Enumerating and estimating maternal and neonatal deaths in the Western Cape Province, South Africa

Damian Jeremia Damian

Thesis presented for the degree of Doctor of Philosophy in the School of Public Health and Family Medicine, University of Cape Town

July 2021

TABLE OF CONTENTS

Contents page

LIST OF TABLES

LIST OF FIGURES

PREFACE

This thesis comprises six results-based chapters written in a manuscript-format potential for publication in peer-review journals. Data analysis and drafting of these manuscripts were done by the candidate during the doctoral degree registration period under the guidance of his doctoral supervisors. Two manuscripts, a systematic review protocol and the actual review for the first objective, as part of this work, have been published and are referenced in a preface to the relevant chapter. Tables and figures are numbered for the whole document, whereas the references are chapter specific. The following are the results chapters presented in the thesis in manuscript format: -

1. Trends in maternal and neonatal mortality in South Africa: A systematic review.
2. Validating the Provincial deterministic data linkage approach through a fully probabilistic record linkage method applied to mortality and pregnancy data, Western Cape, South Africa.
3. A three-source capture-recapture estimate of maternal deaths under-reporting in the Western Cape Province, South Africa.
4. Estimation of neonatal deaths under-reporting using capture-recapture method in the Western Cape Province, South Africa.
5. Factors associated with maternal mortality in the Western Cape province, South Africa.
6. Factors associated with neonatal mortality in the Western Cape province, South Africa.

DECLARATION

This thesis is presented in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in the School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town. The work on which this thesis is based is original research and has not, in whole or in part, been submitted for another degree at this or any other university. The contents of this thesis are entirely the work of the candidate, or in the case of multi-authored published papers, constitutes work for which the candidate was the lead author. I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publications:

1. **Damian, D. J.,** Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review. *Syst. Rev.* (2019).

2. **Damian, D. J.,** Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review protocol. *Syst. Rev.* **6**, 165 (2017).

SIGNATURE: Signed by candidate **DATE:** 10th July 2020

STUDENT NAME: Damian Jeremia Damian **STUDENT NUMBER:** DMNDAM001

ACRONYMS AND ABBREVIATIONS

AIC	Akaike Information Criterion
AIDS	Acquired Immune Deficiency Syndrome
ANC	Antenatal Clinic
ART	Antiretroviral Therapy
BIC	Bayesian Information Criterion
CEMD	Confidential Enquiry into Maternal Deaths
CI	Confidence Interval
CRC	Capture-recapture
CHIP	Child Problem Identification Program
DHA	Department of Home Affairs
DHS	Demographic Health Survey
DOH	Department of Health
DSS	Demographic Surveillance System
HIV	Human Immunodeficiency Virus
IMMR	Institutional Maternal Mortality Ratio
LMIC	Low- and Middle-Income Countries
MDG	Millennium Development Goal
MICS	Multiple Indicator Cluster Survey
MMEIG	Maternal Mortality Estimation Inter-agency Group
MMR	Maternal Mortality Ratio
MNH	Maternal and Neonatal Health

NCCEMD	National Committee for Confidential Enquiry into Maternal Deaths
NCEMD	National Confidential Enquiry into Maternal Deaths
NIH	National Institute of Health
NMR	Neonatal Mortality Ratio
PHDC	Provincial Health Data Centre
PMI	Patient Master Index
PMTCT	Prevention of Mother-to-Child Transmission
PPIP	Perinatal Problem Identification Programme
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analyses.
RAMOS	Reproductive-Age Mortality Studies
SA	South Africa
SAVVY	Sample Vital Registration with Verbal Autopsy
SSA	Sub-Saharan Africa
TAG	Technical Advisory Group
TB	Tuberculosis
UCT	University of Cape Town
UI	Uncertainty Interval
UN-IGME	United Nations Inter-agency Group for Child Mortality Estimation
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund
WCDOH	Western Cape Department of Health
WHO	World Health Organisation

DEFINITION OF TERMS

Akaike Information Criterion (AIC) – the measure of model performance that account for model complexity and number of estimated parameters i.e., AIC=2k-2ln($\hat{L}$).

Deterministic/exact record linkage - the use of unique identifier or exact matching of defined identifiers (such as names, dates of birth, addresses etc.) to link records from different datasets.

Fuzzy matching - is a special case of record linkage using computer-assisted translation to process word-based *matching* queries when finding correspondences between segments of a text and entries in a database.

Maternal deaths - the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to, or aggravated by the pregnancy, or its management ,but not from accidental or incidental causes.

Maternal Mortality Ratio: Number of maternal deaths per 100,000 live births:

*(Number of maternal deaths / Number of live births) * 100,000*

Modulus 10 validation – a mathematical formula used to determine accuracy of various identification numbers i.e., SA ID.

Neonatal deaths - the death of a baby within the first 28 days of life from any cause

Neonatal Mortality Rate: Number of babies dying before 28 days of life per 1,000 live births:

*(Number of infant deaths / Number of live births) * 1,000*

Probabilistic Record Linkage - the process of linking records from different data sources that have a high probability of belonging to the same individual when a unique identifier or identifier combinations are unavailable, unreliable or of insufficient quality[3,8,13,15].

PHDC/Provincial linkage - is a decision rule-based deterministic linkage approach used by the Western Cape Provincial Health Data Centre which uses fuzzy matching to generate confidence match scores that categorise records as exact, probable or possible matches.

Soundex algorithm – a phonetic algorithm producing a character string to identify a set of names phonetically alike.

ACKNOWLEDGEMENTS

The success of this project has been made possible by the contributions of many people in several different ways; I therefore express my sincere thanks to them all. I wish to thank the HIV- Social Behavioural Sciences Training and Research Program, particularly Professor John Bartlett - the program director, for providing partial financial support for my studies. My sincere gratitude is extended to my supervisor Prof. Andrew Boulle for his tireless efforts and support, constructive criticisms and guidance in writing this document, as well as my co-supervisor Greg Distiller, for his technical supports throughout the PhD programme. Special thanks to my mentor Prof. Sia E. Msuya for her unceasing support and professional advice. My final thanks are extended towards various Master of Public Health courses convenor's, librarians, the staff of the Provincial Government of the Western Cape - Department of Health Data Centre, and my fellow PhD candidates for their professional inputs and various supports.

ABSTRACT

Measuring and monitoring progress towards global development goals requires valid and reliable estimates of maternal and child mortality. This thesis has aimed at enumerating and estimating maternal and neonatal deaths from 2010 to 2013 in the Western Cape Province; and determining factors associated with these outcomes during the same period. This thesis comprises nine chapters, of which six present the research findings. The first results chapter has presented the findings from a systematic review, determining trends of maternal and neonatal mortality from 1990 to 2015 in South Africa. The review found that estimates of maternal and neonatal mortality are widely divergent across data sources and estimation methods, with conflicting trends over the analysis period. The second results chapter compared the performance of the existing decision-rule based linkage approach (provincial linkage) which uses fuzzy linkage to an independent fully probabilistic record linkage (PRL) implementation for identifying mortality records across the Western Cape Provincial Health Data. The PRL was shown to be a feasible method for future implementation, while the existing linkage performed similarly to the independent linkage exercise, providing reassurance on the adequacy of the linked datasets on which the subsequent chapters were based. The third and fourth results chapters involved the applications of three-source capture-recapture methods, to estimate maternal and neonatal mortality under-reporting in the Western Cape province. Based on these models, maternal and neonatal mortality under-reporting were estimated at 45.6% and 17.7% over the full 4-year period respectively. The last two results chapters focused on determining factors associated with maternal and neonatal mortality in this setting and exploring whether the estimates of association were altered through using an expanded number of outcome events based on

database linkage across multiple data sources. Most findings were consistent with known associations, as well as estimates from single-source analyses in the same setting. The thesis concludes that estimates of maternal and neonatal mortality are widely divergent in South Africa, and single-source reporting likely under-estimates the event rates. The application of capture-recapture methods is a viable approach in South Africa to resolve the problems of under-ascertainment in estimation of these outcomes.

CHAPTER 1: INTRODUCTION

1.1 Background

Maternal and neonatal mortality are important health challenges and insufficient progress has been made in addressing them. In 2015, about 303,000 women globally died during pregnancy, childbirth, or within 42 days of termination of a pregnancy from causes related to pregnancy or its management[1]. Of these, almost 99% (302,000) were from developing countries. Sub-Saharan Africa (SSA) accounted for approximately 66% (201,000) of the global maternal deaths[1].

The Millennium Development Goal (MDG) target 5A aimed at reducing the maternal mortality ratio (MMR) by three-quarters between 1990 and 2015[2,3]. However, a decline of only 44% in the MMR has been achieved i.e., 216 maternal deaths per 100,000 live births in 2015 from 385 maternal deaths per 100,000 live births in 1990[1]. Despite the reduction in global maternal mortality in 2015, developing regions had roughly 20 times higher MMR compared to the developed world[1]. Additionally, SSA recorded the highest MMR within the developing regions (546 maternal deaths per 100,000 live births)[1].

About 70%-80% of the global maternal deaths are due to direct causes such as maternal bleeding/haemorrhage, hypertensive disorders, sepsis (infection), eclampsia, obstructed labour, unsafe abortion and other direct causes. Only 20%-30% were due to indirect causes, such as anaemia, cardiovascular disease, malaria, HIV/AIDS and TB[4–9]. SSA remains the region mostly affected by HIV/AIDS, contributing about 85% of all HIV/AIDS related maternal deaths worldwide[1,7].

Evidence from the literature has shown that interventions to reduce maternal mortality also prevent neonatal mortality[10–15]. Globally, there were an estimated 2.7 million neonatal deaths in 2015[16]. Even though the neonatal mortality rate dropped from 36 deaths per 1,000 live births in 1990 to 19 deaths per 1,000 live births in 2015, it still accounted for 45% of all under-five child deaths[16]. Nearly three quarters of all neonatal deaths occurred during the first week of life; while 30%-40% occurred within the first 24 hours[16–21]. Low and middle-income countries (LMIC) accounted for the vast majority (85%) of these deaths[16–21].

Most neonatal deaths are considered preventable with proven, existing, and cost-effective interventions available[20]. Globally, preterm birth complications (35%); intrapartum-related complication (24%) and sepsis (15%) were the main causes of neonatal deaths[20–23]. Malnutrition and HIV/AIDS are also important contributory causes of neonatal mortality in SSA settings[8].

Compared to other SSA countries, South Africa has been reported to have a relatively low MMR (138 maternal deaths per 100,000 live births) with a high contribution from HIV (32%)[1]. However, the country did not achieve the MDG targets for maternal and child health by 2015. Between 1990 and 2015, the country recorded only moderate progress in achieving MDG 4 (reducing by two-thirds, between 1990 and 2015, the under-five mortality rate) with no progress in achieving MDG 5A (reducing by three quarters, between 1990 and 2015, the maternal mortality ratio) [1,4,7,8,23–28]. Progress is difficult to assess as the country has widely divergent estimates of maternal and neonatal mortality from institutional reporting and global metrics, mainly due to estimation techniques, variable data sources and the definitions used[29].

Monitoring progress towards MDG 4 and 5 required valid and reliable estimates of maternal and child mortality. Estimation of these outcomes in developing countries is challenging due to lack of accurate, valid and reliable data[4,27,30–34]. Lack of accurate and reliable estimates of neonatal and maternal mortality in South Africa has a substantial impact in monitoring trends, health planning and evaluation of health interventions. As described in the literature review, several methods for measuring and estimating maternal and neonatal mortality exist, and can be broadly categorized as empirical or analytical approaches[35]. However, the choice of estimation method primarily depends on the availability, coverage and quality of data in a given country.

Due to varying estimates of maternal and neonatal mortality provided by different sources in South Africa and the Western Cape province, there is value in applying additional estimation techniques which utilise multiple data sources to correct for under-reporting of these outcomes. Unique opportunities for estimating maternal and neonatal mortality are provided in the country, and particularly in the Western Cape Province, by having multiple data sources, which facilitates the application of capture-recapture methods. The available data sources often include unique identifiers, enabling potential data linkage.

Moreover, it is important to understand the factors contributing to these deaths. Given the HIV epidemic in the country and the massive uptake of HIV treatment and Prevention of Mother-to-Child Transmissions (PMTCT), it is also important to understand the ongoing relationship between HIV and HIV interventions, and maternal and neonatal outcomes.

1.2 Literature review

This section presents the details of the literature search strategy; estimates of MMR and NMR; the comparative estimates of MMR and NMR worldwide; trends in MMR and NMR in South Africa; factors associated with MMR and NMR and the estimation methods used in Low- and Middle-Income Countries (LMIC). Given that many studies provide estimates across multiple countries, in order to facilitate comparison of estimates, three countries (South Africa, Tanzania and Malawi) were chosen to illustrate the variability in estimates, based on availability of additional single-country studies, and similarities and differences to the South African context. The literature review in this chapter, while based on a systematic search strategy, is summative and illustrative in order to inform the current study, and not intended to reflect the full breadth of the articles reviewed. A formal systematic review is included as a thesis chapter, focussing specifically on the South African estimates of maternal and neonatal mortality.

The bulk of this review was undertaken at the time of initial protocol development in 2015. While some additional literature has been included, the review should be read as current at the time the study was undertaken. Relevant data and articles published while the analyses were underway are discussed in the concluding chapter.

