increased substantially since the end of the 1990’s and by 2006 was approximately 88%, 62% and 75% for infant, child and under-five year old age groups respectively.

The only other empirical sources of available data since the Community Survey are the 2011 Census questions related to the SBH, deaths in the household and the previous birth technique. Unfortunately, assessment of the SBH do not produce a reliable mortality trend. While the data on births in the preceding 24 months question were analysed, the estimate yielded an unrealistically low IMR of 26.6 around the end of 2010 and therefore cannot be used to assess completeness. The completeness of registered deaths is therefore assessed by way of comparison with the mortality produced from the deaths in the household described in (section 5.4).

These are the most recent empirical estimates of infant and under-five mortality and are presented along with other available national estimates since 1996 (Figure 7-8) in order to assess the level and trend in death registration.

Interestingly, the sources of all the retrospective estimates are the summary birth history data from the Censuses and the Community Survey other than the only full birth history estimate from 1998 DHS. The first four point estimates depict an under-five mortality rate of between 70 and 75 per 1,000, from 1994 to 2006. The 1996 SBH and the 1998 FBH estimates differ from those published elsewhere for a few reasons. The SBH have been adjusted for the misreporting of still births as live births that died resulting in a 17% decrease in level and the FBH rates were calculated for single year periods as opposed to the five-year average presented in the survey report (Nannan, Timaeus, Laubscher et al. 2007).
The two trend lines, represent the ratio of deaths under age five registered in the year of death as a proportion of the census births derived by reverse survival (solid line) and the adjusted DHIS births (diamond line). The final and most recent points derived from the household deaths uses a life-table approach (see section 3.8) to calculate $q_0^1$ (cross) and $q_0^5$ (star). Interestingly, the infant and under-five estimates up to 2007 suggest that the mortality remained more or less constant since 1994, whereas the two under-five trends calculated from the births illustrate an increase, and then a decrease in the mortality, undoubtedly due to higher incompleteness of registration of deaths in the earlier years. The point estimates suggest further that under-five mortality dropped from 75 to 46 per 1,000 live births and the IMR from 50 to 34 between 2006 and 2011.

The Rapid Mortality Surveillance (Dorrington, Bradshaw, Laubscher et al. 2014) and the second South Africa National Burden of Disease Study 1997–2010 (2nd NBD) (Msomburi, Pillay-van Wyk, Dorrington et al. 2014) serve as standards against which the mortality schedule proposed in this thesis can be compared. The thesis makes two adjustments to the mortality rates designated adjustment 1 and adjustment 2 and has utilised slightly different data from the NBD, namely, the more recent release of the household deaths (Stats SA 2014a). The comparison is shown in Figure 7-9.
The method used to estimate the mortality presented in the next section of the thesis is explained in the section 3.11.3.1. The mortality estimation is based on the expected number of deaths at two points in time, namely (a) the mortality produced by Darikwa in 2006.6 and (b) the mortality based on a life-table calculation of household deaths occurring in the year preceding the census (mid-point at 2011.27). The probability of dying by calendar year for the next five years is calculated from the VR deaths corrected for completeness. This is achieved by linear interpolation to calculate the annual change in completeness between 2006 and 2011. This was then applied to the VR deaths in the intervening years to produce a series of adjusted under-1 and under-5 deaths by calendar year. Figure 7-9 compares this trend (adjustment 1 - solid line) with the 2nd NBD trend (diamond) the Rapid Mortality Surveillance trend (circle) and an alternative trend based on adjusted household deaths in 2011 (adjustment 2 - dashed line).

Of note is the fact that all three estimates are quite consistent in the overall under-5 mortality, but the level of infant and child mortality are higher and lower respectively than the comparative estimates. Since the estimates are produced by linear interpolation between 2006 and 2011, the lack of agreement in terms of the components of the under-5 age group are probably explained by the common data error of heaping deaths at age one, particularly in enquiries where no probes are included in the questionnaire (Curtis 1995).

Figure 7-9: Comparison of mortality derived from adjusted vital registration applying alternative assumptions for estimating completeness
Legend: NBD 2014 (diamond); RMS 2014 (circle); adjustment 1 - mortality produced from adjusted VR deaths assuming completeness based on the 2007 CS and the reported HH deaths in 2011 (solid line); adjustment 2 - mortality produced from adjusted VR deaths assuming non-changing completeness based on the 2007 CS and the reported HH deaths in 2011 Census (dashed line).
A consequence of this form of heaping, is that it leads to an over-estimate of
deaths under age-one, and a corresponding under-estimate of completeness; the 1-4 year
old deaths would be too low and the corresponding completeness too high. Examining
this possibility shows that the completeness estimated from the reported deaths would
imply a rather unrealistic decline in the completeness of infant deaths from 87% to 72%
in five years (Table 7-5). Redistribution of the infant deaths to the child deaths solves
the problem of the seemingly unrealistic decline in completeness of infant deaths and
produces a more plausible estimate of completeness of child deaths, one that remained
stable between 2006 and 2011 at 55% and 53% rather than to have increased from 55%
to 61%. The redistribution based on the reasonableness of the implied completeness of
infant and child deaths since the 2007 Community Survey is presented in Table 7-5.

Table 7-5: Comparison of the implied completeness of death registration in 2011
estimated from the 2011 Census and the implied completeness estimated from the 2007
Community Survey

<table>
<thead>
<tr>
<th>Age group</th>
<th>% completeness in 2006 based on 2007 Community Survey</th>
<th>% completeness in 2011 based on 2011 Census</th>
<th>% completeness in 2011 implied by redistribution of infant and child deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infant</td>
<td>87</td>
<td>72</td>
<td>86</td>
</tr>
<tr>
<td>Under-five</td>
<td>76</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>Child</td>
<td>55</td>
<td>61</td>
<td>53</td>
</tr>
</tbody>
</table>

Although nationally the redistribution and resulting levels by age appear plausible,
these averages mask the variation in the completeness of death registration by province,
and although an alternative age-distribution of deaths was proposed, the ‘true’ ratio
of $q_0$ to $q_1$ and how this has changed over time in South Africa is still an outstanding
issue. The differences in completeness by province are assessed in Chapter 8.

7.5 **Age distribution of deaths within childhood**
The knowledge about the relationships between the components of under-five mortality
as mortality declines is based on a substantial body of empirical evidence, and the
mortality experience embedded in MLT’s all of which are representative of AIDS-free
populations. However, in populations affected by HIV/AIDS, the typical mortality
shape from birth to age five is affected by an increase in deaths at some stage during the
age-interval, although there are significant differences due to complex factors related to
transmission such as the age of transmission, viral-load and breastfeeding practices,
which influence the increase in age-specific rates. Epidemiological evidence supporting
increased deaths during the first two years of life have been linked to infection during
delivery. For the second to fifth year of life the HIV-infection through breastfeeding is the major cause (Luzuriaga, 2000). In South Africa, the identification of an AIDS signature around two to three months (Bourne, Thompson, Brody et al. 2009) further specifies the age-specific increase within infancy.

Linking the age-at-death to the causes of death it is well established that, relative to the overall under-five mortality rate, high levels of death during infancy are associated with congenital abnormalities and conditions originating in the perinatal period causing deaths in the neonatal period. In contrast, high child deaths are representative of an older-age pattern of mortality whereby the underlying causes of death are childhood infections associated with environmental factors such as sanitation, nutrition, immunisation and other health care practices and confirmed by many others (Galley and Woods 1999; Mosley and Chen 1984; Preston 1976). Given South Africa’s unique mortality experience, assessment of the CRVS using the standard tools and approaches recommended by the World Health Organisation (WHO 2010) present challenges, as they do not account for HIV/AIDS.

The establishment of what has occurred since that time would be informative. However, since then the AIDS epidemic matured and the introduction of PMTCT and ARV programmes in 2002 and the pneumococcal and rotavirus vaccines are all contributing factors to the behaviour of these indices.

Against the backdrop of mortality reversal caused by HIV infection, coupled with the mis-certification of HIV-related deaths, establishing the epidemiological expectation is problematic, the following strategy is employed to facilitate an assessment of epidemiological consistency of South African CRVS for children. First, a demographic analysis of the observed age-pattern that aims to evaluate the changes in the age-specific rates reported in censuses and surveys against Princeton model life tables. Second, a demographic analysis of the model values and empirical values standardised to the level of mortality estimated by the Thembisa model values and evaluated against the Princeton model life tables. Third, an examination of CRVS cause of death information in the form of National Burden of Disease estimates is undertaken as these estimates best demonstrate the probable age-cause trend and therefore provide an epidemiological expectation under the said conditions.

35 National Burden of Disease estimates are adjusted for completeness of registration and for the misclassification of causes of death and internal consistency.
7.6 **Demographic analysis of the coherence of the age pattern**

Two investigations that have used graphical comparison of the relationship between \( q_0 \) and \( q_1 \) against those of a standard to examine age-specific rates during childhood, one by Guillot and colleagues (Guillot, Gerland, Pelletier *et al.* 2012) and the other by Jasseh (2003).

Guillot *et al.* (2012) investigated the relationship between \( q_0 \) and \( q_1 \) for a given level of \( q_0 \) in many regions of the world by evaluating how well the empirical combinations of \( q_0 \) and \( q_1 \) are described by the mortality patterns embodied in the United Nations and the Princeton (Coale and Demeny 1966) model life tables. They conclude that the model life tables do not cover the entire range of age-patterns found in the selected empirical data, but that, the most divergent patterns were found in Sub-Saharan Africa and were due to the relatively high values of \( q_1 \) relative to \( q_0 \). The 1998 SADHS dataset outlines a childhood mortality pattern that transformed from one very close to the Princeton West to North and back to West as overall childhood mortality started increasing in the ten year period prior to the survey (Jasseh 2003: 142).

As a first step in exploring this relationship, we need to examine the behaviour of the infant and child mortality trends in the wake of the changes in overall under-five mortality since the 1998 DHS. This is done by considering the empirical estimates from surveys and censuses. Figure 7-10 shows three graphs spanning the period 1998 to 2011. The provincial relationships between \( q_0 \) and \( q_1 \) are compared to those of the Princeton models. From the first panel showing the 1998 DHS, it can be observed that most points lie between the West and North patterns. The second panel is based on the Community Survey and shows that all of the provincial points shifted towards the North pattern over the period. The final panel presents the most recent \( q_0 \) and \( q_1 \) combinations based on the provincial deaths reported by households in the 2011 Census and the 2011 national household death estimate adjusted for heaping at age one. Again there is more of a shift towards higher values of \( q_1 \) relative to \( q_0 \), illustrated by the fact that most of the points are positioned above the North family, but some have reverted towards the West model. The survey results also highlight the range in the level of mortality and the variation in age-specific rates of the provinces.
Figure 7-10: Observed $1q_0$ and $4q_1$ relative to the level of $5q_0$ relationships from the 1998 DHS, 2007 Community Survey and the 2011 Census compared with the Princeton model life tables.
Next, we consider the combinations of $4q_1$ and $1q_0$ from 1986 to 2011 based on modelled values from international agencies, compared against the Princeton model life tables. Figure 7-11 presents the relationship between $1q_0$ and $4q_1$ for $5q_0$ the same as that from the Thembisa model for each year (Johnson, Dorrington and Moolla 2016), for the: IGME(2015), WPP(2015) and IHME(2013), model values 1986-2011, compared against the Princeton MLT’s.

![Figure 7-11](image.png)

**Figure 7-11: Comparison of the relationship between $1q_0$ and $4q_1$ from international agencies at a level of $5q_0$ from the Thembisa model for each year 1986-2011 against the Princeton families of model life tables**

The representations of 1986-2011 values are plotted chronologically from the right to left.