1.2.1 Literature search strategy

Literature searches were conducted in several electronic databases including Medline, Africa-Wide Information, Scopus, Web of Science and CINAHL. Keywords used included maternal, neonatal, mortality, death, estimate, developing countries, Africa and South Africa. Searches

were limited to articles written in English. Reference lists of retrieved articles were also screened for additional publications (Appendix 1).

1.2.2 Estimation of maternal and neonatal mortality

Globally, limited guidance exists for estimating maternal mortality at country and/or sub-national level. The Technical Advisory Group (TAG) of the United Nations Inter-agency Group for Child Mortality Estimation (UN-IGME) i.e., UNICEF, WHO, the World Bank and the United Nations Population Division, provide guidance on methods of estimating neonatal mortality. Estimation procedures are recommended through the use of civil registration and sample surveys due to the representative nature of these data. In HIV-affected countries (HIV prevalence >5%), TAG has recommended estimation of non-HIV neonatal mortality and HIV-neonatal mortality separately, thereafter adding the estimates to obtain the total estimated neonatal mortality rate[21,36]. The methods of estimating MMR and NMR are briefly described in Chapter 2.

1.2.2.1 Country and sub-national estimates of maternal mortality

According to Maternal Mortality Inter-Agency Group (MMEIG)[1]; in 2015, Sierra Leone was the country with the highest MMR in SSA (MMR=1360 per 100,000 live births; 80% Uncertainty Interval (UI): 999 to 1980). Afghanistan (MMR=396; 80%UI: 253,620), Yemen (MMR=385; 80%UI 274,582) and Haiti (MMR=359; 80%UI: 236,601) were the countries with the highest maternal mortality rates outside the SSA region. In SSA, Cabo Verde and Mauritius were the countries with the lowest MMR, i.e., MMR=53 (80%UI: 38,77) and MMR=42 (80%UI: 20,95) respectively. Unlike most SSA countries, South Africa had a relatively low MMR (MMR=138, 80%UI: 124,154) of which 32% of the deaths were attributed to HIV[1].

A systematic review done in 2014 in 188 countries ranked MMR by country slightly differently and showed substantial variations in MMR by countries[7]. Globally, South Sudan and the Central African Republic recorded the highest MMR in the year 2013 (MMR=956; 95%UI: 685,1262 and MMR=911; 95%UI: 578,1293, respectively), in contrast to Iceland which had the lowest MMR of 2.4 maternal deaths per 100,000 live births (95%UI: 1.6,3.6).

According to regions, Bolivia had the highest MMR in South America (MMR=180; 95%UI: 110,257). Haiti and Guyana had the highest MMR in the Caribbean (MMR=333; 95%UI: 219,480 and MMR=118; 95% UI: 76,179, respectively). Similarly, in South Asia, Afghanistan had the highest MMR of 885 maternal deaths per 100,000 live births (95%UI: 508,1445). Additionally, Yemen was the country with highest MMR in North Africa and Middle East (MMR=309; 95%UI: 509,555), while Papua New Guinea was the highest in South-East Asia and Oceania region (MMR=594; 80%UI: 313,1031). Seychelles was the country with lowest MMR in the SSA region (MMR=15.7; 95%UI: 12,21). According to the report, in 2013, South Africa had 174 maternal deaths per 100,000 live births (95%UI: 96,257), with the proportion of maternal deaths attributed to HIV not documented.

The tenth interim report on Confidential Enquiries into Maternal Deaths (CEMD) in South Africa showed an institutional Maternal Mortality Ratio (IMMR) of 140 maternal deaths per 100,000 live births (95% UI: 85,210) in the year 2012. Free State province recorded the highest IMMR in contrast to the Western Cape province which had the lowest IMMR in the country (240.4 vs. 78.6 maternal deaths per 100,000 live births)[37].

Table 1 shows the selected literature on country/sub-national estimates of maternal mortality, focussing on African settings, in particular South Africa and the designated comparator countries described earlier (Tanzania and Malawi).

Table 1: Countries and sub-national estimates of maternal mortality

Author /country	**Study design** **-Sample size** **-Data source**	**Methods of estimation**	**Findings (year):** *Estimate (UI)*
Maternal Mortality Estimation Inter-Agency Group, 2015* (Multi-country)	Report	Bayesian maternal mortality estimation model	**MMR - 2015** South Africa – 138 (124,154) Tanzania – 398 (281,570) Malawi – 634 (422,1080)
Kassebaum et al, 2014** (Multi-country)	-Record review -188 countries -GBD 2013 cause of death database	Statistical model (Cause of death ensemble model)	**MMR - 2013** South Africa – 174 (96,257) Tanzania – 389(227,549) Malawi – 334.7 (226,465)
Colbourn et al, 2013 Malawi	Systematic review	Population based estimate	**MMR-2010** MMR – 484 (national average)
Udjo et al, 2013 South Africa	-Secondary data analysis -Census, sample survey, vital registration	Statistical model (Growth balance method)	**MMR-2007** South Africa-764 Western cape – 102 Eastern Cape – 1639 Free state – 1080 Kwazulu-Natal – 969 Northern cape – 610 Gauteng – 484
Confidential enquiry, 2012 South Africa	Interim report -Institutional reporting	Direct estimate	**IMMR- 2012** South Africa– 146.7 Free state – 240.4 Kwazulu-Natal – 180.6 Eastern cape – 180.4 Northern cape – 173.5 Gauteng – 125.3 Western cape – 78.6
Lozano et al, 2011** (Multi-country)	Systematic review	Statistical model (Ensemble model)	**MMR-2011** South Africa – 91(69,121) Zimbabwe – 329(232,470) Botswana – 575(363,689)

			Tanzania – 417(337,511)
DHS, 2010 (Multi-country)	Survey	Direct sisterhood	**MMR-2010** Malawi – 675(570,780) Tanzania – 454 (353,556)
Hogan et al, 2010** (Multi-country)	Systematic review	Statistical method	**MMR-2008** South Africa–237 (146,372) Malawi – 1140(675,1813) Tanzania – 449 (273,721)

*UI-Uncertainty Interval; *80% Uncertainty Interval; **95% Uncertainty Interval; ND - Not documented*

The maternal mortality estimates are highly variable, ranging from 335 to 1140 deaths per 100 000 live births for Malawi, 389 to 449 deaths per 100,000 live births for Tanzania and 91 to 764 deaths per 100,000 live births for South Africa, noting all within a decade of each other. These are covered in further detail in Chapter 3.

1.2.2.2 Trends in maternal mortality in South Africa

According to the estimates from MMEIG, South Africa has recorded insufficient progress in achieving MDG 5A[1]. In the country, MMR has increased by 27.8%: from 108 maternal deaths per 100 000 live births in 1990 to 138 maternal deaths per 100 000 live births in 2015. The MMR varied substantially from 1990 to 2015, most probably due to the HIV epidemic and the subsequent introduction of treatment for HIV[29]. AIDS-related indirect maternal deaths increased from 1700 in 1990 to 12,000 in 2005, thereafter decreased to 7500 AIDS-related indirect maternal deaths in 2013 likely due to high ART coverage among eligible patients (>80%) in the country[38–41].

A systematic review done by Kassebaum *et al*[7]*;* indicated remarkable changes in MMR in South Africa: from a ratio of 134 (95%UI: 93,175) maternal deaths per 100 000 live births in 1990 to 342 (95%UI: 228,481) maternal deaths per 100 000 live births in 2003, followed by a decline to 174 (95%UI: 96,257) maternal deaths per 100 000 live births in 2013. Comparing MMR just in 2013

and 1990, this is consistent with an increase in deaths from both direct and indirect obstetric causes recorded in SSA during the same period[7].

The study done by Lozano *et al*[42] has provided findings contrary to Kassebaum whereby in South Africa, a decline of MMR from 120 to 91 maternal deaths per 100 000 live births was recorded from 1990 to 2011. The decline in MMR was linked with a reduction in HIV-related maternal deaths which was mainly due to the steady scale-up of ART in the country[38,39,41,42].

There was an increased IMMR recorded by the Confidential Enquiries into Maternal Deaths from 1998 to 2009, and a decrease from 2009 to 2012 in South Africa. The annual rate of change in IMMR in the country from 1998 to 2012 was insufficient in relation to achieving the MDG target 5A. AIDS/HIV related deaths and caesarean-section-associated haemorrhage were the main contributory causes of a slow progress in MDG 5A[37,43,44]. Table 2 shows the selected literature on trends in maternal mortality in South Africa.

Table 2: Trends in maternal mortality in South Africa

Author/Country	**Methods of estimation**	**Findings (year):** *Estimate (UI)*
Maternal Mortality Estimation Inter-Agency Group, 2015* (Multi-country)	Bayesian maternal mortality estimation model	**MMR - 2015** South Africa – 138 **MMR - 2010** South Africa – 154 **MMR - 2005** South Africa – 112 **MMR - 2000** South Africa – 85 **MMR - 1995** South Africa – 62 **MMR - 1990** South Africa – 108
Kassebaum et al, 2014** (Multi-country)	Statistical model	**MMR - 2013** South Africa – 174 (96,257)

		MMR - 2003 South Africa – 342 (228,481) **MMR - 1990** South Africa – 134 (93,175)
Confidential enquiry, 2012 South Africa	Direct	**IMMR- 2012** South Africa – 146.7 **IMMR- 2011** South Africa – 159.1 **IMMR- 2003** South Africa – 120 **IMMR- 1999** South Africa – 80
Lozano et al, 2011** (Multi-country)	Statistical model (Ensemble model)	**MMR-2011** South Africa – 91 (69,121) **MMR - 2000** South Africa – 89 (78,101) **MMR - 1990** South Africa – 119.6 (99,145)
Hogan et al, 2010** (Multi-country)	Statistical model	**MMR-2008** South Africa – 237 (146,372) **MMR - 2000** South Africa – 155 (95,248) **MMR - 1990** South Africa – 121 (73,190)

*UI-Uncertainty Interval; *80% Uncertainty Interval; **95% Uncertainty Interval; ND - Not documented*

Similar to quite divergent point estimates during the same calendar periods, there are also divergent estimates of trends in maternal mortality. The trends in maternal mortality are broadly discussed in the systematic review in Chapter 3.

1.2.2.3 Country and sub-national estimates of neonatal mortality

According to the UN-IGME report, in 2015 there were 11 neonatal deaths per 1,000 live births in South Africa[16]. In comparison to most SSA countries, South Africa had notably low early (0-6 days) and late (7-28 days) neonatal mortality rates[23,42]. Once again, estimates for South Africa are contrasted with those of Malawi and Tanzania for illustrative purposes (Table 3).

Table 3: Countries and sub-national estimates of neonatal mortality

Author /Country	**Study design** **-Sample size** **-Data source**	**Methods of estimation**	**Findings (year):** *Estimate (UI)*
UN-IGME, 2017* (Multi-country)	Report Civil registration Census Sample survey	Statistical model (Bayesian B-splines Bias-reduction model [B3 model])	**NMR - 2015** South Africa – 11 Tanzania – 19 Malawi – 22
Wang et al, 2014** (Multi-country)	-Record review -188 countries -Survey, census, vital and sample registration	Statistical model (Gaussian process regression)	**NMR - 2013** South Africa Early – 11.4 (9.5,13.3) Late – 3.2 (2.5,4.2) Tanzania Early – 18.0 (15.7,20.4) Late – 5.8 (4.8,6.9) Malawi Early – 19.7 (16.9,22.1) Late – 6.5 (5.5,7.4)
Lozano et al, 2011** (Multi-country)	Systematic review	Statistical model (Ensemble model)	**NMR - 2011** South Africa Early – 10.8 (9.8,11.8) Late – 3.5 (3.2-3.9) Tanzania Early – 18.5 (16.4-20.8) Late – 5.8 (5.2,6.6) Malawi Early – 20.6 (17.9,23.6) Late – 6.6 (5.6,7.9)

*UI-Uncertainty Interval; *80% Uncertainty Interval; **95% Uncertainty Interval; ND - Not documented*

Estimates of neonatal mortality are more consistent than those of maternal mortality. However, there are fewer estimates than those of maternal mortality. The available estimates of neonatal mortality are discussed in further detail in Chapter three.

1.2.2.4 Trends in neonatal mortality in South Africa

Regarding neonatal mortality (Table 4), South Africa experienced a moderate (1.6%) annualised average reduction in neonatal mortality between 1990 and 2015[16]. However, this reduction was not enough to achieve the MDG 4 by 2015. Several studies suggested the annualised reduction is associated with the scale-up of child and maternal interventions in the country, massive scale-up of Prevention of Mother-To-Child Transmission of HIV (PMTCT) and the increasing roll-out of ART[22,24,45,46].