All of the trends shown in Figure 7-11 demonstrate a shift between the families, albeit with different starting points, different end-points and different degrees of reversion. According to the 1998 DHS, the ratio lay mid-way between the West and North families. The WPP and IHME 2013 points indicate little or no shift from the DHS in 1998. However, the estimates from Thembisa and IGME show clear shifts in the age-pattern of infant-to-child deaths beyond the North model since the 1998 DHS. While IGME reverts to the North model in 2011, the most recent Thembisa values up to 2011 remain at higher ratios than those of the North model. IGME demonstrates the widest range of ratio values over the period, beginning close to the South family in 1986 and ending at the North family by 2011.
7.7 Establishing an epidemiological expectation in an AIDS environment

An alternative approach is to compare ratios of \( q_1 \) to \( q_0 \) against values standardised with the Thembisa model’s values.\(^{36}\) The next figure compares the North and West family \( q_1 \) to \( q_0 \) ratios corresponding to the Thembisa \( q_0 \) over time, and the Thembisa ratio to ratios from IHME\(^{37}\), IGME\(^{38}\) and the WPP\(^{39}\) standardised to Thembisa \( q_0 \) over time and shown Figure 7-12.

Figure 7-12: Comparison of the standardised ratios of \( q_1 \) relative to \( q_0 \) produced by IHME(2013), IGME(2015) and WPP(2015), against the values represented in Thembisa and the West and North MLT, 1986-2011

Generally, the expectation is that as the under-5 mortality rate increases, the proportion of child deaths relative to infant deaths increases. Largely, the modelled ratio values fall below the North and Thembisa models and above the West model. Thembisa begins to deviate from the West and North models in 1996, increasing more than the other models, while the difference between IHME and WPP ratios is fairly consistent throughout the period. The IGME ratios on the other hand are quite inconsistent between 1991 and 1996 and again in about 2006. Presumably, this is due to a change in their methodology. The difference between the ratios from the West and North models and Thembisa is due to the absence of HIV-related mortality in the Princeton tables.

\(^{36}\) http://www.thembisa.org/content/downloadPage/ThembisaVersions3
\(^{37}\) (http://vizhub.healthdata.org/mortality/
\(^{38}\) http://www.childmortality.org/index.php?r=site/graphRID=ZAF_South_Africa
However, the higher ratios of Thembisa and slower decline compared with the West and North models are likely because of ART impact on the IMR before they impact on deaths above age one.

Figure 7-13: Comparison of the standardised ratios of $q_1$ to $q_0$ produced by the CRVS, RMS, NBD against the values represented in Thembisa and the West and North MLT, 2000-2011

Empirical values of $q_1$ to $q_0$ ratios from the CRVS, the RMS and the Second NBD from 2000 to 2011, standardised to the Thembisa values of $q_0$ over time compared with the West, North and Thembisa ratios are shown in Figure 7-13. All of these trends in the ratio are much closer to the North model and to Thembisa (in terms of level) than were the values from IHME, IGME and WPP. Interestingly, although the level of CVRS ratio is clearly too low, it progresses consistently over time and lies above those of the various agencies, and thus may offer some insight into the contribution of the components of the ratio over this period in spite of the under-reporting bias.

The next step towards establishing an epidemiological expectation is to use CRVS data corrected for incompleteness by age and secondly to apply the NBD cause of death distributions to address the misclassifications and coding errors on the DNF (Pillay-Van Wyk, Laubscher, Msemburi et al. 2014). The last but arguably most severe deficiency in the CRVS to correct for is the number of deaths attributable to HIV/AIDS as an underlying cause of death, (Alkema and You 2012) which the NBD modelled.
Figure 7-14: Trend in infant and child mortality rates by broad cause group, 1997-2010
Source: Msemburi, Pillay-van Wyk, Dorrington et al. 2014

Figure 7-14 shows the NBD results for infants and children aged 1-4, over time. The broad cause of death groups for infants show communicable and HIV/AIDS death rates peak in 2005, thereafter HIV/AIDS death rates decline as overall under-five mortality declines, whereas, the deaths due to communicable diseases begin to decline only four years later. Deaths due to non-communicable diseases and injuries are stable throughout the period regardless of the level of under-five mortality. In contrast, the cause of death pattern for children aged 1-4 years shows a peak in HIV/AIDS deaths in 2004, but stable rates from the communicable, non-communicable and injury deaths.

The epidemiological expectation by age provided by Rao’s analytical framework is relevant to South African data because the cause-proportions (diarrhoea and LRI, perinatal, congenital and other) used in the prediction models were validated with South African 2005 vital registration data (Rao, Adair and Kinfu 2011). Figure 7-14, complies with the general theory that a decline in the overall level of under-five and infant mortality is accompanied by a decrease in infectious diseases. Specific to the rates within infancy, it is clear that HIV/AIDS has not affected the neonatal rate. Bourne et al. (2009) observed an increase in HIV-related deaths reported in the CRVS between 1997 and 2002, with a peak between ages two and three months. This finding was consistent with the emergence of a similar peak at four months found by Newell, Coovadia, Cortina-Borja and colleagues (2004). However, no evidence of increased mortality during the neonatal period in the South African population has been reported, or
observed in the CRVS. On the contrary, even though independent of the careful adjustments applied to the NBD indicate substantial increase in the post-neonatal rate due to HIV/AIDS, the CRVS clearly indicates an increase as well as corresponding decrease in accordance with the expectation from PMTCT occurring later.

7.8 Understanding the pattern in South Africa
Garenne (1982: 260) first noted the presence of a pattern with relatively higher values of \( q_1 \) in a Senegalese surveillance site at the end of the 1970’s and suggests that the resulting epidemiology of this mortality pattern is characterized by a high prevalence of malaria, measles and diarrhoea, all of which contribute to shifting the mortality beyond the first birthday. Since this initial observation, a similar age-pattern has been observed in several demographic surveillance sites on the sub-continent where malaria is endemic (Abdullah, Adazu, Masanja et al. 2007). The reason for South Africa’s relatively high child-to-infant mortality ratios in 2011 are less clear, because there is no reported evidence to suggest children older than one have been more severely affected by HIV-deaths than infants have. However, as the impact of the antiretroviral roll-out (PMTCT and ART to mothers) would first impact on infants, perhaps the higher ratios are a consequence of that effect.

Guillot et al. (2012) suggest that the HIV/AIDS epidemic may offer an alternative epidemiological explanation for the high child relative to infant mortality identified as a more prevalent pattern recently, but conclude that actually the contribution of AIDS remains unclear because of “the complexity of the age of infection among HIV positive babies in the South African context”. Certainly, the South African experience shows excess mortality during the post-neonatal period (Bourne, Thompson, Brody et al. 2009; Newell, Coovadia, Cortina-Borja et al. 2004) and illustrated in Figure 7-6, differences in the cause-profile that have led to the increases in the ratio observed in other countries. However, PMTCT has affected a compression of the under-five mortality rate in the post-neonatal period, in contrast to the prediction of a compression of the under-five mortality rate into the neonatal period (Rao, Adair and Kinfu 2011).

7.9 Conclusion
The chapter has shown that the evaluation of completeness is only as good as the accuracy of the demographic standards against which completeness is measured, but it has also questioned the accuracy of the age-specific mortality data the analysis of the completeness of death registration is based on.
There has been significant improvement in the completeness of birth registration and death registration particularly since 2000. In 2010, the completeness of birth registration (83%) was slightly less than the completeness of registration of infant deaths implied by the redistributed infant and child deaths reported in the 2011 Census (86%). The completeness of reporting of under-5 deaths was 73%. Both the completeness of birth registration and death registration appear to have stabilised nationally since around 2006.

Similarly, the investigation of the impact of AIDS on the age-pattern of childhood mortality has shown that the existing MLT and the model-based mortality standards do not capture the changes in the age-pattern due to HIV, with the exception of the Thembisa model.
8.1 **Provincial estimates of the completeness of birth registration**

UNICEF (2013) highlights the fact that civil registration systems are most wanting in rural settings and South Africa exhibits wide inequalities in this regard associated with the level of development across the nine provinces which is strongly correlated with several development indicators. The next sub-section therefore assesses the completeness of vital registration by province.

Estimation of provincial births was achieved by back-projection of the numbers counted in the 2011 Census by province. Figure 8-1 shows that the annual number of births by province ranged from around 25,000 per annum in the Northern Cape to around 250,000 per annum in KwaZulu-Natal over the last decade.

![Figure 8-1: Provincial births derived by back projecting the recorded province of birth recorded in the 2011 Census, 1996-2010](image)

As pointed out above, the adjusted DHIS corroborates the numbers of births derived by reverse survival nationally indicating that, in aggregate, the adjustment for non-public sector births is appropriate. To obtain the provincial DHIS adjustment factors, the regression equation shown in (section 3.11.2.3) was employed to predict the proportions of private-facility and home-births by province. A provincial comparison of the adjusted and unadjusted DHIS trends, the number of births produced by reverse
survival and the number of births registered in the VR year of birth +1 are shown in Figure 8-2.

The plots reveal that all provinces with the exception of Mpumalanga and North West show that the adjusted DHIS reflects increased coverage, where, in the recent period, the ratio of DHIS adjusted births and births derived from RS is one; in Limpopo and Northern Cape by 2008 and in Eastern Cape and Kwa Zulu-Natal by 2010. Since the DHIS births prior to 2008 are below 100% nationally, this suggests that post-2008 the births from Free State, Gauteng and Western Cape must compensate for the births missing in Mpumalanga, North West and the Eastern Cape and Kwa Zulu-Natal up to 2010. Moreover, although several factors may be responsible for the Census producing higher or lower numbers of births than the adjusted DHIS, the most significant factors lie in errors in either the data sources or errors arising from the possible mismatch between place of birth and place of registration (province on the BNF). It is therefore necessary to investigate each province in more detail.

Since the purpose is to assess completeness of registration of births registered provincially and given the potential error in the DHIS adjustments it is also useful to investigate which empirical estimate provides the most plausible (consistent) trend in relation to the provincial registration pattern represented by the VR +1 trend for each province. An investigation into the number of births by Dorrington and Moultrie (2015) compares the estimates, in addition with the provincial TFR estimates and school enrolment data as an alternative source of corroboration. This thesis utilises their suggested adjustments to the Census estimates, which amount to minor refinements for the Eastern Cape, Gauteng, Limpopo, Northern Cape and North West trends and are summarised in Table 8-1. The trends produced by reverse survival in the remaining provinces Free State, Kwa Zulu-Natal, Mpumalanga and Western Cape were accepted as they were supported by the TFR’s and school enrolment data and therefore assumed to require no further adjustment.
Table 8-1: Adjustments to the provincial number of births derived from reverse survival (RS), Census 2011

<table>
<thead>
<tr>
<th>Province</th>
<th>Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Cape</td>
<td>RS is not supported by TFRs, or by the adjusted DHIS, however it is supported by school data up to 2000, after which census/school attendance is under estimated by around 7%. Indication that DHIS coverage increased since 2004 to near complete by 2010. VR+1 trend matches the RS and DHIS numbers match VR+1 by 2010, thus assume that DHIS coverage such that true DHIS numbers matched VR+1. This suggests that recent RS estimates are too high by about 10%. Thus, RS estimates were adjusted (linearly) by 0% in 1996 to -10% in 2011.</td>
</tr>
<tr>
<td>Gauteng</td>
<td>RS is not supported by TFRs which suggest RS is 10% too low, and also not supported by DHIS (suggests RS 20% too low), but school data support 1996-2004. There is a confused picture, but adjusting RS estimates (linearly) by 0% in 1996 to +10% in 2011 appears reasonable.</td>
</tr>
<tr>
<td>Limpopo</td>
<td>RS is not supported by TFRs as these estimates are about 5% too low in the early years and 5% too high in recent years which is supported by the DHIS data, but not by school data. The evidence is contradictory. RS estimates were adjusted linearly from +5% in 1996 to -5% in 2011.</td>
</tr>
<tr>
<td>Northern Cape</td>
<td>RS is not supported by the TFR, which suggest the RS estimates are 10%-20% too high and is somewhat supported by the DHIS (suggests RS is about 5% too high), and is also supported by school data 1999-2004 before which census/school attendance is underestimated by 5%. The conclusion was to reduce RS by 5% throughout.</td>
</tr>
<tr>
<td>North West</td>
<td>RS supported by TFRs, but is not supported by DHIS (suggests RS 6% too high up to 2009 then 13% too high), and school data do not support (suggest census/school attendance 5% too high). The conclusion was to reduce the RS by 5%.</td>
</tr>
</tbody>
</table>

Source: Dorrington and Moultrie, 2015

The levels of completeness implied by the two empirical estimates, (the one based on the reverse survival and the other, the adjusted DHIS) are shown in Table 8-2 and graphically presented in Figure 8-2. As is apparent from Table 8-2 in most provinces, and more so in the earlier years, the adjusted DHIS under-estimates births and produces implied estimates of completeness of over 100% for the VR +1 registration. We will therefore use the Census derived estimates to measure completeness, and examine the relationship between the alternative sources of births to interpret the birth trend for each province in turn.
Figure 8-2: Provincial comparison of trends in the number of births from various sources
Table 8-2: Implied completeness of registered births registered in the year of birth +1 compared with births derived from the 2011 Census and adjusted DHIS births, 2001-2011

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>Census</td>
<td>60</td>
<td>64</td>
<td>69</td>
<td>75</td>
<td>80</td>
<td>83</td>
<td>80</td>
<td>81</td>
<td>80</td>
<td>82</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>122</td>
<td>121</td>
<td>120</td>
<td>115</td>
<td>111</td>
<td>104</td>
<td>102</td>
<td>94</td>
</tr>
<tr>
<td>FS</td>
<td>Census</td>
<td>66</td>
<td>74</td>
<td>79</td>
<td>80</td>
<td>83</td>
<td>86</td>
<td>82</td>
<td>86</td>
<td>83</td>
<td>86</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>101</td>
<td>103</td>
<td>105</td>
<td>105</td>
<td>107</td>
<td>107</td>
<td>112</td>
<td>112</td>
</tr>
<tr>
<td>GT</td>
<td>Census</td>
<td>84</td>
<td>87</td>
<td>85</td>
<td>90</td>
<td>89</td>
<td>85</td>
<td>83</td>
<td>82</td>
<td>79</td>
<td>76</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>96</td>
<td>97</td>
<td>98</td>
<td>97</td>
<td>93</td>
<td>98</td>
<td>95</td>
<td>-</td>
</tr>
<tr>
<td>KZN</td>
<td>Census</td>
<td>67</td>
<td>81</td>
<td>77</td>
<td>76</td>
<td>80</td>
<td>87</td>
<td>84</td>
<td>86</td>
<td>82</td>
<td>84</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>100</td>
<td>105</td>
<td>110</td>
<td>110</td>
<td>108</td>
<td>107</td>
<td>109</td>
<td>108</td>
</tr>
<tr>
<td>LM</td>
<td>Census</td>
<td>65</td>
<td>67</td>
<td>70</td>
<td>74</td>
<td>77</td>
<td>77</td>
<td>78</td>
<td>82</td>
<td>81</td>
<td>82</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>98</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>95</td>
<td>95</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>MP</td>
<td>Census</td>
<td>57</td>
<td>67</td>
<td>74</td>
<td>83</td>
<td>83</td>
<td>85</td>
<td>86</td>
<td>88</td>
<td>84</td>
<td>86</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>113</td>
<td>110</td>
<td>110</td>
<td>110</td>
<td>113</td>
<td>113</td>
<td>113</td>
<td>113</td>
</tr>
<tr>
<td>NC</td>
<td>Census</td>
<td>72</td>
<td>76</td>
<td>79</td>
<td>84</td>
<td>88</td>
<td>89</td>
<td>90</td>
<td>93</td>
<td>90</td>
<td>91</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>105</td>
<td>107</td>
<td>108</td>
<td>111</td>
<td>112</td>
<td>113</td>
<td>113</td>
<td>114</td>
</tr>
<tr>
<td>NW</td>
<td>Census</td>
<td>78</td>
<td>83</td>
<td>90</td>
<td>93</td>
<td>101</td>
<td>99</td>
<td>100</td>
<td>103</td>
<td>100</td>
<td>97</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>126</td>
<td>130</td>
<td>132</td>
<td>134</td>
<td>132</td>
<td>132</td>
<td>137</td>
<td>134</td>
</tr>
<tr>
<td>WC</td>
<td>Census</td>
<td>85</td>
<td>90</td>
<td>87</td>
<td>95</td>
<td>97</td>
<td>98</td>
<td>92</td>
<td>94</td>
<td>89</td>
<td>87</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>104</td>
<td>111</td>
<td>112</td>
<td>111</td>
<td>110</td>
<td>111</td>
<td>115</td>
<td>114</td>
</tr>
<tr>
<td>SA</td>
<td>Census</td>
<td>73</td>
<td>79</td>
<td>79</td>
<td>82</td>
<td>86</td>
<td>86</td>
<td>84</td>
<td>86</td>
<td>84</td>
<td>83</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DHIS</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>105</td>
<td>107</td>
<td>105</td>
<td>103</td>
<td>102</td>
<td>99</td>
<td>101</td>
<td>102</td>
</tr>
</tbody>
</table>

In terms of migration during the inter-censal period 2001 to 2011, the 2011 Census information (Stats SA 2012a), about the province last moved from, show that the Eastern Cape, Northern Cape, Free State, Kwa Zulu-Natal and Limpopo experienced a net loss of people while, not surprisingly, the Western Cape and Gauteng experienced the greatest gains. However, the North West and Mpumalanga also experienced net gains. This pattern is mirrored by the lifetime migration movements, which also indicate that Eastern Cape, Northern Cape, Free State, Kwa Zulu-Natal and Limpopo have negative net migration indicating more people have left these provinces than moved in over time. Gauteng also experiences the largest net immigration, but North West and Mpumalanga also experience high net movement. Interestingly the report also states that of all the migration into the North West, Free State and Gauteng, 20%, 13% and 22% respectively was non-South Africans.

Eastern Cape - The reverse survival (RS) trend is considerably higher than the adjusted DHIS trend until 2010, indicating that the routine system has failed to capture a significant proportion of births in the earlier years although the system does improve. The reverse survival displays the same trend as the VR, suggesting fertility peaked between 2006 and 2007. The higher number of births produced from RS compared with the number estimated from the fertility rates suggests out-migration of children (mostly
to the Western Cape). Since the 2011 fertility estimate and the adjusted DHIS are consistent, this suggests that the difference between these and the RS estimates is due to out-migration since birth which is confirmed by the inter-censal migration (Stats SA 2012b). Completeness in 2010 was 82%.

*Free State* - Some degree of out migration (probably mostly to Gauteng) is suggested by the fact that RS produces slightly higher births than the measures of fertility. The fact that the adjusted DHIS is higher than the RS may indicate a proportion of births not belonging to the province have occurred in the provincial health facilities. The estimated completeness in 2010 was 86%.

*Gauteng* - All of the trends indicate an increase in births. The RS indicates an increase of 100,000 births between 1996 and 2011. The births produced by RS are lower than the fertility estimates up until about 2006, at which point the number of births implied by the RS and TFR become very similar. It is likely that the adjusted DHIS are higher than the RS due to births from other provinces taking place in health facilities in Gauteng, but may also be due to incorrect estimates of home-based and or private-facility births. The opposite scenario is unlikely, that a higher proportion of women covered by medical aid giving birth in public facilities. An important consideration is Gauteng’s central geographic position regarding births from residents living in other provinces (North West, Limpopo, Mpumalanga and Kwa Zulu-Natal) but also births from non-South Africans. The 2011 Census indicates that Gauteng experienced the largest influx of people into the province and of those 22% were non-South Africans, making the monitoring of births occurring in Gauteng and subsequent registration complicated. Completeness in 2010 was 76%.

*KwaZulu-Natal* - There is little difference in level between all three trends and the estimates derived using fertility rates, although the 2006 TFR does appear slightly low. The RS indicates out-migration and this is supported by the reported net out-migration from KZN (Stats SA 2012b). The number of births increased slightly between 2000 and 2011. The adjusted DHIS appears a suitable tool for monitoring the number of births in the future. Completeness in 2010 was 84%.

*Limpopo* - Again there is little difference between the levels suggested by the estimates, although the increasing VR in recent times does indicate improvement in registration. The small disagreement between the most recent TFR and the adjusted DHIS suggests that the routinely collected birth data in these health facilities are of poor
8.2 Provincial estimates of the completeness of death registration

Knowledge of provincial levels of under-reporting of deaths during childhood are absent from the literature, indicating a lack of research in this area, however, in the form of an assessment of completeness of vital registration between 1997 and 2010, previous research (Pillay-Van Wyk, Laubscher, Msemburi et al. 2014) has produced trends in provincial completeness to produce the Second National Burden of Disease Estimates (NBD).

In the case of estimating completeness of the reported deaths of children under-1 and under-5 years old, the NBD study compared the reported deaths with mortality rates derived from censuses and surveys (Msemburi, Pillay-van Wyk, Dorrington et al. 2014). The NBD used initial national and provincial estimates for the years 1996, 2001,
2007 and 2011 from several sources. The 2001 estimates were primarily those from Dorrington, Moultrie and Timeaus (2004), except for KZN $q_0$, which was estimated as the average of the estimate directly from the reported census household deaths and that derived by Dorrington et al (2004) and the $q_0$ was estimated from $q_0$ in the same ratio as estimated by Dorrington et al. In the case of the North West $q_0$ was estimated directly from the census deaths in the household. The estimates derived by Darikwa (2009) were used for 2006 for each of the provinces. The 2011 estimates derived from census household deaths (Stats SA 2013a) excluded any of the deaths recorded with unspecified age (this was approximately 25% of total deaths) being attributed to childhood deaths. The 1996 estimates produced from the ASSA2008 model were used which were calibrated using birth histories from the 1996 census and 1998 DHS.

To improve the robustness of the NBD estimates, the estimates were smoothed by regressing separately for $q_0$ and $q_0$, provincial estimates for each census/survey set on each other set using orthogonal regression.

The completeness of registered deaths under age-1 and under age-5 was estimated in 1996, 2001, 2006 and 2011 as the ratio of VR deaths to those expected from the annual number of births multiplied by the smoothed estimates of $q_0$ and $q_0$. Estimates of completeness for the intervening years were derived by linear interpolation and the schedule of $q_0$ and $q_0$ estimates assessed for reasonableness and adjusted where necessary. The provincial estimates in each year were adjusted proportionally to ensure that the sum of the estimated deaths were the same as the number of deaths for the country in total, after adjustment for completeness assuming that the national $q_0$ and $q_0$ estimates are more likely reliable than those for the individual provinces are.

Aside from the procedure outlined above, specific adjustments to the trend in completeness were made to the VR data to allow for the following anomalies:

1) Western Cape: A sharp increase in deaths between 1997 and 1999.
3) Northern Cape: a sharp increase in deaths in 2007 and 2008.
5) North West: a significant decline after 2006.
A difference in the VR data used is that the NBD study used late registrations only up until the release of the 2010 data, whereas this thesis was updated with late registrations processed up to 2013 (Stats SA 2014b).

This thesis estimates completeness between 2006 and 2011 by interpolating linearly by single year, which limits the estimation of the trend and it places a lot of reliance on the accuracy of the two data points at the beginning and end of the period. Reviewing factors that may cast doubt on the reliability of these values, beginning with Darikwa’s 2006 mortality level derived from the reported SBH points of mothers over age 20 and the data point from the household deaths. Logistic regression was used to smooth the pairs of SBH and reported household death estimates. The question related to household deaths was phrased in such a way that the reference period could well have been understood to refer to 13 or 14 months rather than the intended 12 months. These data therefore may well have produced a slightly higher estimate than the ‘true’ rate, as noted by Darikwa and Dorrington (2011). Also, another source of potential bias in the SBH data stems from the absence of a suitable adjustment for changing HIV prevalence over time (Darikwa and Dorrington 2011).