Table 4: Trends in neonatal mortality in South Africa

Author	**Methods of estimation**	**Findings (year):** *Estimate (95% UI*)*	**% Annual rate of reduction**
UN-IGME, 2015 (Multi-country)	Report Civil registration Census Sample survey	**NMR - 2013** South Africa – 15 **NMR - 2010** South Africa – 18 **NMR - 1990** South Africa – 18	**1990 – 2015** -1.6%
WHO, 2015 (Multi-country)	Statistical model	**NMR - 2012** South Africa – 15 **NMR - 2005** South Africa – 18 **NMR - 1995** South Africa – 19	ND
Nannan et al, 2012 South Africa	Direct -DHIS -PPIP -ACDIS	**INMR- 2008** KZN– 10.6 **INMR- 2006** KZN – 14.4 **INMR- 2003-2006** KZN – 13.6 **INMR- 2000** KZN – 15	ND
Kerber et al, 2013 South Africa	Mathematical model (UCT Paediatric HIV model)		**1990 – 2011** -2.3% **2006 – 2011** -3.4%

**UI-Uncertainty Interval; ND - Not documented*

Similar to there being relatively consistent point estimates of the NMR during the same calendar periods, estimated trends in neonatal mortality are broadly aligned. The trends in neonatal mortality in South Africa are covered in greater detail in Chapter four.

1.2.3 Methods of estimating MMR and NMR

Achieving the goal of safe motherhood and tracking progress towards MDG 4 and 5, various methods for measuring and estimating maternal and neonatal mortality have been developed, tested and are now widely used. Most approaches have been either empirical measurement, or simulation to compensate for under-ascertainment, with few attempts at analytic approaches to estimate the true burden. These empirical estimation techniques can be further categorised as direct and indirect. Empirical measurement methods include death registration (civil/vital registration and Sample Vital Registration with Verbal Autopsy (SAVVY)); health facility statistics, decennial census, surveys (Demographic Health Survey (DHS) and Multiple Indicator Cluster Survey - (MICS)); surveillance (Confidential Enquiries, Demographic Surveillance System - (DSS) & Active Surveillance of Reproductive Age Female Deaths); Reproductive-Age Mortality Studies (RAMOS) and birth and death record linkage. Analytical measurement methods comprise modelling or simulation i.e., mathematical, ecological or statistical models such as capture-recapture methods[4,27,30,31,34,35,47–51]. Figure 1 below summarises the general methods of estimating maternal and neonatal mortalities.

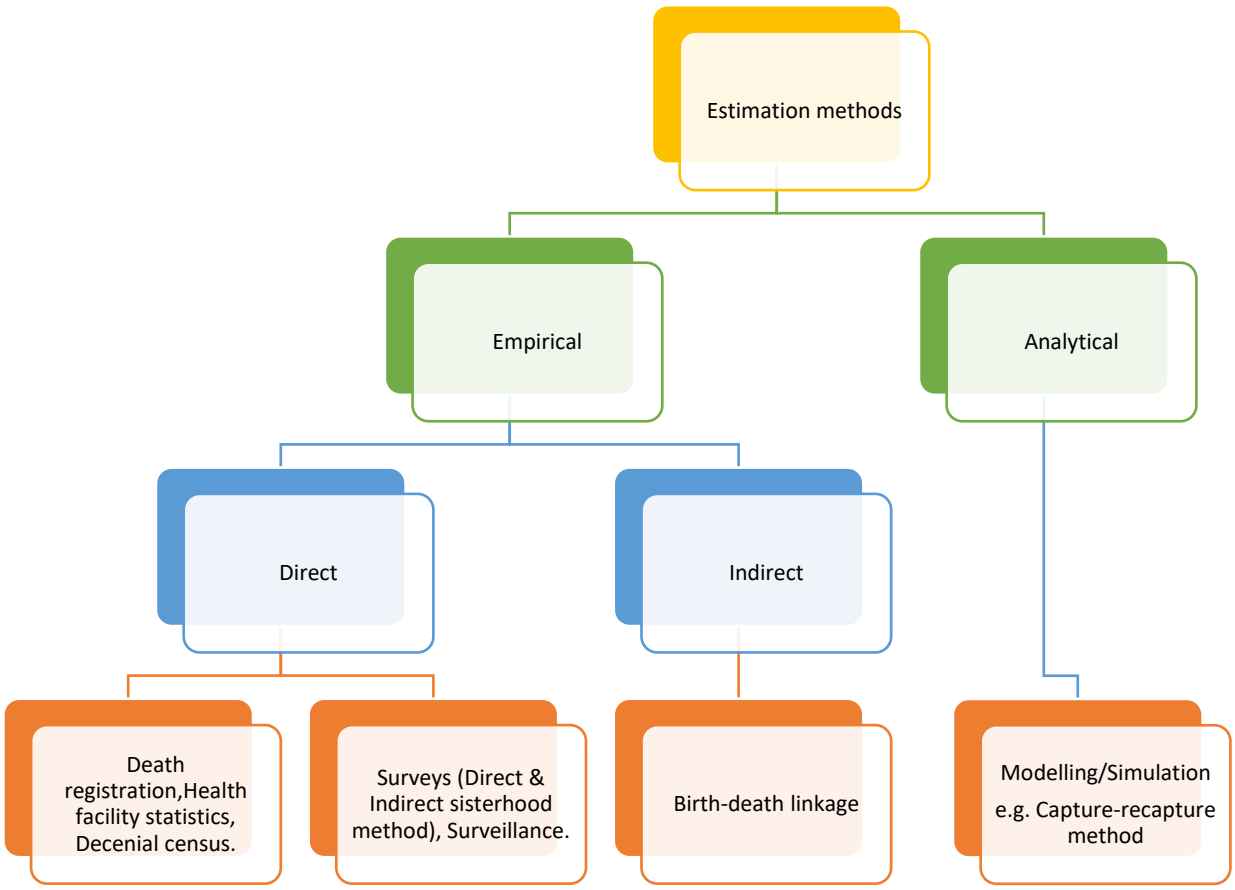


Figure 1: Methods of estimating MMR and NMR

The choice of estimation method primarily depends on the availability, coverage and quality of data in a particular area. In areas with complete death registration systems and accurate ascertainment of cause of deaths, estimates of maternal and neonatal mortality are obtained directly from these data. However, in areas with incomplete data; analytical measurement methods are used for estimation purposes[4,30,52]. Where direct estimation is not possible, or estimation is required across multiple settings, mathematical/statistical modelling is frequently used, which integrates demographic parameters (mortality, fertility and age structure) and calibrates with maternal mortality estimates from similar contexts.

Seychelles and Mauritius are the only African countries having near-complete vital registration systems which facilitate estimation of maternal and neonatal mortality directly[53]. Among developing countries, South Africa has national facility-based mortality audits for maternal, perinatal and child death, focused both on estimating trends, as well as describing avoidable causes amenable to intervention[43,54]. The estimation of maternal and neonatal mortality in the country is often based on the National Confidential Enquiry into Maternal Deaths (NCEMD) which records maternal deaths, and Perinatal Problem Identification Programme (PPIP) which records stillbirths and neonatal deaths[25,37,55].

1.2.4 Factors associated with maternal and neonatal mortality

Several studies have revealed different factors associated with maternal and neonatal mortality ranging from individual, underlying healthcare factors and disease control. The most commonly studied factors associated with maternal mortality include: maternal diseases (HIV/AIDS, malaria, TB, anaemia, sickle cell), maternal age, maternal education, ANC attendance, facility delivery, assistance by a skilled birth attendant, parity and place of residence[37,56–67]. The most common factors associated with neonatal mortality include maternal diseases, neonatal infections, nutrition, gestation age and parity[68–72].

In South Africa, the commonest preventable causes of maternal mortality include non-pregnancy related infections, obstetric haemorrhage and hypertension[37,43,73]. Almost half of maternal deaths in the country are due to avoidable patient-associated factors such as delays in seeking healthcare, transport problems and unsafe abortion by unregistered providers. Administrative factors such as lack of transport between facilities, poor access to intensive care units, poor availability of blood

and inadequately trained medical personnel contribute to more than one third of maternal deaths in South Africa. Health-care provider associated avoidable factors contributing to these deaths include delays in referral, poor clinical assessments, failure to respond to abnormalities, not following standard protocols and poor monitoring of patients[37,43,73].

Most neonatal deaths in South Africa are due to common and often preventable causes such as prematurity, low birth weight and birth asphyxia. The person-associated avoidable factors include delays in seeking medical care during labour, late booking in pregnancy, few ANC visits, and not initiating antenatal care. Healthcare provider associated factors for these deaths include improper monitoring with delayed detection of foetal distress, unattended prolonged second stage of labour, delays in referrals and delays in seeking assistance from a specialist.

Possible avoidable administrative factors include inadequate number of facilities, lack of access to an intensive care unit bed with ventilator, inadequate resuscitation equipment and lack of transport for referral[72].

1.3 Rationale

There are widely divergent estimates of MMR and NMR in South Africa, including of temporal trends in these. This undermines attempts at estimating relative performance in key health outcomes across settings, as well as changes in outcomes due to changes in risk factors or health service delivery. Estimates of MMR and NMR in South Africa rely on direct estimation with likely under-reporting. Corrected estimates of MMR and NMR are mostly produced by

mathematical/statistical models, with few analytical approaches applied to correct for underreporting.

There are no published previous attempts from South Africa at combining multiple empirical data sources analytically to estimate these parameters more accurately. Estimated associations between potential risk factors and MMR and NMR may also be less biased on account of improved enumeration of outcomes when considering an expanded number of events derived from multiple sources. The research purpose, therefore, is to improve the estimation of maternal and neonatal mortality event rates over time, to better inform health service delivery and improve analyses of associations with these outcomes.

1.4 Aims and Objectives

1.4.1 Aim

- To enumerate and estimate maternal and neonatal mortality and determine associations with mortality, from 2009 to 2013 in the Western Cape Province, South Africa.

1.4.2 Objectives

1. To systematically review the trends in estimates of maternal and neonatal mortality from 1990 – 2015 in South Africa.
2. To compare the performance of a decision-rule based linkage approach which uses fuzzy or non-exact comparisons, to a fully probabilistic approach to identifying patients across datasets.
3. To determine the proportions of maternal deaths captured by different data sources from 2009 – 2013 in the Western Cape Province, South Africa.

4. To determine the proportions of neonatal deaths captured by different data sources from 2009 – 2013 in the Western Cape Province, South Africa.
5. To estimate maternal mortality from 2009 – 2013 in the Western Cape Province, South Africa - using capture recapture methods.
6. To estimate neonatal mortality from 2009 – 2013 in the Western Cape Province, South Africa, using capture-recapture methods.
7. To determine associations with maternal mortality from 2009 – 2013 in the Western Cape Province, South Africa.
8. To determine associations with neonatal mortality from 2009 – 2013 in the Western Cape Province, South Africa.

1.4 Public health significance

The corrected estimates of maternal and neonatal mortalities to be obtained from this study will help to track actual progress in MDGs 4 & 5 and will be used in evaluating successes and failures of interventions focusing on reducing maternal and child mortality in the country ,and particularly the Western Cape province over the past decade. This study will provide baseline data for better appreciation of the gap in achieving Sustainable Development Goals (SDGs), and on where interventions to reduce maternal and neonatal mortality should be focusing, based on associations identified with these outcomes.

1.5 Ethical approval

Approval to conduct this study was obtained from the University of Cape Town (UCT) Human Research Ethics Committee (HREC-703/2016). Access to the data was sought through the

Provincial Department of Health Data Access Guidelines following ethical and provincial approval; and was conducted in accordance with the standard operating procedures of the Provincial Health Data Centre and under supervision of the Data Centre staff. Data linkage procedures required using patient identifiers. For this reason, identified data were only accessed at the Western Cape Department of Health offices under supervision; and delinked from clinical data. To ensure privacy and confidentiality, data at all stages were stored on password-protected, encrypted files and only anonymised data were taken off-site. Once the data linkage exercise was complete, data were anonymised, and only anonymised datasets were retained for further analyses.

1.6 Funding

This research was partially supported by the Socio-behavioural Sciences Research to Improve Care for HIV Infection in Tanzania – a National Institute of Health funded program – Grant # D43-TW009595.

1.7 References

1. WHO, UNICEF, UNFPA, World Bank Group & The United Nations Population Division. *Trends in Maternal Mortality: 1990 to 2015. Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*. (2015). doi:ISBN 978 92 4 156514 1

2. Killen, B. *et al.* Health and the Millennium Development Goals. 84 (2005).

3. Oxfam. The Millennium Development Goals. *Monthly Review An Independent Socialist Magazine* **57,** 1–15 (2006).

4. WHO, UNICEF, UNFPA, The World Bank & The United Nations Population Division. *Trends in Maternal Mortality: 1990-2013. Estimates by WHO,UNICEF, UNFPA, The World Bank and the United Nations Population Division*. *WHO, UNICEF, UNFPA, The World Bank The United Nations Population Division* (2014). doi:WHO/RHR/14.13

5. Say, L. *et al.* Global causes of maternal death: A WHO systematic analysis. *Lancet Glob. Heal.* **2,** 323–333 (2014).

6. Khan, K. S., Wojdyla, D., Say, L., Gülmezoglu, a. M. & Van Look, P. F. WHO analysis of causes of maternal death: a systematic review. *Lancet* **367,** 1066–1074 (2006).

7. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384,** 980–1004 (2014).

8. Kinney, M. V *et al.* Sub-Saharan Africa's mothers, newborns, and children: Where and why do they die? *PLoS Med.* **7,** 1–9 (2010).

9. Lumbiganon, P. *et al.* Indirect causes of severe adverse maternal outcomes: a secondary analysis of the WHO Multicountry Survey on Maternal and Newborn Health. *BJOG* **121 Suppl,** 32–39 (2014).