Differences related to the data used in the NBD analysis and in this thesis are that the thesis uses the more recently released data set cleaned to reduce the large proportions of unspecified age and sex (Stats SA 2014a) resulting in a net effect of lower values of deaths in all provinces. The authors of the NBD study (Pillay-Van Wyk, Laubscher, Msemburi et al. 2014) noted that mortality rates in the more recent data set were 3 per 1,000 live births higher in the Western Cape; 2 per 1,000 lower in the Free State and that other changes would not have amounted to more than 1 per 1,000 in the remaining provinces.

In addition, two issues related to data quality, are that neither the SBH or the household deaths questions are designed to produce age-specific rates, but instead to provide ‘pegs’ (empirical point estimates) of the underlying level of under-five mortality. However, the absence of age-specific mortality indices for South Africa since 1998, demands that any available data be utilised despite the limitations of these data.

Despite all of these caveats, the approach used to produce national completeness appears to present a feasible option for the provinces, mainly because the alternative explanation that completeness deteriorated from 2006 to the extent implied by the

---

40 The question read “MORTALITY IN THE LAST 12 MONTHS – (BETWEEN FEBRUARY 2006 AND MARCH 2007). READ OUT: I am now going to ask for information on any member of the household who has passed away in the last 12 months between February 2006 and March 2007. The 2011 Census did reflect a reference period of 12 months.
reporting of the household deaths in 2011 is not plausible. Nevertheless, if one assumes that the reporting of deaths under-five was reasonably accurate in the 2011 Census, the provincial completeness, shown in Figure 8-3, has remained more or less constant for most provinces since 2006. The percentages of completeness by province are shown in Appendix 8.

Hence, applying the same method to the provinces and assuming that the level of under-five mortality is reliable but that a redistribution of infant deaths to the older age group minimises heaping at age one and thereby negates the implied exaggerated improvement of completeness of 1-4 year old deaths over time.

The comparative graphs in Figure 8-4 show that the NBD \( 5q_0 \) trend (cross), the \( 5q_0 \) trend produced in this thesis (adjustment 1 \( 5q_0 \) - solid line) and the magnitude of heaping, indicated by the difference between the reported \( 1q_0 \) (square) and reported \( 4q_1 \) (.circle) and the adjusted trends (adjustment 2 \( 1q_0 \) and adjustment 2 \( 4q_1 \) – dashed line) differ by province. It is hard to think why the error was smaller in the Eastern Cape, Limpopo and the Northern Cape compared with the other provinces.
Figure 8-4: Comparison of reported and adjusted $q_0$, $q_1$ and $q_5$ against the NBD study, by province.
Figure 8-3 reveals that four provinces; the Free State, Gauteng, Northern Cape and North West, exceed 100 per cent completeness (over the 6-year period). This indicates that there were more registered deaths than the empirically derived estimates, while the opposite is apparent in Limpopo and the North West post 2009, which have fewer registered, while the Eastern Cape, KwaZulu-Natal and Mpumalanga have far fewer registered deaths. An increase in completeness of under-five deaths experienced in four provinces could also indicate under-reporting in the census. Over-reporting is indicated by the $q_0$ trend estimated in this research, before any smoothing compared with the NBD trend as illustrated in Figure 7-9. On the other hand, the increased completeness could signify genuine improvement in registration, the difficulty, however, is that verification of either explanation is not possible now, so these uncertainties remain. The NBD study also reported completeness of infant death registration exceeding 100% in the same provinces as those in Figure 8-3 with the addition of the Northern Cape.

Interpreting the changes in mortality and changes in completeness at the same time is challenging because it involves the inter-play of several factors specific to individual provinces such as the province of death, province of residence, province of registration and boundary changes, which have affected some provinces during the period under investigation. Assessment of the proposed age-adjustment is undertaken in Figures 8-5 and 8-6 by comparing the IMR trend produced from the reported household deaths (left panel – solid line) with the IMR trend produced from the reported household deaths that is adjusted for heaping at age one (right panel – solid line ‘adjusted’). The reported and adjusted IMR trends estimated in this thesis are compared against other empirical estimates of IMR from the NBD (hollow-circle); the provincial unadjusted ratios of the reported VR deaths and the reported births in VR births+1 (dashed line); the 1998 full-birth history (solid triangle); the 2006 summary birth history (Darikwa 2009) (solid circle); and the 2011 reported household deaths based on the life-table calculation (solid square). Figures 8-5 and 8-6 show the comparative trends by province, and drawing on the NBD results prior to 2006, are interpreted in turn next. The same provincial comparison for children under-five years of age is shown in Appendix 9.

41 Confidence intervals have not been provided owing to the fact that calculation is complicated by not only for the allowance of PES adjustment to person data, but unknown uncertainty about the estimate of the numbers of deaths reported by households.
Figure 8-5: Comparison of reported (left panel) and age adjusted (right panel) infant mortality against the NBD study and other empirical estimates, by province.
Figure 8-6 continued: Comparison of reported (left panel) and age adjusted (right panel) infant mortality against the NBD study and other empirical estimates, by province.
**Eastern Cape** – Infant mortality remained more or less stable at around 61 per 1,000, from 1995 up until 2006, after which it declined uniformly. Problems in the VR system (most likely to be associated with boundary changes) are indicated by a decline in deaths between 2008 and 2009 and then an increase from 2009-2010. The unadjusted IMR ratio is roughly half the empirical value suggesting high levels of under-reporting confirmed by the estimated 46% of infant deaths reported in the VR in 2011. Completeness of under-1 deaths improved from 30 per cent in 2000, to about 48 per cent in 2006, after which it appears to have remained unchanged. Little improvement is seen in completeness of under-five year old deaths after 2006, when it was 53 per cent in 2011.

**Free State** – Infant mortality was about 52 per 1,000 in 1995, and peaked at about 68 in 2005. Estimated completeness of under-1 deaths was 92% in 2000 and 125% in 2010, indicating that the survey and census estimates are lower than the actual number of deaths reported in the provincial VR system. The NBD notes an unstable completeness during the period reaching 136% in 2008. The completeness of under-5 registration was about 111% in 2011 and the unadjusted ratio produces a rate far higher than produced by empirical estimates, which suggests that the VR system registers more child deaths in this province than were reported in any of the other sources. The proposed adjustment appears appropriate as it offers a closer match to the other estimates.

**Gauteng** – The IMR was 37 per 1,000 in 1995, and peaked in 2003 at just over 40 per 1,000. Registration of under-1 deaths also exceeds 100% in this province (115% in 2000 and 122% in 2010). The inconsistency may be caused by a mismatch of closely related factors; province of death, province of residence and the province of death registration. Completeness of under-five year old deaths was about 111% in 2011. As with the Free State, the provincial boundary changes in 2005 may also be playing a role. The proposed adjustment appears appropriate as it provides better agreement with the other estimates.

**Kwa Zulu-Natal** – The IMR increased from 52 per 1,000 in 1995, peaked around 70 per 1,000 in 2003. The estimated completeness was 42% in 2000 and 45% in 2010. Completeness of under-five deaths was 62% in 2011 and borne out by much lower unadjusted ratios than the empirical estimates. Since the reported IMR in 2011 is clearly too high, perhaps the proposed redistribution of infant deaths is too severe.
**Limpopo** – Mortality appears to have remained stable at 37 per 1,000 between 1995 and 2005. Improvement in the estimated completeness in 2000 from 47% to about 100% in 2010. The unadjusted ratio being very similar to the empirical trend is perhaps suggestive of near complete registration since 2006/2007. The proposed adjustment fits the other data points better than the reported deaths. The estimated completeness of deaths under the age five was about 80% in 2011.

**Mpumalanga** – The IMR increased significantly between 1995 when it was 47 per 1,000 until peaking in 2003 at slightly over 70 per 1,000. Completeness was 40% in 2000 and 67% in 2010 appearing quite stable since 2006. The under-5 completeness was 64% in 2011. Sustained under-reporting is reflected in the unadjusted IMR being lower than the empirical estimates. The adjustment to infant deaths offers closer comparison to the other data points. Noteworthy is the minimal difference between the unadjusted ratio and the estimate produced using the adjusted infant deaths since 2006.

**Northern Cape** – The IMR decreased slightly between 1995 and 2006. There are anomalies in the VR such as the sharp decline in 2005. The odd mortality trend estimated in this research, and replicated by the unadjusted VR ratio is perhaps suggestive of problems with the estimated births and the 2006 mortality estimate appears to be too low. The proposed adjustment does not affect the IMR or $4q_1$ because the completeness of under-5 deaths increased from 90% in 2006 to 96% in 2011, adding further weight to the 2006 mortality estimate being problematic.

**North West** – The IMR was 42 per 1,000 in 1995 and peaked at 64 per 1,000 in 2003. The adjusted infant deaths result in a more plausible IMR trend than that of the reported deaths. The completeness of infant deaths reported in the NBD is marked by fluctuations from 87% in 2000, 139% in 2006 and 110% in 2010, strongly indicating boundary changes. The completeness of deaths under age five in 2011 was 92%.

**Western Cape** – The Western Cape has had the lowest infant mortality in the country since 1995 when it was 30 per 1,000 and 24 per 1,000 in 2006. The NBD reports many years where completeness of reporting of infant deaths exceed 100%, however, the estimated under-5 completeness was under 100% improving from 78% to 86% over the period. The implied difference in the IMR produced from the reported and adjusted household deaths in 2011 demonstrates the misleading potential of the unadjusted ratio.
The method used to assess completeness offers an element of ‘internal consistency’ because of the use of live births in the derivation of the mortality and the inclusion of the late registrations in the process as opposed to relying on model-based estimates. The framework lends itself well to evaluate the reported ratios against empirical estimates and this provides an assessment tool to measure the performance of provincial registration systems. Importantly the comparison of unadjusted mortality ratios with empirical estimates highlights how misrepresentative the reported ratios are, and demonstrates that the practice of presenting reported unadjusted ratios as ‘true’ provincial IMR’s in publications such as the Committee on Morbidity and Mortality (COMMIC 2014), advisory to the Ministry of Health, should be discouraged in future.

8.3 Conclusion

The analysis of provincial mortality reaffirms the known geographic differences in the level of mortality. Careful analysis of the provincial cause of death profile in children by the second South African NBD study (Msemburi, Pillay-van Wyk, Dorrington et al. 2016) provides important information for policy makers. The study attributes the high mortality in KwaZulu-Natal, Mpumalanga, Free State and the Eastern Cape to a predominance of pre-transitional infections and HIV. In contrast, the lower mortality provinces, the Western Cape and Gauteng display proportionally more deaths due to neonatal causes than deaths due to infections. The provincial declines in mortality observed since around 2005 are the result of the interventions to reduce mother-to-child transmission of HIV. In addition, interventions to address diarrhoeal disease and pneumonia in children contributed to the decline the death rates from 2008 and 2010 respectively.

The greatest challenge posed by the research questions concerning the completeness of birth and death registration was the interpretation of the provincial births and fertility trends and the provincial deaths and mortality trends. This is due to the interplay of inter-provincial movement and international immigration into South Africa on the one hand, and the occurrence of a birth or death in a particular province and the registration of that birth or death in perhaps a different province on the other. The most recent information regarding inter-censal and lifetime migration reports that there is considerable interprovincial movement as well as increasing international immigration playing a role in provincial birth trends. This must therefore be evaluated when measuring completeness of provincial civil registration. To address the statistical
challenge of computing meaningful vital rates with the appropriate sub-national populations-at-risk, it would useful to ensure that both vital statistics and the enumerated population are tabulated by usual place of residence.

The provinces have all experienced improvements in the completeness of birth and death registration over time and the relative differences in completeness between the provinces have narrowed over time. The fact that registration of child deaths appears to exceed 100% suggests a lack of correspondence between the province where the death occurred and the province of the registration of birth, possibly because of changing provincial boundaries in the data. Similarly, with the interpretation of the completeness of provincial birth registration, the migration of young children since birth to a different province of residence is a challenge, although the problem is more acute in some provinces than in others.