10. Lassi, Z. S., Salam, R. A., Das, J. K. & Bhutta, Z. A. Essential interventions for maternal, newborn and child health: background and methodology. *Reprod. Health* **11 Suppl 1,** S1 (2014).

11. Lassi, Z. S. *et al.* Essential interventions: implementation strategies and proposed packages of care. *Reprod. Health* **11 Suppl 1,** S5 (2014).

12. Lassi, Z. S., Mansoor, T., Salam, R. A., Das, J. K. & Bhutta, Z. A. Essential pre-pregnancy and pregnancy interventions for improved maternal, newborn and child health. *Reprod.*

Health **11 Suppl 1,** S2 (2014).
13. Gates, M. & Binagwaho, A. Newborn health: A revolution in waiting. *The Lancet* **384,** 2013–2014 (2014).

14. United Nations. *The Millennium Development Goals Report*. (2014).

15. Darmstadt, G. L. *et al.* Evidence-based, cost-effective interventions: How many newborn babies can we save? *Lancet* **365,** 977–988 (2005).

16. UNICEF, WHO, World Bank, U.-D. P. D. *Levels and trends in child mortality 2015*. *WHO* (World Health Organization, 2017).

17. Anthony, D. & Mullerbeck, E. *Committing to Child Survival: A Promise Renewed: UNICEF Progress Report 2013*. (2013). doi:978-92-806-4706-8

18. Wardlaw, T., You, D., Newby, H., Anthony, D. & Chopra, M. Child survival: a message of hope, but a call for renewed commitment in UNICEF report. *Reprod. Health* **10,** 64 (2013).
19. United Nations. *The Millennium Development Goals Report*. (2014).

20. Wardlaw, T., You, D., Hug, L., Amouzou, A. & Newby, H. UNICEF Report : enormous progress in child survival but greater focus on newborns urgently needed. *Reprod. Health* **11,** 1–4 (2014).

21. UNICEF, WHO, The Wolrd bank & UN. *Levels and trends in child mortality : estimates developed by the UN Inter-agency Group for child Mortality Estimation (IGME) - report 2014*. (2014).

22. Lawn, J. E. *et al.* Every Newborn: progress, priorities, and potential beyond survival. *Lancet* **384,** 189–205 (2014).

23. Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384,** 957–979 (2014).

24. Chopra, M. *et al.* Achieving the health Millennium Development Goals for South Africa: challenges and priorities. *Lancet* **374,** 1023–1031 (2009).

25. Chopra, M., Daviaud, E., Pattinson, R., Fonn, S. & Lawn, J. E. Saving the lives of South Africa's mothers, babies, and children: can the health system deliver? *Lancet* **374,** 835–846 (2009).

26. Murray, C. J. L. *et al.* Global, regional, and national incidence and mortality for HIV,

tuberculosis, and malaria during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384,** 1005–1070 (2014).

27. Kjaergaard, J. *Levels & Trends in Child Mortality - Report 2010: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. *Hernia* **14,** (2010).

28. Lawn, J. Maternal and child health—now, then, or when? *The Lancet* **375,** 1957–1958 (2010).

29. Damian, D. J., Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review. *Syst. Rev.* **8,** 76 (2019).

30. Merdad, L., Hill, K. & Graham, W. Improving the Measurement of Maternal Mortality: The Sisterhood Method Revisited. *PLoS One* **8,** (2013).

31. Graham, W. J. Now or never: The case for measuring maternal mortality. *Lancet* **359,** 701–704 (2002).

32. Berg, C. J., Danel, I. & Mora, G. Guidelines for Maternal Mortality Epidemiological Surveillance. *Pan Am. Heal. Organ.* 50 (1996).

33. Walker, N., Hill, K. & Zhao, F. Child Mortality Estimation: Methods Used to Adjust for Bias due to AIDS in Estimating Trends in Under-Five Mortality. *PLoS Medicine* **9,** (2012).

34. Silva, R. Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories. *PLoS Med.* **9,** e1001296 (2012).

35. Graham, W. J., Ahmed, S., Stanton, C., Abou-Zahr, C. & Campbell, O. M. R. Measuring maternal mortality: an overview of opportunities and options for developing countries. *BMC Med.* **6,** 12 (2008).

36. UNICEF, WHO, World Bank, U. P. D. Estimates Developed by the UN Inter-agency Group for Child Mortality estimation. 1–32 (2012).

37. Pattinson, R., Fawcus, S. & Moodley, J. *Tenth interim report on Confidential Enquiries into Maternal Deaths in South Africa*. (2012).

38. WHO, UNICEF & UNAIDS. *Global HIV/AIDS Response Progress Report 2011*. *Vasa* (2011). doi:ISBN 978 92 4 150298 6

39. UNAIDS. Access to Antiretroviral Therapy in Africa, 2013. *Status Rep. Prog. Towar. 2015 targets* 1–10 (2013).

40. Johnson, L. F. Access to antiretroviral treatment in South Africa, 2004 - 2011. *South. Afr.*

J. HIV Med. **13,** 22–27 (2012).

41. Bradshaw, D. & Dorrington, R. E. Maternal mortality ratio - trends in the vital registration data. *Sajog-South African J. Obstet. Gynaecol.* **18,** 38–42 (2012).

42. Lozano, R. *et al.* Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: An updated systematic analysis. *Lancet* **378,** 1139–1165 (2011).

43. Moodley, J. *et al.* The confidential enquiry into maternal deaths in South Africa: a case study. *BJOG An Int. J. Obstet. Gynaecol.* **121,** 53–60 (2014).

44. NCCEMD. *Saving Mothers 2005-2007 : Fourth Report on Confidential Enquiries into Maternal Deaths in South Africa*. *Health (San Francisco)* (2008).

45. Kerber, K. J. *et al.* South African child deaths 1990-2011: have HIV services reversed the trend enough to meet Millennium Development Goal 4? *AIDS* **27,** 2637–48 (2013).

46. Nannan, N. *et al. Under-5 mortality statistics in South Africa: Shedding some light on the trends and causes 1997-2007.* (2012).

47. WHO. *Beyond the numbers: reviewing maternal deaths and complications to make pregnancy safer. British medical bulletin* **67,** (2004).

48. Graham, W. The sisterhood method for estimating maternal mortality. *Mothers Child.* **8,** 1–2 (1989).

49. Qomariyah, S. N. *et al.* An option for measuring maternal mortality in developing countries: a survey using community informants. *BMC Pregnancy Childbirth* **10,** 74 (2010).

50. Yazdizadeh, B. *et al.* Application of capture-recapture for fine-tuning uncertainties about national maternal mortality estimates. *Int. J. Prev. Med.* **5,** 624–631 (2014).

51. Stanton, C., Abderrahim, N. & Hill, K. DHS Maternal Mortality Indicators : An Assessment of Data Quality and Implications for Data Use. *DHS Anal. Rep.* **4,** (1997).

52. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20,** 3123–3157 (2001).

53. Setel, P. W. *et al.* A scandal of invisibility: making everyone count by counting everyone. *Lancet* **370,** 1569–1577 (2007).

54. Bradshaw, D. *et al.* Every death counts: use of mortality audit data for decision making to

save the lives of mothers, babies, and children in South Africa. *Lancet* **371,** 1294–1304 (2008).

55. Pattinson, R. & Rhoda, N. *Saving Babies 2012-2013: Ninth report on perinatal care in South Africa*. *RC Pattinson* (2014).

56. Menéndez, C. *et al.* An autopsy study of maternal mortality in Mozambique: The contribution of infectious diseases. *PLoS Med.* **5,** 0220–0226 (2008).

57. Høj, L., da Silva, D., Hedegaard, K., Sandström, A. & Aaby, P. Factors associated with maternal mortality in rural Guinea-Bissau. A longitudinal population-based study. *BJOG* **109,** 792–9 (2002).

58. Magadi, M., Diamond, I. & Madise, N. Analysis of factors associated with maternal mortality in Kenyan hospitals. *J. Biosoc. Sci.* **33,** 375–89 (2001).

59. Nair, M. *et al.* Factors associated with maternal death from direct pregnancy complications: a UK national case-control study. *BJOG* **122,** 653–62 (2015).

60. Asamoah, B. O. *et al.* Distribution of causes of maternal mortality among different socio-demographic groups in Ghana; a descriptive study. *BMC Public Health* **11,** 159 (2011).

61. Ahmed, Y. *et al.* A study of maternal mortality at the University Teaching Hospital, Lusaka, Zambia: The emergence of tuberculosis as a major non-obstetric cause of maternal death. *Int. J. Tuberc. Lung Dis.* **3,** 675–680 (1999).

62. Geelhoed, D. W. *et al.* Trends in maternal mortality : a 13-year hospital-based study in rural Ghana. **107,** 135–139 (2003).

63. Laopaiboon, M. *et al.* Advanced maternal age and pregnancy outcomes: a multicountry assessment. *BJOG* **121 Suppl,** 49–56 (2014).

64. Berhan, Y. & Berhan, A. Review Antenatal Care as A Means of Increasing Birth in The Health Facility and Reducing Maternal Mortality : A Systematic Review. 93–104

65. Olsen, B. E. *et al.* Causes and characteristics of maternal deaths in rural northern Tanzania. *Acta Obstet. Gynecol. Scand.* **81,** 1101–1109 (2002).

66. Tunçalp, Ö. *et al.* Education and severe maternal outcomes in developing countries: a multicountry cross-sectional survey. *BJOG* **121 Suppl,** 57–65 (2014).

67. Souza, J. P. & WHO Multicountry Survey on Maternal. The World Health Organization Multicountry Survey on Maternal and Newborn Health project at a glance: the power of collaboration. *BJOG* (2014). doi:10.1111/1471-0528.12690

68. UNICEF, WHO, World Bank, U. *Levels & Trends in Child Mortality in 2014- Report: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. (2014).

69. Mmbaga, B. T. *et al.* Causes of perinatal death at a tertiary care hospital in Northern Tanzania 2000-2010: a registry based study. *BMC Pregnancy Childbirth* **12,** 139 (2012).

70. Schmiegelow, C. *et al.* Factors associated with and causes of perinatal mortality in northeastern Tanzania. *Acta Obstet. Gynecol. Scand.* **91,** 1061–1068 (2012).

71. Yego, F. *et al.* A retrospective analysis of maternal and neonatal mortality at a teaching and referral hospital in Kenya. *Reprod. Health* **10,** 13 (2013).

72. Pattinson, R. C. (Mr. R. U. for M. and I. H. C. S. on B. of P. U. *Saving Babies 2010-2011 : Eighth report on perinatal care in South Africa*. (2013).

73. Allanson, E. R., Muller, M. & Pattinson, R. C. Causes of perinatal mortality and associated maternal complications in a South African province: challenges in predicting poor outcomes. *BMC Pregnancy Childbirth* **15,** 37 (2015).

CHAPTER 2: Data Sources, Setting and Common Methods

2.1 Study setting

This study was conducted in the Western Cape province, South Africa. Over decades, the Western Cape province has been performing relatively well in most key health indicators as compared to other provinces[1]. According to recent statistics, the province has an estimated population of about 6.5 million, with a total fertility rate estimated at 2 births[2]. Additionally, life expectancy at birth is estimated at 67 years for males and 72 years for females[2]. About 89% of pregnant women attend antenatal care visits four times or more; 94.3% of women receive antenatal care from a skilled provider; 99.2% of deliveries are assisted by skilled birth attendants; 90.8% of women attend postnatal care within two days after delivery; 43.5% of children receive all age-appropriate vaccinations; and immunisation coverage (BCG, three doses of DTaP-IPV-Hib, and one dose of measles vaccine) is estimated at 87.7%[3,4]. Regarding health expenditure, in 2015-2016, the province ranked second-highest in total health expenditure in the country[5].

2.2 Data sources and types

Data used for the analyses of Chapters 4 to 8 come from the Western Cape Department of Health (WCDOH) routine administrative and clinical data and the South African Department of Home Affairs (DHA). These data were extracted from the Western Cape Provincial Health Data Centre (PHDC) during the analysis period, where they had previously been linked.

In the Western Cape province, the PHDC was formally established in 2015 (although also incorporating historical data from prior to 2015). Among its core functions, the PHDC supports patient care either through clinician-interfaces or patient line-listing by identifying patients who

require interventions. Furthermore, it acts as a functional health information exchange, providing a consolidated data environment with which multiple systems can interact based on the interoperability standards, to ensure data completeness and continuity.

The PHDC has four key databases maintained on the Microsoft SQL Server database platform i.e., Patient, Clinical, Archive and Staging databases[6,7]. The Patient database holds demographic and other patient identification information of all individuals issued with a valid unique health identifier; the Clinical database stores clinical information, such as episodes, encounters, pharmacy, laboratory, etc. The Archive database stores raw/unprocessed data from various sources for back-ups; and the Staging database holds newly received data from various data sources prior to processing. Each database has different tables, which can also be referred to as datasets.

At the Data Centre, the individual patient-level data are assigned an internal anonymous-unique identifier across datasets. Currently, over 90% of patient-mapping is based on the unique identifier also known as Patient Master Index (PMI). However, this is not the case for all datasets, including the death certificates which do not record a folder number. The absence of a folder number which doubles as a unique identifier, allows linkage to be done on the combinations of other identifiers. High levels of individual patient-level data linkable on the PMI provide opportunities for clinical management, healthcare administration, business intelligence and epidemiological research activities[6,7]. Figure 2 shows how patients maintain a unique health identifier (PMI) across multiple domains to facilitate linkage of patient data.

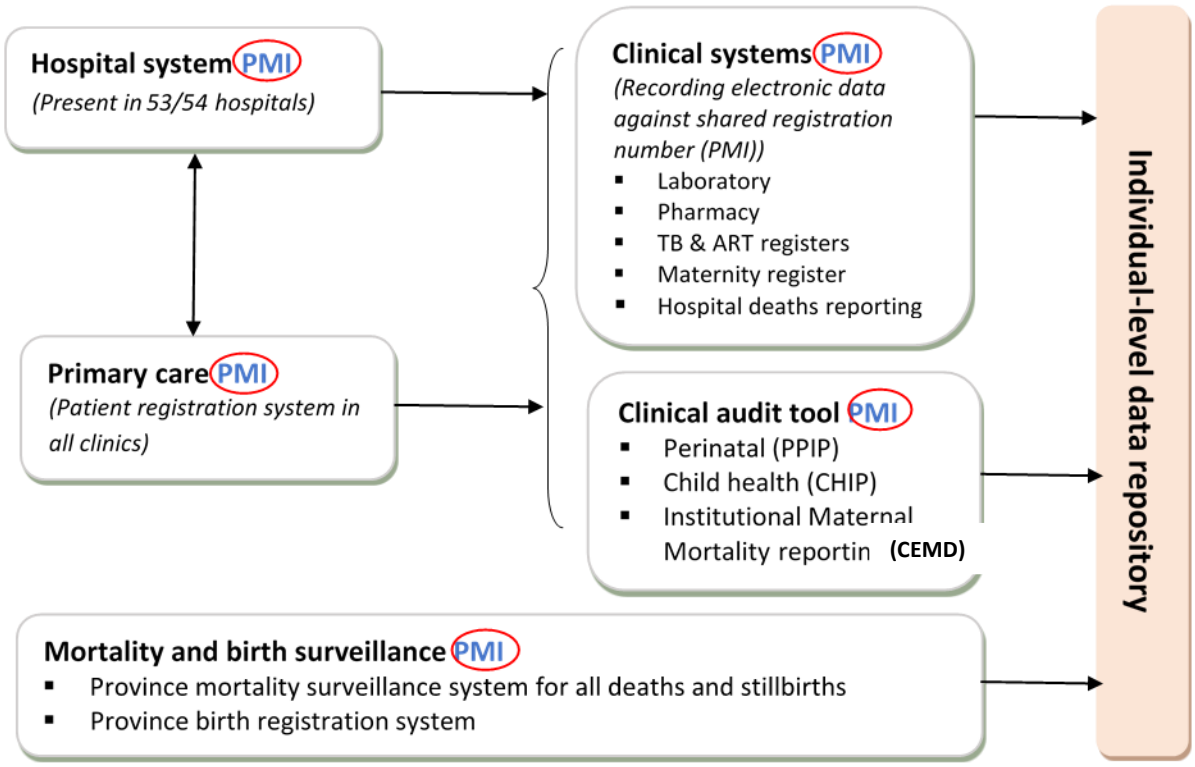


Figure 2: Sharing of unique health identifier (PMI) across multiple domains[a]

2.2.1 Data from the Department of Health

2.2.1.1 Data from the Provincial Health Data Centre (PHDC)

Over the past two decades, the WCDOH has progressively established an electronic health-data system available at an individual patient-level. This involved the assignment of a unique health identifier when patients register at any of the primary care or hospital - provincial or city facilities for the first time. This common patient identifier was issued by the hospital information system (Clinicom™) and referred to as the Clinicom/folder number or patient-master index (PMI). During

PPIP - Perinatal Problem Identification Programme
CHIP - Child Problem Identification Program
CEMD - Confidential Enquiry into Maternal Deaths

this period, patients increasingly shared the same folder number in all services and in all systems, i.e., at public hospitals and primary care clinics; clinical systems (laboratories, disease registries, pharmacies); clinical audit tools (PPIP, CHIP and CEMD) and surveillance systems (mortality and birth). This enabled the patient registration system, when searching for new patients through a web-service to see whether a patient was already registered in the system before issuing a new number. To date, at least 11 million patients have been registered using the PMI through primary care registration systems or the hospital information system[7].

2.2.1.2 Institutional reporting – Confidential Enquiry into Maternal Deaths (CEMD)

In South Africa, the CEMD has been operational since 1998, following the initiation of the National Committee for Confidential Enquiry into Maternal Deaths (NCCEMD) and the policy making maternal deaths notifiable by law (1997)[8]. As one of the clinical audit tools, CEMD performs routine surveillance of maternal deaths at both district and national levels. The enquiry process begins with a discussion at a facility level, where the maternal death has occurred, in order to identify the causes and avoidable factors. Thereafter, the provincial maternal and child health coordinator is notified through a duly filled out maternal death notification form to which is attached the relevant copies of all clinical records. Teams of independent assessors appointed by the province with relevant professional experience (i.e., medical officers, obstetricians, midwives and anaesthetists) assess a case and by using a structured assessor's form, they identify the causes and the avoidable factors. After the assessment process is complete, data are entered into an electronic system known as Maternal Morbidity and Mortality Audit System (MAMMAS) at a provincial level. At a national level, these data are used to produce annual interim reports

and comprehensive triennial reports known as "Saving Mothers Reports"[8]. In the Western Cape province, although the assessments of avoidable factors are strictly confidential, the enumeration of maternal deaths includes the PMI, and these can therefore be linked to data from the clinical system.

2.2.1.3 Clinical audit tool i.e., Perinatal Problem Identification Program (PPIP)

The Perinatal Problem Identification Program (PPIP) is a national audit process of every neonatal death and stillbirth. The audits are digitised through a computer-based standard software package, bearing the same name (PPIP), which facilitates the perinatal and neonatal death's audit, analysis and reporting. This is a user-friendly tool, used for collecting the data on the number of deliveries, stillbirths, early and late neonatal deaths, causes of death, missed opportunities, avoidable factors and substandard care; from maternity and neonatal wards in the public facilities[9].

The programme was designed and initially developed in the 1990s, tested extensively, and subsequently adopted by the National Department of Health (NDOH), as a mandatory facility-based surveillance programme to improve the quality of maternal and child care in the country[10]. The programme informs the regional and national cause of death estimates; and it is used to triangulate and assess the vital registration data, as reported by Statistics South Africa (Stats SA)[11,12]. These data are used to produce annual reports known as "Saving babies Reports"[12–18]. In the WC province, the programme has been adopted as a perinatal audit tool since 2000; and it has recently been implemented in all the public facilities offering delivery services[10]. A recent

report shows that PPIP covers more than 95% of the institutional deliveries captured by the District Health System (DHS) in the WC province, based on comparing the aggregate number of records in each system[14]. Using the shared PMI, these data were linked to those from the clinical system (Figure 2).

2.2.2 Data from the Department of Home Affairs (DHA)

2.2.2.1 Death certificates with known cause-of-death

In South Africa, the registration of deaths is governed by the Births and Deaths Registration Act, 1992 (Act No. 51 of 1992), as amended[19]. Using the death notification form DHA-1663, the Department of Home Affairs register all deaths and stillbirths. Statistics South Africa (Stats SA) routinely collects the dully filled death notification forms from the Department of Home Affairs head office for data processing and analysis as well as report writing and the dissemination. Causes of death statistics are compiled using the 10th revision of the International Classification of Diseases and Related Health Problems (ICD-10) in accordance with the World Health Organization regulations, i.e., every member nation should classify and code causes of death [20].

The amendment to the South African Births and Deaths Registration Act in February 2014 restricted access to the cause-of-death data for confidentiality purposes[19]. This legislation has subsequently limited mortality data linkage across the systems[21]. Due to this limitation, our analysis used DHA data from 2010 to 2013.

2.3 Pregnancy and mortality ascertainment

Evidence from multiple sources (such as facility visits, laboratory and pharmacy records) were used to ascertain pregnancy. These included previous birth records, patient encounters at the maternity ward, antenatal laboratory screening tests, e.g., Rhesus antibodies, ICD10 codes specifying inpatient admissions, or the procedures related to pregnancy, dispensed non-specific pregnancy-related drugs or drugs specific to termination of pregnancy and/or ongoing care specific to pregnancy. The strength of evidence and the combination of multiple evidence were used to obtain confidence scores in the respective episode (pregnancy).

The ascertainment of maternal deaths was done using the following sources: death certificates (vital registration) mentioning pregnancy; hospital deaths reporting indicating maternal deaths, irrespective of this being mentioned on the death certificate, with pregnancy ascertained from multiple sources e.g., admitted in maternity ward, having laboratory results specific to pregnancy or delivery; and institutional reporting (as collected for the confidential enquiry).

The ascertainment of neonatal deaths includes the use of death certificates, hospital deaths reporting and PPIP which records all neonatal deaths along with factors which might have been associated with the outcome.

2.4 Pregnancy and mortality data linkage (PHDC linkage)

Analyses in Chapters 7 and 8 used linked mortality and pregnancy data for ascertaining factors associated with these deaths. The raw pregnancy and mortality data were extracted from separate tables, together with the demographic data, from the PHDC SQL Server database; and

thereafter linked using the provincial linkage algorithm. The PHDC linkage algorithm, also known as provincial linkage, is a decision rule-based deterministic linkage approach that uses fuzzy matching to merge datasets in the province. It includes a set of rules forming criteria to link records. The linkage algorithm categorises records as exact, probable or possible matches, with corresponding confidence match scores, ranging from 1 to 10. In all analyses, records matched with high (>7) "confidence score" were selected and used for further analysis. After ascertaining maternal and neonatal deaths and associated causes, there was a unique opportunity to link known deaths captured by the hospital death reporting system and CEMD/PPIP with deaths reported to the Department of Home Affairs, using the PMI. The PHDC linking criteria are presented in Appendix 2.

2.5 Methods of estimating MMR and NMR

As briefly discussed in Chapter 1, estimation of maternal and neonatal mortality can be done empirically or analytically (through modelling/simulation) (Figure 1). In South Africa, the estimates of maternal and neonatal mortality (as reported by both global agencies and in-country institutional reporting) are mainly ascertained using empirical estimation procedures, with limited applications of analytical estimation techniques such as capture-recapture (CRC) methods[8,15,22–29]. This method is unique in that it statistically operationalises the overlap across sources to estimate the actual size, rather than relying on complete case ascertainment. The details of capture-recapture methods are presented below.

2.5.1 Closed population capture-recapture models: Historical overview and applications

Closed population capture-recapture models were originally developed for wildlife applications and for centuries they have been used to estimate the size of unknown populations. The idea underlying this method was first used in the 1600's by John Graunt to produce statistically based estimates of population size and the effect of the plague in London[30,31]. Subsequently, it was used by Pierre-Simon Laplace in 1786 to estimate the population size of France[31,32].

This approach was developed further by Petersen and Dahl during the period 1860 - 1920 when estimating fish populations[31], and the method has since been extensively developed in the field of wildlife biology. Recently, there has been an increasing application of capture-recapture methodologies in human populations, particularly in medicine, demography and epidemiology; mostly in estimating registration completeness, prevalence and incidence of health conditions/diseases, and for detecting hidden populations[33,34,43,35–42].

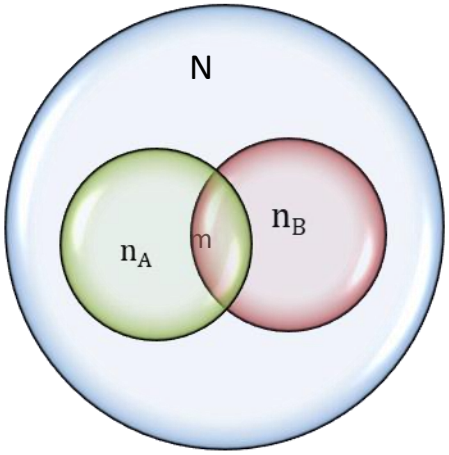
In ecology, capture-recapture involves sampling the respective animal population, uniquely marking or tagging individuals that are caught, releasing them to mix in the population, and after a specified period, taking a new sample. The repeated samples allow one to estimate the probability of detection and consequently, the size of the population of interest. Taking a canonical example whereby a sample of 100 elephants is captured, marked/tagged and released. After allowing them to mix with others, another sample of 100 elephants is captured from the same area of which 10 are re-captures (those previously marked – thereby implying that 1/10th

of elephants in the population are captured); an estimate of the population size $\widehat{N}$ can be obtained as follows: -

$$\widehat{\mathrm{N}} = 100 * \frac{100}{10}$$

Assuming certain assumptions are met, from the above equation, the estimate of elephants in the population is 1000.