Evaluation of birth and death registration shows there has been great improvement in the completeness of both registration systems particularly since 2000; however, levels of completeness appear to have levelled off since 2006. Nationally 86% and 83% of births were registered in the year of birth +1 in 2008 and 2010 respectively. The provincial variation ranged between 100% and 80% between 2008 and 2010. In comparison, registration of deaths under the age of five exhibited a lower national average of 75% in 2008 and 72% in 2011 and wider provincial variation of 113% to 52%. Unlike the births where late birth registrations improve in the levels of completeness year on year, late death registrations have little impact on the levels of completeness.

The extensive and sustained under-reporting of deaths (Table 7-7) in the Eastern Cape, KwaZulu-Natal, Limpopo and Mpumalanga is concerning. In this regard, South Africa could consider exploring the option to proactively search for deaths and live births as proved fruitful for Brazil (Szwarcwald, Germano de Frias, deSouza Junior et al. 2014). Between 2006 and 2011, completeness averaged 50%, 62%, 77% and 66% in these provinces respectively and is strongly associated with general levels of development such as areas of land designated rural and therefore geographically remote, making accessibility to administrative offices challenging for poor communities. Although Appendix 8 illustrates the over-reporting implied by the under-5 household deaths in several provinces as well as for the country as a whole, the table also highlights provinces where completeness exceeded 100%. Between 2006 and 2011 on average about 74% of deaths were registered nationally, implying inter-
provincial displacement. This observation is partially dependent on the fact that provincial tabulations were generated using ‘province of death’, however, in the case that the deceased died in a province other than the one in which they lived ‘province of usual residence’ would be more suitable to ascertain completeness by province. Again, there is wide provincial variation in this regard. For example, in 2011 the proportions of deceased children who did not die in the province in which they lived are as follows: EC 18%, NW 16%, WC 12%, LP 11%, GT 10%, KZN 9%, MP and NC 8%, FS 7%. Municipal boundary changes have in some instances affected provincial boundary changes since 2000, and this may also have led to the non-correspondence of the province of death and the province of usual residence, but according to Statistics South Africa these would mostly have occurred in 2005 and 2011 (Stats SA 2012b).

Even considering these two reasons for lack of correspondence provincially, there is a strong indication that non-negligible proportions of child deaths are never registered in a number of provinces. Detailed investigation of province-specific factors will be a necessary first step going forward and gaining understanding for monitoring provincial CRVS because the extent of under-reporting in some of the provinces suggests the reasons are structural and will therefore require concerted interventions in those localities.
9.1 Introduction
The overall aim of this research was to assess how childhood mortality can be best monitored in low and middle-income countries, using South Africa as a case study. The four sources of data for measuring childhood mortality identified in the literature review were the full birth history; the census questions – the population count, the summary birth history, deaths in the household and the preceding birth technique; demographic surveillance and civil registration and vital statistics. The performance of each data source was evaluated using the appropriate methods of analysis including a discussion of the strengths and limitations. Major findings and conclusions of have already been presented at the end of each chapter; hence, this discussion will aim to integrate the previous findings into coherent themes that contribute to answering the research question.

After summarising the findings of Chapters 4 to 8, for ease of comparison, I present a constructed national under-five mortality trend using the 1998 full birth history data, adjusted vital registration deaths and births derived from the 2011 Census. This is compared with estimates produced by international agencies and estimates produced by a local epidemiological and demographic model and the Rapid Mortality Surveillance. The issue of data quality was a prominent crosscutting issue identified throughout the thesis. This is discussed further in section 9.4. The scope of the CRVS for monitoring childhood mortality is discussed in section 9.5. The contribution of the thesis to existing knowledge and areas of further research to strengthen the measurement of childhood mortality are presented before the conclusions.

9.2 Summary of the findings
The investigation of household surveys in Chapter 4 showed that the strength of the full birth history data collected from women of reproductive age lies in their ability to provide age-specific rates of childhood mortality particularly during infancy. The FBH is the optimal instrument for estimating the IMR, early and late neonatal mortality rates and the post-neonatal mortality rate in the absence of complete vital registration. The detailed information collected in the FBH, however, requires well-executed fieldwork in order to ensure reliable estimates. This was the case with South Africa’s first DHS, which yielded high quality data in contrast to the second DHS, where the
under-reporting of births in the five years prior to the survey severely affected the estimates of fertility and child mortality.

The assessment also identified that administering a pregnancy history as opposed to a birth history provides a robust tool for minimising the misclassification of stillbirths as neonatal deaths. Evidence from the 1998 DHS suggests that probing for stillbirths misclassified as live births that experienced early neonatal death improves the accuracy of the IMR, strongly supported by evidence from two recent studies in Mali and Nigeria (Adewemimo, Kalter, Perin et al. 2017; Liu, Kalter, Chu et al. 2016). The assessment of a pregnancy history also generates unique empirical information about the stillbirth rate, which, globally, is rare but also provides important information about maternal health status. None of these indicators can be substituted by any other approach than a fully functioning CRVS.

An important objective of the chapter was to investigate the impact of HIV on the FBH information. The application of two approaches to correct for mother’s survivor bias in the FBH data - one based on prevalence and the other based on incidence - showed that the method using incidence data consistently under-estimates the bias compared with the prevalence method. The take-home message is that because the HIV epidemic invalidates the assumption of no selection bias operating from the reports of mothers on the survival of their children, retrospective childhood mortality data require an adjustment to correct for mother’s survivor bias in South Africa.

In Chapter 5, the performance of the Census questions in South Africa’s three censuses in monitoring the level and trend in childhood mortality was evaluated. The Census offers internal consistency that is not afforded by other data sources. The 1996 Census SBH questionnaire produced national, provincial and population group trends in infant and under-five mortality. Apart from the misclassification of stillbirths reported as live births that died shortly after birth, there is consensus that these were good quality data, and have been extensively used. The quality of the 2001 Census data on the other hand, was found to be so poor that no SBH estimates were possible (Dorrington, Moultrie and Timaeus 2004).

The chapter also reviewed the performance of the census questions in the 2007 Community Survey, namely the analysis of Darikwa (2009) who produced national and provincial infant and under-5 mortality trends by synthesizing the SBH and deaths in the household data. Successful application of the Ward and Zaba (2008) adjustment
for mother’s survivor bias, an adjustment proposed by Darikwa and Dorrington (2011) to allow for changing prevalence over time and a third adjustment for the impact of ART were also demonstrated with these data.

The 2011 Census measurement of the IMR based on the survival status of the last birth produced an underestimate of the mortality in South Africa. Reverse survival of the enumerated population under-fifteen provides an estimate of the trend of births, while the reported deaths in the last 12 months allows for provincial point estimates of infant, child and under-five mortality around mid-2010 to be made. These two data sources also permit the comparison of provincial life-table measures and the calculation of rates using the births. Examination of the differences produced by the two methods can offer insight as to the probable source of error. Utilising the re-released data to construct provincial infant and 1-4 year old mortality trends, it is clear that, nationally, the reported household deaths are acceptable, but provincially there is strong evidence of heaping at age one, producing, in some instances, unrealistic infant and child mortality rates. The deaths in the household data were, however, re-released after correction for extensive missing age and sex, which amounted to about 20 per cent in the original dataset. These errors were apparently due to scanning errors, and the final data were released with approximately 22 per cent less deaths than the original data set, causing concern about the reliability of these data. In any event, aside from the uncertainty due to these procedural scanning errors, examination of the distribution of deaths reported by month of death suggest that recall error is responsible for an over-estimation of overall deaths under the age of five years.

The proposed third adjustment for ART to the SBH data is appropriate; however, after removal of all the inconsistencies, these data produce an implausible mortality trend suggesting the data quality is problematic.

The chapter concludes that migration of children in the first five years of life is an important consideration when studying the mortality of children provincially, however, there are strong indications that the census question for investigating this was also affected by scanning errors, making it difficult to correct for the movement of children.

The continuous demographic surveillance of an open cohort for monitoring childhood mortality was investigated in Chapter 6. It was found that the FBH data together with the episode data replicate the national trend – namely an increase in childhood mortality in the early 1990’s due to the HIV epidemic and the subsequent
decline starting around 2003 due to the PMTCT intervention both occurring slightly earlier in this rural population than other areas in the country.

The chapter highlighted the importance of methodological rigor in measuring child mortality. Interrogating the 2000 to 2007 prospective trend showed that the 61% decline in the neonatal rate, contributing to a 58% overall decline in the IMR is suspect because it coincides with a methodological change which dropped the pregnancy notification form in 2003. An important finding was that the neonatal deaths in particular were missed from the routine surveillance after this change. This hypothesis was explored by modelling the age and sex pattern over time and this showed that the reported decline produced an odd neonatal pattern for boys indicating errors in the reported data.

These data also permit the investigation of the correlates of childhood mortality, although their retrospective nature may require careful interpretation of the socioeconomic information associated with deaths further back in the past because changes in household conditions affecting the older children.

The main criteria used to assess the CRVS, as a tool for monitoring childhood mortality, was the completeness of both birth registration and death registration. It was concluded that there has been impressive improvement in the registration of births and deaths. Nationally in 1996 only 33% of birth registrations occurred by the second birthday and 58% occurred by the fifth birthday. By 2011, 83% of registrations took place by the second birthday and about 90% of registrations occurred by the fifth birthday.

The assessment of the completeness of registered deaths was undertaken with less certainty around the numbers of deaths by age because the available independent estimates post the 1998 DHS are not the ideal instruments to collect age-specific rates; and the reported overall deaths under the age of five in the 2011 Census appeared to have heaping at age one. Nonetheless, the assessment of completeness suggests that the completeness of under-five death registrations has probably declined (slightly) from 2006.

Given the paucity of information about the distribution of $i_{0}$ and $i_{1}$ relative to $i_{0}$, the aim of investigating the age-pattern of deaths over time was to evaluate the coherence of the age-pattern in the CRVS. Comparison of the provincial relationships between $i_{0}$ and $i_{1}$ between 1998 and 2011 against those of the Princeton models show a shift towards higher values of $i_{1}$ relative to $i_{0}$. The data points closer to the
West model at the beginning of the period, however they move towards the North model and beyond in recent years, although some data points revert to ratios of the West. These graphical comparisons highlight the range in the level of mortality and the provincial variation in age-specific rates. Comparison of the observed relationships between $\text{q}_0$ and $\text{q}_1$ with IGME, IHME and the Thembisa models, indicate that the assumptions of the Thembisa model provide the closest fit regarding the age-pattern of mortality.

Establishing an epidemiological expectation in the current AIDS era is challenging due to the age-specific changes caused by the effect of interventions. Assessment of the epidemiological consistency of CRVS data indicates substantial increases in the post-neonatal rate due to HIV/AIDS and corresponding decreases in accordance with the expectation from PMTCT. South Africa displays the compression of the under-five mortality rate into the post-neonatal period, in contrast to the prediction of a compression of the under-five mortality rate into the neonatal period, proposed by Rao and colleagues (2011).

Chapter 8 presented the challenge of monitoring childhood mortality for the provinces from CRVS. The provinces experienced the same pattern of improvement in the completeness of registration as did the country as a whole over time, and the relative differences in the completeness between the provinces narrowed markedly, so that by 2011 the range of completeness in births registered by the second birthday was 74% in Gauteng to 88% in the North West. It was recognised that using reverse survival to estimate provincial completeness of birth registration is challenging because children are not necessarily registered in the province in which they were born.

The analysis revealed that improvement in the provincial mortality since 2006, has more-or-less stagnated. The comparative mortality analysis showed the hazard of using unadjusted mortality ratios as representative of the ‘true’ mortality rate in the context of changing completeness of death registration. The chapter highlights a gap in our understanding around the migration patterns of children that would help in the interpretation of the provincial registration of young children.