In human populations, using the capture-recapture method involves listing individuals, who are registered (captured) on different databases, and identifying those who appear on more than one source (overlaps) to estimate the population size. The basic idea of how the models work can be demonstrated simply with a two-source example. In general, the model is usually estimated with a multinomial likelihood. When having two data sources (samples) namely A and B as follows: -



Whereby n_A is the size of sample A, n_B the size of sample B and m the number of overlapping cases between A and B; the capture-recapture estimate ($\widehat{N}$) is given by the following Petersen-Lincoln equation:

$$\widehat{N} = \frac{n_A \times n_B}{m}$$

To reduce bias due to small sample size, Chapman modified the above Petersen-Lincoln estimator by adding a small number as follows[44]:

$$\widehat{N} = \frac{(n_A + 1)(n_B + 1)}{m + 1} - 1$$

The 95% Confidence Interval (CI) of the population estimate is calculated using the equation:

$$95\%CI = \widehat{N} \pm \sqrt{\left[\frac{(n_A + 1)(n_B + 1)(n_A + m)(n_B + m)}{(m + 1)^2(m + 2)}\right]}$$

2.5.2 Underlying assumptions of CRC

Regardless of the number of data sources to be used for estimation purposes, the application of capture-recapture methods requires a set of assumptions about the population of interest to be met. The key assumptions of traditional capture-recapture methods, as described below, have been presented in different ways by different authors[45–47]. These are:

1. The population is closed i.e., no addition or deletion.
2. Individuals sampled in different sources can be identified and matched.
3. Each individual has an equal chance of being included/captured in each sample i.e., equal catchability probability/capture homogeneity.
4. The presence of an individual in one sample does not influence the presence of the same individual in another sample i.e., independence between sources

2.5.3 Violation of assumptions

In general, violating these assumptions has been associated with under-estimation or over-estimation of the population size. For instance, violating the assumption of a closed population in a capture-recapture approach, leads to over-estimation of the population size[47]. Likewise, if individuals found in a given data source cannot reliably be identified and matched, this leads to over-estimation of the population size due to a reduced number of individuals re-captured, as a result of missing the true matches[47].

The standard capture-recapture method also assumes equal capture probability among individuals in the population. Violation of this assumption occurs when some individuals have low or high probability of being found in each source, leading to an underestimation or overestimation of population size respectively[33,48]. Lastly, the models assume data sources are independent i.e., inclusion of an individual in one source does not influence his/her inclusion in the other data source. If data sources are not independent, the traditional methods can bias (underestimate or overestimate) the true size[46]. A positive dependence leads to underestimation of the true size and vice versa. Violation of independence assumption is much more likely the case with epidemiological compared to ecological data due to the fact that, inclusion of an individual in one source sometimes depends to its inclusion in the other source.

Since it is frequently not possible to meet all assumptions, sophisticated statistical approaches have been developed and widely used for controlling potential bias. When **capture probabilities depend on individual characteristics** (covariates), two methods have been proposed to eliminate bias i.e., the use of multinomial logit models which relate the characteristics of individuals

captured to their probabilities of being captured. The other commonly used approach is based on stratifying data according to the individual characteristics, thereafter estimating stratum sample estimates and finally pooling the estimates to obtain the overall numbers of cases[33,48–50].

Nonetheless, the standard assumptions include independence of samples/sources, and this usually applies in a wildlife population but not in the epidemiological setting. If there is dependence, more than two sources are required so that the **dependence** can be estimated, and consequently an unbiased estimate of population size can be obtained. Two data sources provide insufficient data for estimating dependence statistically since four parameters are required to be estimated (population size, two mean capture probabilities and a measure of dependence)[36,46,51]. In a *multiple system estimation*, various models incorporating dependence among data sources have been proposed, particularly ecological models, log-linear models and the sample coverage approach. As briefly discussed below, the methodology and assumptions of these techniques have been broadly reported elsewhere[34,42,46,52–54].

2.5.4 Estimating dependence

Having dependence between data sources in a dual system estimation is associated with obtaining a biased population estimate[46]. However, under the multiple system estimation the effect of dependencies can be controlled statistically through the use of ecological models, log-linear models and the sample coverage approach[32,39,52,54–60]. Below, are the details of these modelling approaches.

2.5.4.1 Log-linear models

Log-linear models are common methods for analysing discrete data particularly multiway contingency tables. This approach was initially proposed by Fienberg for the application to human populations, and extended by Cormack, to control for dependence between sources[54,58]. The log-linear method models the logarithm of the expected value of each observable category in a contingency table. In its most basic form, this method involves fitting various log-linear models to the observed cells of an incomplete 2^n contingency table (n - number of data sources) whereby the cell corresponding to individuals not listed by any sources is missing. Selection of the model giving the best fit to the data is usually done using the deviance statistic and Akaike Information Criterion (AIC). Once the model with the best fit has been chosen, it is projected onto the unobserved cell, assuming that there are no three data-source (overall) interactions. The interaction terms in the model correspond to the local dependencies between sources.

Assuming that three data sources are used, to control for dependence involves fitting the following log-linear model[46]:

$$logE(Z_{ijk}) = u + u_1 I(i = 1) + u_2 I(j = 1) + u_3 I(k = 1) + u_{12} I(i = j = 1) + u_{13} I(i = k = 1) + u_{23} I(j = k = 1) + u_{123} I(i = j = k)$$

Where:

$logE(Z_{100}) = u + u_1$

$logE(Z_{010}) = u + u_2$

$logE(Z_{001}) = u + u_3$

$logE(Z_{110}) = u + u_1 + u_2 + u_{12}$

$logE(Z_{101}) = u + u_1 + u_3 + u_{13}$;

$$logE(Z_{011}) = u + u_2 + u_3 + u_{23}$$

$$logE(Z_{111}) = u + u_1 + u_2 + u_3 + u_{12} + u_{13} + u_{23} + u_{123}$$

Assuming $u_{123} = 0$, allows the estimation of the missing cases as a function of the fitted values of other categories. Whereby, the fitted values are determined by the log-linear models.

$$\hat{z}_{000} = \frac{\hat{z}_{001} * \hat{z}_{010} * \hat{z}_{100} * \hat{z}_{111}}{\hat{z}_{110} * \hat{z}_{011} * \hat{z}_{101}}$$

The independent model i.e., the model including only main effects is estimated using the following equation: -

$$logE(Z_{ijk}) = u + u_1 I(i = 1) + u_2 I(j = 1) + u_3 I(k = 1)$$

The details and applications of log-linear models for estimating dependence between sources have been widely discussed elsewhere[32,34,46,48,52,54,59,60].

2.5.4.2 Ecological models

In order to relax the "equal catchability assumption" of the capture-recapture method, Pollock proposed a set of models that specify different forms of capture probabilities for biology and wildlife applications[61]. These included: a null model with no source of variation (subscript *0*), and seven models incorporating up to three sources of variation in capture probabilities, as depicted below i.e., temporal effects – where there is variation in capture probabilities, due to time (subscript *t*), behavioural response - capture probabilities, depending on previous capture history (subscript *b*) and heterogeneity between individuals (subscript *h*)[62]. Except model M_t and the null model ,which assumes that samples are independent, model M_h accounts for heterogeneity in

capture probabilities; models M_b and M_{tb} account for local dependencies, as well as model M_{bh} and M_{tbh} accounts for both types of dependency.[46,62]

When applying the method to human populations, models incorporating behavioural responses (M_b, M_{bh} and M_{tbh}) have limited use, due to the lack of a sequential order in the data sources/lists[63]; therefore, they are not appropriate for applications to epidemiology. Notwithstanding this, models allowing for heterogeneous capture probabilities (M_h, M_{bh} and M_{tbh}) are recommended, only when at least five sources/lists are available. Model M_t is the same as a log-linear model with only the main effects; consequently, they are expected to produce the same estimates, when fitted to the same data. Table 5 depicts the types and specifications of the ecological models.

Table 5: Closed population ecological models according to types of capture probabilities

SN	Model	Log-linear/Logit models	Estimation method
1	M_0	$log(P_{ij}) = \alpha$ $logit(P_{ij}) = \alpha$	All estimators listed below
2	M_t	$log(P_{ij}) = \beta_j$ $logit(P_{ij}) = \beta_j$	Binomial models/ MLE/Conditional MLE
3	M_b	$log(P_{ij}) = \alpha + \gamma Y_{ij}$ $logit(P_{ij}) = \alpha + \gamma Y_{ij}$ $(\alpha_i \equiv \alpha)$	MLE/Martingale method/Log-linear models
4	M_h	$log(P_{ij}) = \alpha_i$ $logit(P_{ij}) = \alpha_i$	Non-parametric MLE/Bootstrap estimator/Martingale method/ Jack-knife
5	M_{th}	$log(P_{ij}) = \alpha_i + \beta_j$ $logit(P_{ij}) = \alpha_i + \beta_j$	Sample coverage/mixed logit models/ log-linear models/latent class models
6	M_{tb}	$log(P_{ij}) = \beta_j + \gamma Y_{ij}$ $logit(P_{ij}) = \beta_j + \gamma Y_{ij}$	Martingale method/MLE and maximum quasi likelihood/Bayes approach/log-linear models
7	M_{bh}	$log(P_{ij}) = \alpha_i + \gamma Y_{ij}$ $logit(P_{ij}) = \alpha_i + \gamma Y_{ij}$	Jack-knife/Sample coverage/Non-parametric MLE/generalised removal
8	M_{tbh}	$log(P_{ij}) = \alpha_i + \beta_j + \gamma Y_{ij}$ $logit(P_{ij}) = \alpha_i + \beta_j + \gamma Y_{ij}$	MLE using mixture/GLM with covariates /Conditional sample coverage/Estimating functions

2.5.4.3 Sample coverage approach

The sample coverage approach was initially developed and proposed by Turing and Good for estimations of animal population size[64,65]. The method enables the quantification of both the degree of overlap and the degree of dependence, which leads to a valid estimate of population size. Modelling dependence in this approach involves the use of the Coefficient of Covariation (CCV)[46,63,66]. For a three-source scenario, an estimate of population size based on the sample coverage approach when assuming independence (no local dependence and heterogeneity exists) is given by: -

$$\widehat{N}_0 = \frac{\mathrm{D}}{\hat{C}}$$

Where:

$D = M - \frac{1}{3}(Z_{100} + Z_{010} + Z_{001})$ → Average of the overlapped cases

M → Number of individuals ascertained in at least one list

$\hat{C} = 1 - \frac{1}{3}\left(\frac{Z_{100}}{n_1} + \frac{Z_{010}}{n_2} + \frac{Z_{001}}{n_3}\right)$ → An estimator of sample coverage

When the estimated sample coverage (overlapping information) is large enough (at least 55%)[67] and dependence between sources exists, an estimate of population size is given by: -

$$\widehat{N} = \left[\frac{Z_{+11} + Z_{1+1} + Z_{11+}}{3\hat{C}}\right] \Big/ \left\{1 - \frac{1}{3\hat{C}}\left[\frac{(Z_{1+0} + Z_{+10})Z_{11+}}{n_1 n_2} + \frac{(Z_{10+} + Z_{+01})Z_{1+1}}{n_1 n_3} + \frac{(Z_{01+} + Z_{0+1})Z_{+11}}{n_2 n_3}\right]\right\}$$

When dependence exists and the estimated sample coverage (overlapping information) is too low, the 'one-step' estimator $\widehat{N}_1$ (i.e., obtained through one iterative step) is usually used; since the data contain insufficient information to estimate the population size accurately.[46,66] Under this scenario, the estimate of population size is given by the formula below: -

$$\widehat{N}_1 = \frac{\mathrm{D}}{\hat{C}} + \frac{1}{3\hat{C}}\left[(Z_{1+0} + Z_{+10})\hat{\gamma}_{12} + (Z_{10+} + Z_{+01})\hat{\gamma}_{13} + (Z_{01+} + Z_{0+1})\hat{\gamma}_{23}\right]$$

Where:

$\hat{\gamma}_{12} = \widehat{N}' \frac{Z_{11+}}{n_1 n_2} - 1$; $\hat{\gamma}_{13} = \widehat{N}' \frac{Z_{1+1}}{n_1 n_3} - 1$; $\hat{\gamma}_{23} = \widehat{N}' \frac{Z_{11+}}{n_2 n_3} - 1$ → CCV estimates

and

$$\widehat{N}' = \frac{\mathrm{D}}{\hat{C}} + \frac{1}{3\hat{C}}\left[\left(Z_{1+0} + Z_{+10}\right)\left(\frac{\mathrm{D}}{\hat{C}}\frac{Z_{11+}}{n_1 n_2} - 1\right) + \left(Z_{10+} + Z_{+01}\right)\left(\frac{\mathrm{D}}{\hat{C}}\frac{Z_{1+1}}{n_1 n_3} - 1\right) + \left(Z_{01+} + Z_{0+1}\right)\left(\frac{\mathrm{D}}{\hat{C}}\frac{Z_{11+}}{n_2 n_3} - 1\right)\right]$$

The 'one-step' estimator is used as an upper bound for negatively dependent samples as well as a lower bound for positively dependent samples[46]. However, most epidemiological datasets are known to have a net positive dependence[68]. Therefore, the 'one-step' estimator is usually used as a lower bound. A log-transformation is used to construct the confidence intervals with a bootstrap resampling method proposed to obtain the estimated standard errors[69,70].

2.6 Capture-recapture modelling

The analyses of Chapters 5 and 6 fitted three-source capture-recapture models to estimate maternal and neonatal mortality under-reporting respectively. Before applying capture-recapture methods, all the assumptions underlying the method were assessed. The evaluation of dependence between sources was done by fitting the log-linear models, ecological models and sample-coverage approaches. The model giving the best fit to the combined data (full four-years) was selected and fitted to the annual data, in order to provide comparable annual estimates. The capture-recapture modelling accounting for dependence was conducted using three different software implementations (R, Stata and CARE1) for comparison purposes.