9.3 Trends in childhood mortality in South Africa
The thesis reviewed all empirical information pertaining to national and provincial levels of childhood mortality between 1996 and 2011. There were seven occasional opportunities for empirical measurement (census and surveys) during the fifteen years together with the continuous source of vital registration data. The national trend in
estimates produced from these eight data sources is shown in Figure 9-1, where the ‘reconstructed’ trend is compared against those of IGME, IHME, WPP and Thembisa. The constructed trend comprises 25 years of estimates due to the retrospective nature of the 1998 DHS pregnancy histories. These estimates provide a valuable record of South Africa’s child mortality decline during the period when global achievements in child survival were attributed to the effects of the GOBI-FFF policy.

Figure 9-1: Historical trend in under-5 mortality in South Africa 1975-2011, compared with IGME, IHME, WPP and Thembisa 2015 modelled trends and the RMS 2014

IGME, Thembisa estimates and the RMS have the trend most similar to the constructed estimates. The most significant difference between the estimates is at the beginning of the 1990’s in 2005 suggesting the models lacked sufficient information with which to calibrate the impact of HIV on childhood mortality. The models converge after the mortality peaked in 2005, so that by 2011 the estimates range from 58 (WPP), 56 (IHME), 50 (Thembisa and IGME), 45 (the reconstructed estimate) to 40 per 1,000 live births (RMS).

The trend attests to rapid changes in under-5 mortality over time causing two distinct reversals. The first reversal signified by an increase in the rate during the early 1990’s due to HIV/AIDS and the second reversal, by a decrease from around 2005

---

42 My estimates produced from the 1998 DHS (1975-1998) and adjusted VR (2000 – 2011) using the methods and assumptions to derive national and provincial estimates presented in chapters 7 and 8.
due to provision of ART. Dividing the trend into periods defined by the data source responsible for producing the estimate, the following observations can be made:

1) For the period 1975 to 1998, there was good corroboration between the 1996 SBH and 1998 FBH to produce under-5 mortality rates. Specifically the attributes of the FBH permitting age-specific rates and due to the retrospective nature allow for the 25-year trend although the points further back become less reliable. Downward trends, probably associated with primary health care initiatives and declines in fertility rates, were reversed in the 1990’s by the AIDS epidemic and vertical transmission of HIV from mother-to-child.

2) The 2001 Census failed to produce estimates from the SBH and the 2003 DHS failed to produce estimates from the FBH, in other words, both enquiries, which should have served this period, failed in their intended purpose. Estimates of $q_0$ and of $s_0$ were derived from deaths in the household from Census 2001, although these data are of questionable quality. The observed increases in childhood mortality due to mother-to-child transmission peaked between 2003 and 2005.

3) The 2007 Community Survey produced a retrospective trend from 1997 to 2006 of $i_q$ and $s_q$ by synthesizing estimates from the SBH and household deaths. At the beginning of this period, the second reversal in mortality is observed. By 2011, despite the devastating effects of the HIV/AIDS epidemic, child mortality declined to levels below those in the early stages of the epidemic. The interventions to reduce mother-to-child transmission, diarrhoeal disease and pneumonia in children have all played a role in the prevention of child deaths, although the latter two interventions were instituted around 2010.

4) The trend in Figure 9-1 uses these mortality estimates to derive correction factors to adjust the VR data between 1998 and 2006.

Chapter 5 concluded that the proposed adjustments to the 2011 Census SBH estimates to account for changes in the mortality of mothers introduced by ART are appropriate for inclusion in a revised methodology of the SBH. The death in the household question was used to estimate completeness of under-1 and under-5 death registrations in the VR deaths in 2011. Linear interpolation was used to estimate completeness of deaths between 2006 and 2011. The accuracy of these data are
compromised by heaping of deaths at age one, and although the level of under-5 mortality appears reasonable, there is uncertainty around the distribution of infant and child deaths. The census population count allowed a series of births by province to be derived allowing mortality rates to be calculated.

The 2012 HSRC survey also administered a FBH; however, these data were not reported. During this time, the only available estimates were derived from models, which included long periods with poor agreement between the various models primarily due to lack of empirical data for purposes of calibration.

Despite the poor quality of data, the robust nature of the methods allow for an array of applications to reduce the uncertainty around the estimates such as the synthesis of estimates (Pillay-Van Wyk, Laubscher, MSEMBURI et al. 2014) however, regardless of innovation, there is no substitute for empirical estimates of early and late neonatal mortality and post-neonatal mortality. The questionable quality of some of the data sets also compromise the accuracy of estimates provincially, and disaggregation below national level increases the uncertainty around the estimates.

9.4 The pervasive challenge of data quality
The thesis has identified poor data quality as a crosscutting issue in all the data sets subsequent to the 1996 Census and 1998 DHS. Chapters 4 and 5 identified coverage and content errors as well as a wide range of other problems affecting the quality of information from the census and DHS questionnaires. These included birth transference to minimise the workload of fieldworkers, badly designed questions in the case of the Community Survey household deaths affecting the reference period.
However, the most important demographic information for mortality analysis is date of birth and date of death. In the case of children, the accurate recording of these details has more bearing on the accuracy of the information than at older ages, a fact that should be emphasised more during field-worker training.

Data quality is an obvious concern in all data collection exercises, however, the ubiquitous and generic nature of the range of errors warrants concerted effort to improve fieldwork training. The extent to which missing information has featured as an underlying reason for problematic results is beyond comprehension, as is the extent to which data were of such poor quality that these were deemed unusable and were simply never reported (SBH in 2011 Census and FBH in the HSRC 2012). The reluctance of agencies to review the quality of information, to learn from the experience, and implement quality control measures to minimise errors is a concern to
public policy. The errors ostensibly because of scanning a dirty back page containing the mortality schedule in 2001 occurred again in 2011 (Stats SA 2014a) because the provision of a protective cover sheet in the 2011 Census was not considered. It is problematic that after the error caused 2.5 million false mortality records in the 2001 (Dorrington, Moultrie and Timaeus 2004) that the same error was repeated in the following census.

Field-worker training appears to be central to many of the issues identified and this can be improved by, for example, putting in place and monitoring the quality-assurance measures pointed out in Chapters 4 and 5, in addition to the management teams of the census and surveys also needing a degree of scrutiny or supervision. Following the countless texts on data collection procedures is surely a good place to start. The challenge therefore must be to institute mechanisms of assuring high standards of data capture and processing, for example by ensuring strong managerial support from well-trained supervisors in addition to the practice of data editing simultaneous to the initial fieldwork enables a feedback mechanism and should safeguard the accuracy of these data. The role of census and DHS’s to provide periodic calibration for CRVS is essential.

Lastly, perhaps some sort oversight from the research community in the training process of fieldworkers would be a valuable contribution. The poor quality of data are unacceptable to the public purse and South Africa has to do better.

9.5 **The scope of CRVS for monitoring childhood mortality**

In settings where CRVS are deficient, it will always be necessary for censuses, surveys and DSS’s to provide empirical data that can substitute elements of information where CRVS is incomplete. While it is clear that the last birth technique understates mortality and that, the estimates based on household deaths are prone to heaping on age one, the SBH requires further development for South African conditions to account for changes in the assumptions of mothers’ survivor bias. The propensity for children to be under-counted in censuses has resulted in the derivation of the number of births from the population count being under-utilised; however, there is no evidence of an under-count of children in the 2011 data. These estimates of births have enriched the empirical sources of this important indicator, as does the provision of age-specific fertility rates from the FBH and the CEB and born in the last 12 months from censuses and surveys for purposes of triangulation.
The strengths of the FBH data lie in their design to provide fine age-group specific mortality rates in early childhood. The ability to link these data to key covariate-indicators such as the mother, household, area, health and service indicators provides detailed information about the determinants of childhood mortality, is an important dimension of these data. Although appropriate methods to account for mother’s survivor bias in the FBH and the SBH are available, in the era of treatment, the challenge is to understand how the correction for mother’s survivor bias is affected by the impact of ART and PMTCT treatments.

In the case of South Africa, surveillance data from selected populations can provide useful benchmarks, but cannot provide nationally or provincially representative indicators. These data do, however, provide opportunity to detect changes over short periods and due to the relatively short follow-up time are well suited to study seasonal changes in childhood mortality. The verbal autopsy information has been used to correct specific causes of death known to be under-reported in the South African CRVS (Joubert, Bradshaw, Kabudula et al. 2014). Similarly Hallett et al. used HDSS data to determine the correction factors needed to adjust for mother’s survivor bias in Zimbabwe (Hallett, Gregson, Kurwa et al. 2010).

Chapter 7 documented the marked improvement in the completeness of birth registration and in the completeness of death registration, showing that by 2012 South Africa is on the cusp of being able to use unadjusted CRVS to monitor childhood mortality. Chapter 8 showed that it is more challenging to utilise CRVS as a monitoring tool provincially, because, as we have seen, with the possible exception of the Western Cape, adjustments to both the numerator and the denominator are necessary. The provincial ranking of the completeness of births and deaths was in some cases surprising, and this attests to the fact that it is important to monitor trends in basic demographic indicators on an ongoing basis. The magnitudes of the improvements have to some users, suggested (COMMIC 2014; Department of Health 2011a) that birth and death registration data can be used without adjustment. Comparison of the unadjusted ratios with the estimated mortality trends demonstrate how misleading and unrepresentative the unadjusted provincial ratios are, hence this practice needs to be discouraged.

The assessment of completeness by age is a robust method for monitoring purposes going forward as it allows for the inclusion of late registrations, by updating the tabulations (produced by Stats SA) of births by year of birth and year of
registration. Completeness is sensitive to the length of the data-capture interval of the reported numbers of events annually (timeliness) and the consistent practice by the statistical agency responsible for collation of statistics is critical. The chapter also noted the need for clarification of terms and definitions across agencies and ministries within countries.

Periodic review of available data to monitor progress towards improving CRVS is also necessary. For example, UNICEF’s Multiple Indicator Cluster Survey presently designed to estimate the proportion of births registered by the fifth birthday, would be more versatile if it aimed to capture births registered by the first or second birthday.

Findings from both the chapters assessing the CRVS in particular have important lessons for other countries of similar income and development level as South Africa. The assessment of the completeness of birth and death registration, found that derivation of independent estimates of the recent trends in births and deaths was critical to the evaluation of completeness, rather than to utilise estimates produced by statistical models. Certainly, the take home message from the thesis is that one should investigate all available data sources suitable for triangulating numbers of births and deaths. We saw the strength of the reverse survival in producing estimates of the number of births; however, the need for triangulation with other data sources was crucial to verify an unexpected trend. By all accounts, the 2011 South African Census appears not to have undercounted children under fifteen years old.

The Sustainable Development Goal project (Mikkelsen, Phillips, AbouZahr et al. 2015) calls for the assessment of CRVS at a sub-national level to ensure that no one is left behind because the monitoring of equity is central to the project. Chapter 8 highlighted the wide differences between the nine provinces in terms of levels of fertility and mortality as well as differences in civil registration. It is clear that South Africa must strengthen provincial CRVS systems, in order to highlight geographic areas with particular challenges around improving coverage of civil registration. The setting up of Home Affairs mobile offices in large hospitals must be a contributing factor to the increased proportions of birth registrations since 2000. The coincidence of the implementation of the Child Support Grant and increased birth registration is highly suggestive that the grant played a role in improving birth registration. These two interventions present possible mechanisms for improvement that other countries may consider exploring under their own conditions. Last, other countries with high-mobility populations may find it would useful to investigate the
tabulation of vital statistics and the enumerated population by usual place of residence as well as by place of occurrence.

9.6 Contributions to knowledge
This thesis reviewed the literature and developed an analytical framework to investigate the data sources and methods of analysis available to monitor childhood mortality. The framework developed by Graham and colleagues (Graham, Ahmed, Stanton et al. 2008), for monitoring maternal mortality was adapted for child mortality. It includes the routine data sources such as CRVS, health service data and census as well as special data sources such as surveys and longitudinal studies. The methods of demographic analysis for the different data types are included in the framework as well as the step of synthesising data using demographic projection approaches and statistical approaches. Although the framework was developed for the case study of South Africa, much of the content is generic and would be useful for other countries to map out their specific data sources and analytical approaches for monitoring childhood mortality.