2.7 Maternal and neonatal mortality estimation

After adjusting for under-reporting, maternal and neonatal deaths were estimated annually and combined using the capture-recapture method. Exploratory heterogeneity graphs (fi plot) and boxplots were used to graphically explore Pearson residuals of the log-linear models. Evaluation of heterogeneity in capture probabilities and dependence between data sources involved fitting three classes of models i.e., ecological models, log-linear models and the sample coverage

approach. Assessing the goodness-of-fit of the fitted models involved the use of likelihood ratio statistics (G^2), the Pearson goodness-of-fit statistics (χ^2) and distribution of the Pearson residuals as presented by boxplots. The model with the fewest number of parameters, giving the best fit to the data (based on G^2 and χ^2 statistics), having the lowest Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) was selected and used for estimation of maternal mortality underreporting. The reported 95% CIs are profile likelihood intervals since intervals based on asymptotic theory are not expected to perform well and so profile likelihoods are used.

2.8 Estimation of completeness

Completeness was estimated for each individual data source and combined, as a proportion of the estimated maternal/neonatal deaths. Completeness of a source was obtained by dividing the number of maternal/neonatal deaths registered by that source over the total number of deaths estimated by the capture-recapture method respectively. The overall completeness level was also obtained by dividing the number of unique maternal/neonatal deaths registered by the three sources over the estimated maternal/neonatal deaths from the best-fitting log-linear capture recapture model.

2.9 Multivariable modelling

Chapters 7 and 8 used the final linked dataset to determine independent predictors of maternal and neonatal mortality, based on the final linked dataset of pregnant women and births, and on *a priori* specification of potential associations and causal pathways. Two predictive models were fitted, with the first model fitting deaths only registered by the CEMD (PPIP for neonatal mortality estimation) while the other model fitted deaths registered by all sources in the province i.e.,

hospital death reporting, death certificates and CEMD (PPIP for neonatal mortality). We anticipated this would give a snapshot of a possible different pattern of factors associated with maternal and neonatal mortality when deaths from other sources are ignored since CEMD/PPIP is often treated as a more reliable and accurate source of maternal/neonatal deaths data in the province and country at large.

2.10 Missing data

Due to the routine nature of data used, not all cases had all required information owing to insufficient documentation in most secondary data. Therefore, the analyses have been limited to cases with available data. Analyses were conducted to compare whether those who had missing data were different from those who did not, and the potential for bias resulting from differential missingness of data.

2.11 References

1. Coovadia, H., Jewkes, R., Barron, P., Sanders, D. & Mcintyre, D. The health and health system of South Africa: historical roots of current public health challenges. *Lancet* **374**, 817–834 (2009).

2. Statistics South Africa. *Statistical release P0302: Mid - year population estimates 2017*. (2017).

3. National Department of Health (NDoH), Statistics South Africa Council (Stats SA), South African Medical Research (SAMRC) & ICF. *South Africa Demographic and Health Survey 2016: Key Indicators.* (2017).

4. Western Cape Government. *City of Cape Town Socio-Economic profile*. (2016).

5. Statistics South Africa. *Financial statistics of provincial government 2015/2016*. (2017).

6. Western Cape Provincial Health Data Centre. *Interoperability to support patient care the Western Cape Provincial Health Data Centre*. (2018).

7. Western Cape Provincial Centre Health Data Centre. *Strategic approach to patient - lvel health data harmonisation and integration Western Cape Department of Health International Classification of Diseases*. (2013).

8. Moodley, J. *et al.* The confidential enquiry into maternal deaths in South Africa: a case study. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 53–60 (2014).

9. MRC Unit for Maternal and Infant Health Care Strategies. The Perinatal Problem Identification Program (PPIP). *PPIP v3* (2004). Available at: https://www.ppip.co.za/contact-us/. (Accessed: 28th June 2018)

10. Rhoda, N. *et al.* Experiences with perinatal death reviews in South Africa - the Perinatal Problem Identification Programme: scaling up from programme to province to country. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 160–166 (2014).

11. Kerber, K. J. *et al.* South African child deaths 1990-2011: have HIV services reversed the trend enough to meet Millennium Development Goal 4? *AIDS* **27**, 2637–48 (2013).

12. Pattinson, R. C. (Mr. R. U. for M. and I. H. C. S. on B. of P. U. *Saving Babies 2010-2011 : Eighth report on perinatal care in South Africa*. (2013).

13. National perinatal morbidity and mortality committee. *National Perinatal Morbidity and Mortality Committee Report 2014*. (2014).

14. Pattinson, R. & Rhoda, N. *Saving Babies 2012-2013: Ninth report on perinatal care in South Africa*. *RC Pattinson* (2014).

15. Chopra, M., Daviaud, E., Pattinson, R., Fonn, S. & Lawn, J. E. Saving the lives of South Africa's mothers, babies, and children: Can the health system deliver? *Lancet* **374**, 835–846 (2009).

16. Health, I. & Strategies, C. *Saving Babies 2001 2nd Perinatal Care Survey of South Africa*. (2001).

17. Pattinson, R. C. *et al. Saving Babies 2002 Third Perinatal Care Survey of South Africa*. (2002).

18. Pattinson, R. C. & Tlebere, P. M. *The how, when, where, what and why babies die in South Africa*. *Saving babies 2003: Fourth perinatal care survey of South Africa* (2003).

19. Republic of South Africa. Births and Deaths Registration Act No. 51 of 1992. Regulations on the Registration of Births and Deaths. *Government Gazette* 74 (2014).

20. WHO. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. (2016).

21. Groenewald, P. *et al.* The importance of identified cause-of-death information being available for public health surveillance, actions and research. *South African Med. J.* **105**, 528–530 (2015).

22. Bradshaw, D. *et al.* Every death counts: use of mortality audit data for decision-making to save the lives of mothers, babies, and children in South Africa. *Lancet* **371**, 1294–1304 (2008).

23. Souza, J. P. *et al.* Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *The Lancet* **381**, 1747–1755 (2013).

24. UNICEF, WHO, World Bank group & United Nations. *Levels & Trends in Child Mortality - Report 2015: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. *UNICEF* (2015). doi:10.1371/journal.pone.0144443

25. WHO, UNICEF, UNFPA, World Bank & United Nations Population Division. Trends in maternal mortality 1990 to 2015. (2015). doi:10

26. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384**, 980–1004 (2014).

27. Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 957–979 (2014).

28. Saleem, S. *et al.* A prospective study of maternal, fetal and neonatal deaths in low- and middle-income countries. TT -. *Bull. World Health Organ.* **92**, 605–612 (2014).

29. UNICEF, WHO, World Bank, U. *Levels & Trends in Child Mortality in 2014- Report: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. (2014).

30. Amstrup, S. C., McDonald, T. L. & Manly, B. F. J. *Handbook of capture-recapture analysis*. (Princeton University Press, 2005).

31. Goudie, I. B. J. & Goudie, M. Who captures the marks for the Petersen estimator? *J. R. Stat. Soc. Ser. A (Statistics Soc.* **170**, 825–839 (2007).

32. Cormack, R. . The statistics of capture-recapture methods. *Ocean. Mar Biol Ann Rev.* 455–506 (1968).

33. Hook, E. B. & Regal, R. R. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* **17**, 243–264 (1995).

34. International Working Group for Disease Monitoring and Forecasting. Capture-Recapture and Multiple-Record Systems Estimation II: Applications in Human Diseases. *Am. J. Epidemiol.* **142**, 1059–1068 (1995).

35. Héraud-Bousquet, V. *et al.* A three-source capture-recapture estimate of the number of new HIV diagnoses in children in France from 2003-2006 with multiple imputation of a variable of heterogeneous catchability. *BMC Infect. Dis.* **12**, 251 (2012).

36. Hay, G. The selection from multiple data sources in epidemiological capture-recapture studies. *J. R. Stat. Soc. Ser. D (The Stat.* **46**, 515–520 (1997).

37. McCarty, D. J., Tull, E. S., Moy, C. S., Kwoh, C. K. & LaPorte, R. E. Ascertainment corrected rates: applications of capture-recapture methods. *Int. J. Epidemiol.* **22**, 559–65 (1993).

38. Aghaei, A., Ahmadi-Jouibari, T., Baiki, O. & Mosavi-Jarrahi, A. Estimation of the gastric cancer incidence in Teheran by two- source capture-recapture. *Asian Pac. J. Cancer Prev.* **14**, 673–677 (2013).

39. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20**, 3123–57 (2001).

40. Abubakar, I. Assessing the coverage of TB notification using capture recapture methods.

41. Coeli, C. M., Veras, R. P. & da Silva Freire Coutinho, E. [Capture-recapture methodology: an option for surveillance of non-communicable diseases in the elderly]. *Cad. Saude Publica* **16**, 1071–82

42. Domingo-Salvany, a *et al.* Analytical considerations in the use of capture-recapture to estimate prevalence: case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *Am. J. Epidemiol.* **148**, 732–740 (1998).

43. Qomariyah, S. N. *et al.* An option for measuring maternal mortality in developing countries: a survey using community informants. *BMC Pregnancy Childbirth* **10**, 74 (2010).

44. Chapman, D. G. *Some properties of the hypergeometric distribution with applications to zoological sample censuses*. *Univ. California Publ. Stat.* **1**, (1951).

45. Rossi, G., Pepe, P., Curzio, O. & March, M. Generalized Linear Models and Capture-Recapture Method in a Closed Population: Strengths and Weaknesses.

46. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20**, 3123–3157 (2001).

47. VALID International Ltd. *Notes on using capture-recapture techniques to assess the sensitivity of rapid case-finding methods*. (2006)

48. Hook, E. B. & Regal, R. R. Effect of variation in probability of ascertainment by sources ('variable catchability') upon 'capture-recapture' estimates of prevalence. *Am.J.Epidemiol.* **137**, 1148–1166 (1993).

49. Alho, J. M. Logistic regression in capture-recapture models. *Biometrics* **46**, 623–35

(1990).

50. Tilling, K. & Sterne, J. a. Capture-recapture models including covariate effects. *Am. J. Epidemiol.* **149**, 392–400 (1999).

51. Ismail, A. How many data sources are needed to determine diabetes prevalence by capture-recapture? *Int. J. Epidemiol.* **29**, 536–541 (2000).

52. Yip, P. *et al.* Capture-recapture and multiple-record systems estimation I: History and theoretical development. (1995).

53. Simondon, F. & Khodja, H. Capture-recapture method for estimating misclassification errors: Application to the measurement of vaccine efficacy in randomized controlled trials. *Int. J. Epidemiol.* **28**, 113–116 (1999).

54. Fienberg, S. E. The Multiple Recapture Census for Closed Populations and Incomplete 2 k Contingency Tables. *Biometrika* **59**, 591 (1972).

55. Oosterlee, A., Vink, R. M. & Smit, F. Prevalence of family violence in adults and children: estimates using the capture-recapture method. *Eur. J. Public Health* **19**, 586–91 (2009).

56. Nance, M. L., Carr, B. G., Kallan, M. J., Branas, C. C. & Wiebe, D. J. Variation in pediatric and adolescent firearm mortality rates in rural and urban US counties. *Pediatrics* **125**, 1112–8 (2010).

57. Lum, K., Price, M., Guberek, T. & Ball, P. Measuring Elusive Populations with Bayesian Model Averaging for Multiple Systems Estimation: A Case Study on Lethal Violations in Casanare, 1998-2007. *Stat. Polit. Policy* **1**, (2010).

58. Cormack, R. M. Log-Linear Models for Capture-Recapture. *Biometrics* **45**, 395 (1989).

59. Coull, B. A. & Agresti, A. The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies. *Biometrics* **55**, 294–301 (1999).

60. Agresti, A. Simple Capture-Recapture Models Permitting Unequal Catchability and Variable Sampling Effort. *Biometrics* **50**, 494 (1994).

61. Pollock, K. H. Review Papers: Modeling Capture, Recapture, and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: Past, Present, and Future. *J. Am. Stat. Assoc.* **86**, 225–238 (1991).

62. Baillargeon, S. & Rivest, L.-P. Rcapture : Loglinear Models for Capture-Recapture. *J. Stat. Softw.* **19**, 1–31 (2007).

63. Chao, A. Capture-Recapture for Human Populations. *Wiley StatsRef Stat. Ref. Online* 1–16 (2015). doi:10.1002/9781118445112.stat04855.pub2

64. Good, I. J. Turing's anticipation of empirical bayes in connection with the cryptanalysis of the naval enigma. *J. Stat. Comput. Simul.* **66**, 101–111 (2000).

65. GOOD, I. J. The population frequencies of species and the estimation of population parameters. *Biometrika* **40**, 237–264 (1953).

66. Chao, A. & Tsay, P. K. A Sample Coverage Approach to Multiple-System Estimation with Application to Census Undercount. *J. Am. Stat. Assoc.* **93**, 283–293 (1998).