The review of literature demonstrated the fact that although South Africa is a data abundant country, for long periods, either reliable estimates of childhood mortality were completely lacking, or there was considerable uncertainty around the estimates. Investigating the four primary data sources for producing childhood mortality indicators the thesis has documented the strengths and weaknesses of each data source and highlighted the methodological challenges related to these data. The analyses have highlighted that the accurate reporting of early neonatal deaths requires attention to the design of questionnaires. In the case of the DHS, the pregnancy history questionnaire includes questions which probe the respondent’s answers in order to illicit better information regarding the age at death, and therefore aids in distinguishing still births from early neonatal deaths. Similarly, the pregnancy notification form in the demographic surveillance area, serves to alert the fieldworker for the need to ask follow-up questions about the pregnancy. The decrease in the number of reported neonatal deaths coincided with a methodological change whereby the pregnancy notification form was no longer administered. Hence, significantly fewer neonatal deaths were reported in the second half of the period, compared with the first half of the period, suggesting demographic surveillance sites miss early neonatal deaths.
In reviewing the outcome of the data collection processes, the thesis unearthed a worrying record of poor or extremely poor data quality. This suggests a need to strengthen oversight and supervision of fieldworkers and support staff in the field during censuses and surveys in the future. Developing mechanisms to communicate these findings to forums of policy formulation and the research community are needed. A more collaborative approach in the design of the questionnaires and the training of field workers, as is more common in qualitative research, would benefit the process.

9.7 Future research directions
As global efforts to assess completeness of birth registration for individual countries intensify, the advantages of empirical estimates of births are clear. Investigation into the performance of reverse survival to derive estimates of births from recent census data appears a worthwhile start in the pursuit of these data. Administrative data such as the school attendance data played an important role in triangulating the age distribution of the population in order to verify the estimated births.

South Africa would benefit from research into why survey and census field work are not adequate, and how field work can be improved.

Some progress was made in this thesis around the assessment of the age pattern and the link between age and cause of death. More work in this area could be fruitful.

Despite the marked improvement in the CRVS systems, persistent under-registration of births and deaths in some provinces requires investigation into common practices and procedures in those provinces would be provide insight about the obstacles, which communities are most affected and would inform interventions designed to aid these sections of the community. Similarly, investigation into the excess registration of deaths in some provinces is also needed.

Further investigation is needed to assess the performance of linkage of the birth and death register information. To date, these data have only been investigated separately. Linkage could be done through the National Population Register. However, this would necessitate an amendment to the registration system to include the registration of individuals whose birth had not been registered before death.

The Lancet 2015 series on counting births and deaths points out that collaboration is essential for the success of the CRVS project. Collaboration between countries and the international agencies driving the CRVS agenda to develop the competencies of demographic analysis required to monitor births and deaths will be
beneficial, as will exchanges about the lessons learnt. The establishment of reference groups of interested parties with a research agenda would be a useful mechanism to foster research to strengthen CRVS.

The very important finding that in the highly mobile rural setting children living away from their mothers have an increased mortality risk should be further investigated to understand the social, economic and cultural dynamics of this phenomenon.

9.8 Conclusions
A critical appraisal of South Africa’s experience in the measurement of childhood mortality over the last twenty years has facilitated a detailed review of all the evidence to answer the question of how childhood mortality can be best monitored in upper-middle income countries with limited resources. This thesis has shown that although South Africa is a data abundant country, there is a dearth of key demographic and health indicators, particularly at sub-national level.

This thesis has produced convincing evidence to support the recommendation that strengthening the CRVS system to further improve the registration of births and deaths is the best way forward for monitoring childhood mortality provincially and is sustainable. A DHS would be required to benchmark neonatal, post-neonatal, infant and child mortality rates provincially, every five years. The Rapid Mortality Surveillance project attests to the fact that nationally, vital registration is robust enough to monitor accurately infant and under-5 mortality with minimal adjustment, however, the dearth of age-specific rates within infancy is severely debilitating and can only be resolved with a DHS. Without a DHS, there would be too much uncertainty associated with monitoring these indices provincially. This is the future challenge, and the 2016 SADHS provides opportunity to benchmark age-specific mortality on a provincial basis and going forward.
REFERENCES


Department of Women Children and Persons with Disabilities. 2012.


Fottrell, F, P Byass and Y Berhane. 2008. "Demonstrating the robustness of population surveillance data: implications for error rates on demographic and mortality estimates", BioMed Central: Medical Research Methodology 8(13)


Demographic and Health Survey and from a Demographic Surveillance System", *Acta Tropica* **98**:212-218.


Nhacolo, AQ, DA Nhalungo, CN Sacoor, JJ Aponte et al. 2006. "Levels and trends of demographic indices in southern rural Mozambique: evidence from demographic surveillance in Manhica district", BMC public health 6(291)


STATA. 2007. *Stata Release 10*. 4905 Lakeway Drive College Station, Texas 77845-4512, USA: StataCorp LP.


Appendix 1: Extract of pregnancy history questionnaire, SADHS 1998 (Chapter3)

Now I would like to ask you about all of your pregnancies, whether born alive, born dead, or lost before full term, starting with the first one you had.

RECORD ALL THE PREGNANCIES. RECORD TWINS AND TRIPLETS ON SEPARATE LINES.

<table>
<thead>
<tr>
<th>214</th>
<th>215</th>
<th>216</th>
<th>217</th>
<th>218</th>
<th>219</th>
<th>220</th>
<th>221</th>
</tr>
</thead>
<tbody>
<tr>
<td>Think back to the time of your (first/last) pregnancy.</td>
<td>Was that a single or multiple pregnancy?</td>
<td>Was the baby born alive, born dead, or lost before full term?</td>
<td>Did the baby cry, move, or breathe when it was born?</td>
<td>What was the name given to that child?</td>
<td>Is (NAME) a boy or a girl?</td>
<td>In what month and year was (NAME) born?</td>
<td>Is (NAME) still alive?</td>
</tr>
</tbody>
</table>

01 SINGLE ... 1 MULTIPLE 2
- BORN ALIVE ................ 1 YES ... 1 BOY ... 1 MONTH ... 1 YEAR ... 1 NO ... 2 GIRL ... 2 YEAR ... 2
- BORN DEAD ................... 2 NO ... 2
- LOST BEFORE FULL TERM 3 (SKIP TO 225)

02 SINGLE ... 1 MULTIPLE 2
- BORN ALIVE ................ 1 YES ... 1 BOY ... 1 MONTH ... 1 YEAR ... 1 NO ... 2 GIRL ... 2 YEAR ... 2
- BORN DEAD ................... 2 NO ... 2
- LOST BEFORE FULL TERM 3 (SKIP TO 225)

IF BORN ALIVE AND STILL LIVING:

IF BORN ALIVE BUT NOW DEAD:

IF BORN DEAD OR LOST BEFORE FULL TERM:

222
- How old was (NAME) at his/her last birthday? RECORD AGE IN COMPLETED YEARS

223
- Is (NAME) living with you?

224
- How old was (NAME) when he/she died? IF 1 YR., PROBE: How many months did he/she die from diarrhea?

224A
- Did (NAME) die from diarrhea?

225
- In what year and month did this pregnancy end?

226
- How many months did the pregnancy last? RECORD IN COMPLETED MONTHS.

228
- From year of this pregnancy subtract year of previous pregnancy. Is the difference 2 or more years?

229
- Were there any other pregnancies between the previous pregnancy mentioned and this pregnancy?

01 AGE IN YEARS
- YES ... 1 NO ... 2 NEXT PREG.

02 AGE IN YEARS
- YES ... 1 NO ... 2 NEXT PREG.
Appendix 2: Fertility questions, 2011 Census (Chapter 3)
Appendix 3: STATA programme for the calculation of under-five mortality component death probabilities (Chapter 3)

/*set output error
global varparms "*"
capture program drop childmor
program define childmor
version 6.0
syntax [varlist]*/
gen failure = V008-b3
replace failure = failure-1 if V016<16 & failure>0
drop if failure>239
gen bcohort=mod(failure,12)
label var bcohort "years since birth"
label def period 0 "1" 1 "2" 2 "3" 3 "4" 4 "5" 5 "6" 6 "7" 7 "8" 8 "9" 9 "10" 10 "11" 11
"12" 12 "13" 13 "14" 14 "15" 15 "16" 16 "17" 17 "18" 18 "19" 19 "20"
label value bcohort period
replace failure = b7 if b5==0
replace failure = failure+0.5
gen idnum = bidx + V003*100 + V002*10000 + V001*10000000
stset failure, f(b5==0) id(idnum)
tab failure bcohort
keep $varparms b5 bcohort wwhr1 V005 WostdWht failure
number _st _d _t t0
stsplitage_gp, at(1 3 6 12 24 36 60 120)
drop if age_gp==120
gen expos=age_gp
replace expos=0.083333 if expos==0
replace expos=0.166667 if expos==1
replace expos=0.25 if expos==3
replace expos=0.50 if expos==6
replace expos=1 if (expos==12|expos==24)
replace expos=2 if expos==36
replace expos=5 if expos==60
*recode expos 0=0.083333 1=0.166667 3=0.25 6=0.5 12/24=1 36=2 60=5
replace expos=expos/2 if b5==.
gendth=0
replacedth=1 if b5==0
gen weight=V005/1000000
collapse (sum) deaths=dthpyrs=expos [pweight=WoStdWht],
by(age_gpbcohort $varparms)
labvarage_gp "age of child"
labdefagegp 0 "0 mth" 1 "1-2 mths" 3 "3-5 mths" 6 "6-11 mths" 12 "1 year" 24
"2 years" 36 "3-4 years" 60 "4-5 years"
labvalage_gp
*end
list
*quietly childmor $varparms
Appendix 4: Deaths in the household in the past 12 months question, Census 2011 (Chapter 3)
Appendix 5: Cohort life table of all children up to age five, January 1st 2000 to December 31st 2007 (Chapter 6)

<table>
<thead>
<tr>
<th>Age Group</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;27 days</td>
<td>47,047</td>
<td>48,715</td>
<td>48,051</td>
<td>46,160</td>
<td>49,168</td>
<td>48,793</td>
<td>46,772</td>
<td>46,196</td>
</tr>
<tr>
<td>1-11 months</td>
<td>577,920</td>
<td>599,869</td>
<td>4</td>
<td>579,414</td>
<td>9</td>
<td>0</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>12-23 months</td>
<td>640,894</td>
<td>669,428</td>
<td>1</td>
<td>690,420</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>24-35 months</td>
<td>682,780</td>
<td>688,664</td>
<td>9</td>
<td>704,762</td>
<td>1</td>
<td>6</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>36-47 months</td>
<td>718,855</td>
<td>728,137</td>
<td>724,32</td>
<td>720,17</td>
<td>711,92</td>
<td>699,05</td>
<td>730,61</td>
<td>710,52</td>
</tr>
<tr>
<td>48-59 months</td>
<td>701,042</td>
<td>761,290</td>
<td>2</td>
<td>737,221</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Person-years exposure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1579.7</td>
<td>6</td>
<td>1517.59</td>
<td>8</td>
<td>1561.4</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>1802.3</td>
<td>4</td>
<td>1792.88</td>
<td>7</td>
<td>1761.2</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>1903.2</td>
<td>1</td>
<td>1834.05</td>
<td>7</td>
<td>1808.5</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>1962.5</td>
<td>2</td>
<td>1886.75</td>
<td>9</td>
<td>1928.2</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1984.4</td>
<td>7</td>
<td>1994.90</td>
<td>7</td>
<td>1973.0</td>
<td>7</td>
<td>7</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>2108.0</td>
<td>5</td>
<td>2085.73</td>
<td>7</td>
<td>2024.5</td>
<td>5</td>
<td>1</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>2192.66</td>
<td>1</td>
<td>2019.78</td>
<td>3</td>
<td>2197.4</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

Probability of dying per 1,000 live-births

| NN | 13.49 | 9.94 | 17.57 | 20.22 | 5.55 | 6.21 | 4.54 | 5.25 |
| PNN | 55.16 | 51.61 | 46.60 | 45.70 | 35.14 | 34.10 | 27.25 | 22.75 |
| $q_0$ | 66.34 | 59.69 | 62.15 | 63.79 | 39.87 | 39.51 | 31.30 | 27.61 |
| $d_0$ | 30.70 | 41.19 | 44.84 | 37.92 | 32.84 | 31.75 | 27.03 | 24.89 |
| $s_0$ | 95.01 | 98.43 | 104.21 | 99.29 | 71.41 | 70.00 | 57.48 | 51.81 |
Appendix 6: Coefficients of the Model 6 (age*year*sex) (Chapter 6)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Girl 1-11 months</td>
<td>1.00</td>
<td>2.93</td>
<td>1.56</td>
</tr>
<tr>
<td>Girl &lt; 28 days</td>
<td>1.074</td>
<td>2.93</td>
<td>0.06</td>
</tr>
<tr>
<td>Girl 12-59 months</td>
<td>-2.378</td>
<td>0.09</td>
<td>0.14</td>
</tr>
<tr>
<td>Boy &lt; 28 days</td>
<td>0.605</td>
<td>1.83</td>
<td>0.87</td>
</tr>
<tr>
<td>Boy 1-11 months</td>
<td>-0.123</td>
<td>0.88</td>
<td>0.59</td>
</tr>
<tr>
<td>Boy 12-59 months</td>
<td>-1.978</td>
<td>0.14</td>
<td>0.09</td>
</tr>
</tbody>
</table>

**Year**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>1.00</td>
<td>2001</td>
<td>-0.059</td>
<td>0.94</td>
</tr>
<tr>
<td>2002</td>
<td>-0.542</td>
<td>0.58</td>
<td>0.37</td>
<td>0.91</td>
</tr>
<tr>
<td>2003</td>
<td>-0.231</td>
<td>0.79</td>
<td>0.53</td>
<td>1.20</td>
</tr>
<tr>
<td>2004</td>
<td>-0.484</td>
<td>0.62</td>
<td>0.40</td>
<td>0.96</td>
</tr>
<tr>
<td>2005</td>
<td>-0.453</td>
<td>0.64</td>
<td>0.42</td>
<td>0.97</td>
</tr>
<tr>
<td>2006</td>
<td>-0.743</td>
<td>0.48</td>
<td>0.30</td>
<td>0.77</td>
</tr>
<tr>
<td>2007</td>
<td>-1.038</td>
<td>0.35</td>
<td>0.21</td>
<td>0.59</td>
</tr>
</tbody>
</table>

**Girls <28 days**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>-0.870</td>
<td>0.42</td>
<td>0.14</td>
<td>1.27</td>
</tr>
<tr>
<td>2002</td>
<td>0.401</td>
<td>1.49</td>
<td>0.59</td>
<td>3.79</td>
</tr>
<tr>
<td>2003</td>
<td>0.048</td>
<td>1.05</td>
<td>0.41</td>
<td>2.67</td>
</tr>
<tr>
<td>2004</td>
<td>-0.459</td>
<td>0.63</td>
<td>0.20</td>
<td>1.96</td>
</tr>
<tr>
<td>2005</td>
<td>-0.990</td>
<td>0.37</td>
<td>0.10</td>
<td>1.41</td>
</tr>
<tr>
<td>2006</td>
<td>-0.678</td>
<td>0.51</td>
<td>0.13</td>
<td>1.96</td>
</tr>
<tr>
<td>2007</td>
<td>-0.750</td>
<td>0.47</td>
<td>0.10</td>
<td>2.30</td>
</tr>
</tbody>
</table>

**Girls 12-59 months**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.601</td>
<td>1.82</td>
<td>0.96</td>
<td>3.47</td>
</tr>
<tr>
<td>2002</td>
<td>1.130</td>
<td>3.10</td>
<td>1.58</td>
<td>6.07</td>
</tr>
<tr>
<td>2003</td>
<td>0.728</td>
<td>2.07</td>
<td>1.07</td>
<td>4.00</td>
</tr>
<tr>
<td>2004</td>
<td>0.835</td>
<td>2.30</td>
<td>1.15</td>
<td>4.60</td>
</tr>
<tr>
<td>2005</td>
<td>0.723</td>
<td>2.06</td>
<td>1.03</td>
<td>4.11</td>
</tr>
<tr>
<td>2006</td>
<td>0.832</td>
<td>2.30</td>
<td>1.10</td>
<td>4.80</td>
</tr>
<tr>
<td>2007</td>
<td>0.839</td>
<td>2.31</td>
<td>1.05</td>
<td>5.10</td>
</tr>
</tbody>
</table>

**Boys <28 days**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.360</td>
<td>1.43</td>
<td>0.53</td>
<td>3.86</td>
</tr>
<tr>
<td>2002</td>
<td>1.307</td>
<td>3.70</td>
<td>1.43</td>
<td>9.56</td>
</tr>
<tr>
<td>2003</td>
<td>1.233</td>
<td>3.43</td>
<td>1.38</td>
<td>8.54</td>
</tr>
<tr>
<td>2004</td>
<td>-0.230</td>
<td>0.79</td>
<td>0.22</td>
<td>2.85</td>
</tr>
<tr>
<td>2005</td>
<td>0.303</td>
<td>1.35</td>
<td>0.45</td>
<td>4.07</td>
</tr>
<tr>
<td>2006</td>
<td>0.094</td>
<td>1.10</td>
<td>0.30</td>
<td>4.00</td>
</tr>
<tr>
<td>2007</td>
<td>0.781</td>
<td>2.18</td>
<td>0.67</td>
<td>7.09</td>
</tr>
</tbody>
</table>

**Boys 1-11 months**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>-0.018</td>
<td>0.98</td>
<td>0.56</td>
<td>1.73</td>
</tr>
<tr>
<td>2002</td>
<td>0.661</td>
<td>1.94</td>
<td>1.07</td>
<td>3.50</td>
</tr>
<tr>
<td>2003</td>
<td>0.079</td>
<td>1.08</td>
<td>0.60</td>
<td>1.95</td>
</tr>
<tr>
<td>2004</td>
<td>0.050</td>
<td>1.05</td>
<td>0.56</td>
<td>1.98</td>
</tr>
<tr>
<td>2005</td>
<td>-0.085</td>
<td>0.92</td>
<td>0.49</td>
<td>1.72</td>
</tr>
<tr>
<td>2006</td>
<td>0.051</td>
<td>1.05</td>
<td>0.53</td>
<td>2.09</td>
</tr>
<tr>
<td>2007</td>
<td>0.269</td>
<td>1.31</td>
<td>0.64</td>
<td>2.69</td>
</tr>
</tbody>
</table>

**Boys 12-59 months**

<table>
<thead>
<tr>
<th>Year</th>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.157</td>
<td>1.17</td>
<td>0.64</td>
<td>2.13</td>
</tr>
<tr>
<td>2002</td>
<td>0.791</td>
<td>2.20</td>
<td>1.18</td>
<td>4.13</td>
</tr>
<tr>
<td>2003</td>
<td>0.235</td>
<td>1.27</td>
<td>0.68</td>
<td>2.36</td>
</tr>
<tr>
<td>2004</td>
<td>0.313</td>
<td>1.37</td>
<td>0.71</td>
<td>2.64</td>
</tr>
<tr>
<td>2005</td>
<td>0.325</td>
<td>1.38</td>
<td>0.73</td>
<td>2.64</td>
</tr>
<tr>
<td>2006</td>
<td>0.526</td>
<td>1.69</td>
<td>0.85</td>
<td>3.36</td>
</tr>
<tr>
<td>2007</td>
<td>0.874</td>
<td>2.40</td>
<td>1.18</td>
<td>4.87</td>
</tr>
</tbody>
</table>

**Constant**

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Relative Risk</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.722</td>
<td>1.00</td>
<td>2.93</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Appendix 7: Death Notification Form (1663) (Chapter 7)

<table>
<thead>
<tr>
<th>A. PARTICULARS OF THE DECEASED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Name of deceased (full name)</td>
</tr>
<tr>
<td>2. Identification of deceased (date of birth)</td>
</tr>
<tr>
<td>3. Residence of deceased (place)</td>
</tr>
<tr>
<td>4. Date of death (day, month, year)</td>
</tr>
<tr>
<td>5. Cause of death (disease or injury)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. ORIGIN OF DECEASED</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Race (White, Black, Indian, Asian, Other)</td>
</tr>
<tr>
<td>2. Sex (Male, Female)</td>
</tr>
<tr>
<td>3. Marital status (Single, Married, Divorced, Widowed)</td>
</tr>
<tr>
<td>4. Place of birth (country, city/town)</td>
</tr>
<tr>
<td>5. Place of residence (country, city/town)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. OCCUPATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Occupation (nature of work)</td>
</tr>
<tr>
<td>2. Skills or qualifications (education, training)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D. OTHER INFORMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Dates of employment (start, end)</td>
</tr>
<tr>
<td>2. Other relevant details (medical condition, disabilities)</td>
</tr>
</tbody>
</table>

---

*Note: All information must be filled in accurately. Any discrepancies or omissions may result in penalties.*
Death Notification Form page 2
Death Notification Form page 3
Appendix 8: Table of implied provincial completeness of registered deaths under age five (Chapter 8)

Implied provincial completeness of registered deaths under age five measured against estimates derived from the 2007 Community Survey and the 2011 Census

<table>
<thead>
<tr>
<th>Province</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Cape</td>
<td>45.5</td>
<td>47.5</td>
<td>49.4</td>
<td>51.4</td>
<td>53.3</td>
<td>52.9</td>
</tr>
<tr>
<td>Free State</td>
<td>120.9</td>
<td>118.4</td>
<td>115.9</td>
<td>113.5</td>
<td>111.0</td>
<td>110.5</td>
</tr>
<tr>
<td>Gauteng</td>
<td>103.4</td>
<td>105.2</td>
<td>107.0</td>
<td>108.8</td>
<td>110.6</td>
<td>111.0</td>
</tr>
<tr>
<td>KwaZulu-Natal</td>
<td>63.8</td>
<td>63.4</td>
<td>63.0</td>
<td>62.5</td>
<td>62.1</td>
<td>62.0</td>
</tr>
<tr>
<td>Limpopo</td>
<td>75.0</td>
<td>76.1</td>
<td>77.2</td>
<td>78.3</td>
<td>79.4</td>
<td>79.6</td>
</tr>
<tr>
<td>Mpumalanga</td>
<td>72.1</td>
<td>69.8</td>
<td>67.5</td>
<td>65.3</td>
<td>63.0</td>
<td>63.5</td>
</tr>
<tr>
<td>Northern Cape</td>
<td>90.5</td>
<td>92.0</td>
<td>93.5</td>
<td>95.0</td>
<td>96.5</td>
<td>96.1</td>
</tr>
<tr>
<td>North West</td>
<td>121.7</td>
<td>114.8</td>
<td>107.8</td>
<td>100.9</td>
<td>93.9</td>
<td>92.3</td>
</tr>
<tr>
<td>Western Cape</td>
<td>78.2</td>
<td>80.2</td>
<td>82.2</td>
<td>84.1</td>
<td>86.1</td>
<td>85.7</td>
</tr>
<tr>
<td>South Africa</td>
<td>76.2</td>
<td>75.4</td>
<td>74.5</td>
<td>73.7</td>
<td>72.8</td>
<td>72.0</td>
</tr>
</tbody>
</table>
Appendix 9: Comparison of estimated under-5 mortality and unadjusted VR ratio against the NBD study, by province (Chapter 8)