67. Chao, A., Tsay, P. K., Shau, W. Y. & Chao, D. . Population size estimation for capture-recapture models with applications to epidemiological data. *Proc. Biometrics Sect. Am. Stat. Assoc.* 108–117 (1996). doi:10.1080/02664760120011572

68. Hook, E. B. & Regal, R. R. Accuracy of alternative approaches to capture-recapture estimates of disease frequency: internal validity analysis of data from five sources. *Am. J. Epidemiol.* **152**, 771–9 (2000).

69. Efron, B. & Tibshirani, R. *An introduction to the bootstrap*. (Chapman & Hall, 1993).

70. Chao, A. Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. *Biometrics* **43**, 783 (1987).

CHAPTER 3: Trends in maternal and neonatal mortality in South Africa: A systematic review

The chapter comprises the pre-publication version of the second paper, with additional edits for style, coherence with the other chapters, and clarity. The contents of this chapter have resulted in the following publications: -

3. **Damian, D. J.,** Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review. *Syst. Rev.* (2019).

4. **Damian, D. J.,** Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review protocol. *Syst. Rev.* **6**, 165 (2017).

3.1 Abstract

Background: Measuring and monitoring progress towards Millennium Development Goals (MDG) 4 and 5 required valid and reliable estimates of maternal and child mortality. In South Africa, there are conflicting reports on the estimates of maternal and neonatal mortality, derived from both direct and indirect estimation techniques. This chapter aimed to systematically review the estimates of maternal and neonatal mortality in the period from 1990 to 2015 in South Africa; and determine trends over this period.

Methods: National representative studies reporting on maternal and neonatal mortality in South Africa were included for synthesis. Searches for eligible studies were conducted in five databases: Medline, Africa-Wide Information, Scopus, Web of Science and CINAHL. Searches were restricted to articles written in English and presenting data covering the period between 1990 and 2015. Reference lists of retrieved articles were screened for additional publications; and grey literature was searched for relevant documents for the review. Three independent reviewers were involved in study selection, data extractions and for achieving consensus.

Results: In total, 969 studies were retrieved and 670 screened for eligibility, yielding 25 studies reporting the data on maternal mortality; and 14 studies on neonatal mortality. Most of the studies had low risk of bias. Estimates from the institutional reporting differed from the estimates included in international metrics. Modelled estimates were widely divergent from the estimates obtained through empirical methods. In the last two decades, both maternal and neonatal mortality appear to have increased up to 2009, followed by a decrease, which was more pronounced in the case of maternal mortality.

Conclusion: Estimates from both global agencies and in-country institutional reporting, although widely divergent, indicate that South Africa has not achieved MDG 4a and 5a goals; but it has made significant progress in reducing maternal and neonatal mortality. To obtain more accurate estimates, it may be possible to apply additional estimation techniques, which utilise multiple data sources to correct for the under-reporting of these outcomes.

Systematic review registration: PROSPERO CRD42016042769

Keywords: Maternal mortality, neonatal mortality, Millennium development goals, South Africa.

3.2 Background

Monitoring progress towards MDG 4 and 5 (reducing child and maternal mortality between 1990 and 2015) requires valid, reliable and internationally comparable estimates of maternal and child mortality in the country. Various methods for measuring and estimating maternal and child mortality have been developed, tested, and widely used[1–8]. Estimating these outcomes in developing countries is challenging; due to the lack of accurate, valid and reliable data[8–14].

Recent estimates from the United Nations Inter-agency Group for Child Mortality Estimation (UN-IGME) and Maternal Mortality Estimation Inter-agency Group (MMEIG) indicated that South Africa did not achieve the MDG 4a and 5b targets by 2015 (reducing by three-quarters the maternal mortality ratio (MMR) and reducing by two-thirds the under-five mortality rate in the period between 1990 and 2015 respectively)[9,14]. The country had conflicting estimates of maternal and neonatal mortality reported by different sources, with wide uncertainty intervals[9,14–18].

South Africa is unusual among developing countries in that, national facility-based mortality audits are carried out for maternal, perinatal and child deaths[19,20]. The estimation of maternal and neonatal mortality in the country is often based on the vital registration, National Confidential Enquiry into Maternal Deaths (NCEMD), which records maternal deaths, and the Perinatal Problem Identification Program (PPIP) ,which records stillbirths and neonatal deaths[19–22]. The CEMD data provides the maternal deaths from the routine surveillance of maternal deaths at a facility level; whereas vital registration data derive the deaths, from the causes of deaths, as well as surveys and censuses that provide maternal deaths from the collected pregnancy-related

data at a household level. Nonetheless, Stats SA are the custodians of vital registration and the Department of Health are the custodians of the confidential enquiries.

The country provides unique opportunities to estimate these outcomes empirically, analytically or through modeling, by having multiple data sources with wide coverage[1–5,21,23,24]. However, there are widely divergent estimates, wherein the two most frequently cited estimates are from institutional reporting and WHO metrics, which makes it difficult to understand the trends in these outcomes, and to assess the successes or failures of interventions focusing on reducing maternal and child mortality in the country over the past decades. The reasons for divergent estimates between institutional reporting and WHO metrics, or among global metrics, can be explained by estimation approaches, sources and quality of the data used[23,25,26].

Monitoring maternal and neonatal mortality in South Africa over the past two decades is of high importance, given the introduction of the Termination of Pregnancy Act in 1996, which has reduced the extent of abortion related maternal morbidity and mortality as well as the context of high HIV prevalence and its associated mortality in women during pregnancy and childbirth[27,28]. There has also been a massive uptake of HIV treatment and the Prevention of Mother-to-Child Transmissions (PMTCT) of HIV, which currently stands at over 90% by some estimates[29,30].

There have been limited attempts to review maternal and neonatal mortality estimates in South Africa to facilitate understanding of trends during the MDG period. This review is expected to provide the context for understanding inconsistencies in the reported estimates of maternal and neonatal mortality by the in-country institutional reporting and the global agencies, by

ascertaining estimation methods, data sources and quality, the sampling methods and definitions used, to better inform comparisons across such estimates. It will also create a framework against which to compare the estimates generated for the Western Cape Province in subsequent chapters.

Aim

This review aimed to synthesise estimates of maternal and neonatal mortality for the period 1990 to 2015 in South Africa and to determine temporal trends during this period.

3.3 Methods

Protocol and registration

The review protocol was registered with the PROSPERO database in 2016, with a registration number CRD42016042769; and it has already been published[31]. The presentation and reporting of the results in this review followed the systematic review reporting standard (PRISMA-P)[32]. To ensure transparency, a PRISMA flow chart was used and a table indicating all the included studies was presented[33].

Eligibility criteria

The population for eligible studies included pregnant women and neonates for ascertaining maternal and neonatal mortality respectively. All studies that are nationally representative, reports providing national-level data (and the trends thereof) and vital registration data were eligible for this review. Searches were restricted to studies written in English and being conducted in South Africa or which have used South African data, and multicentre studies including South

Africa, reporting data covering the period 1990 to 2015. No restrictions on the date of publication were made, in order to include articles reporting data from 1990 – 2015, which were published after 2015.

Information sources

Separate searches for the two outcomes (maternal and neonatal mortality) were conducted in the following electronic databases: Medline, Africa-Wide Information, Scopus, Web of Science and CINAHL. The last search was carried out on 18th August 2017. No restrictions on the date of publication were made. Additional searches for conference abstracts and proceedings were made. Reference lists of retrieved articles were also screened for additional publications. Reports by government or other agencies were included based on publications and a number of data sources reported by them were included. Contacts with experts in the field of study were made, in order to identify any additional relevant articles.

Search

The searches in the fore-mentioned electronic databases were conducted from August 2016 to August 2017. All searches were restricted to articles written in English. In particular, the search strategy used in Medline database was as follows: (("mothers"[MeSH Terms] OR "mothers"[All Fields] OR "maternal"[All Fields]) OR ("infant, newborn"[MeSH Terms] OR ("infant"[All Fields] AND "newborn"[All Fields]) OR "newborn infant"[All Fields] OR "neonatal"[All Fields])) AND (("mortality"[Subheading] OR "mortality"[All Fields] OR "mortality"[MeSH Terms]) OR ("death"[MeSH Terms] OR "death"[All Fields])) AND (estimation[All Fields] OR estimates[All Fields]) AND ("South Africa"[Mesh] OR ("south africa"[MeSH Terms] OR ("south"[All Fields] AND

"africa"[All Fields]) OR "south africa"[All Fields])) AND (("1990/01/01"[PDAT]: "3000/12/31"[PDAT]) AND "humans"[MeSH Terms] AND English[lang])".

Study selection

Search outputs were managed in EndNote reference manager. Any duplicate records were removed before the screening process took place. When the same article was captured in different journals, or the same results were presented with different main authors, the most detailed publications were selected for review. Three independent reviewers were involved in the screening and selection of articles to be included in a quantitative (narrative) synthesis. This involved an assessment of the articles based on titles and abstracts, and a full text review, using Covidence software (https://www.covidence.org/). For an article to be eligible for inclusion in the systematic review, two reviewers had to agree to include it. A third reviewer was consulted in case of any difference of opinion between the two reviewers. This followed, whenever they failed to reach a consensus after a joint examination of the different estimate of those which required third review.

Data collection process

Analysis of the full text was conducted for all eligible articles. Two authors extracted data independently using a pre-agreed data abstraction template. In the case of discrepancies in the extracted data between authors, consensus was sought before involving a third author for resolving the disparities. During the data extraction process, study authors/investigators were contacted to provide extra information when there were insufficient information/data reported in the article.

Data items

The following information were extracted for eligible studies: first author's name; year of publication; year of death (maternal and neonatal); number of pregnant women; number of live-births; maternal deaths; neonatal deaths; definition of maternal death; definition of neonatal death; maternal mortality ratio/rate (if reported, and by year); the neonatal mortality rate (if reported, and by year), sampling method, estimation method used, and an indicator variable showing whether the records were complete.

The main outcomes in this review were maternal and neonatal mortality. Maternal death/mortality was defined as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes[34]. Maternal mortality ratio (MMR) was defined as the number of maternal deaths per 100,000 livebirths.

Neonatal death/mortality was also defined as the death of live-born within the first 28 days of life. Neonatal mortality rate was defined as the number of infant deaths within the first 28 days of life per 1,000 livebirths.

Risk of bias in individual studies

Assessment of risk of bias was done at a study and outcome level. Two authors assessed study quality based on the following quality assessment criteria: 1) definition of maternal mortality; 2) definition of neonatal deaths; 3) completeness of ascertainment of maternal and neonatal

mortality; 4) completeness of ascertainment of live births; 5) sampling technique/design; and 6) data quality. Studies were assessed based on each criterion and were rated as "high risk of bias" or "low risk of bias" accordingly. Studies rated as "high risk of bias" on any criterion were assigned an overall rating of "high risk of bias"; while the overall rating of "low risk of bias" was only assigned in studies with "low risk of bias" in all criteria. For model-based estimations, risk of bias was assessed based on the input data used. Reports by government and other agencies such as Stat SA, National Department of Health and WHO were assessed using similar criteria as empirical studies. Table 6 shows the assessment criteria of risk of bias in individual studies.

Table 6: Risk of bias assessment criteria for individual studies

No.	Criteria	Attributes	Risk of bias
1.	Definition of maternal mortality	❖ ICD-10 maternal death definition[35], or similar	Low
		❖ No or unclear definition provided	High
2.	Definition of neonatal mortality	❖ Death of live-born within the first 28 days of life, or similar	Low
		❖ None, or unclear definition provided	High
3.	Completeness of ascertainment of maternal and neonatal deaths	❖ Prospective recording of mortality data, ❖ Mixed methods cross-referencing facility records, ❖ Demographic surveillance system with frequent rounds, ❖ Survey based on recall of maternal or neonatal deaths ≤ 6 months previously.	Low
		❖ Survey using direct or indirect sisterhood estimation methods, ❖ Demographic surveillance system with infrequent rounds.	High
4.	Completeness of ascertainment of live births	❖ Prospective recording of births data, ❖ Use of census <5 years old for live births.	Low
		❖ Use of census ≥5 years old for live births. ❖ Live births data source not stated or unclear	High
5.		❖ Census,	Low

	Sampling technique/design	❖ Vital registration, ❖ Survey using nationally representative sample. ❖ Systematic analysis involving the use of data collected from the above method(s)	
		❖ Design or sampling techniques not stated or unclear, ❖ Provincial or sub-national sample used.	High
6.	Data quality	❖ Data provide enough information for the study	Low
		❖ Insufficient data provided or unclear	High

Summary measures

Data were presented as ratios for maternal mortality and rates for neonatal mortality with their corresponding confidence or uncertainty intervals.

Synthesis of results

Data were entered and analysed using STATA software version 14.1 (Stata Corp, College Station, Texas). Data were presented as MMR or NMR, in tables and graphs to depict trends over time. The reasons for study exclusions were clearly documented.

3.4 Results

Study selection

As presented in Figure 3 below, a total of 948 studies were identified through the literature search and 21 additional studies were identified through screening of reference lists. After removing the duplicates, 670 studies were screened for eligibility. A total of 608 studies were excluded after screening the titles and abstracts since they reported irrelevant information. Sixty-two abstracts were shortlisted for full-text review and 39 studies met the inclusion criteria for

analysis. Of studies included in the review, 25 reported data on maternal mortality[28,36,45–54,37,55,56,38–44] and 14 on neonatal mortality[14,23,62–65,41,46,56–61].

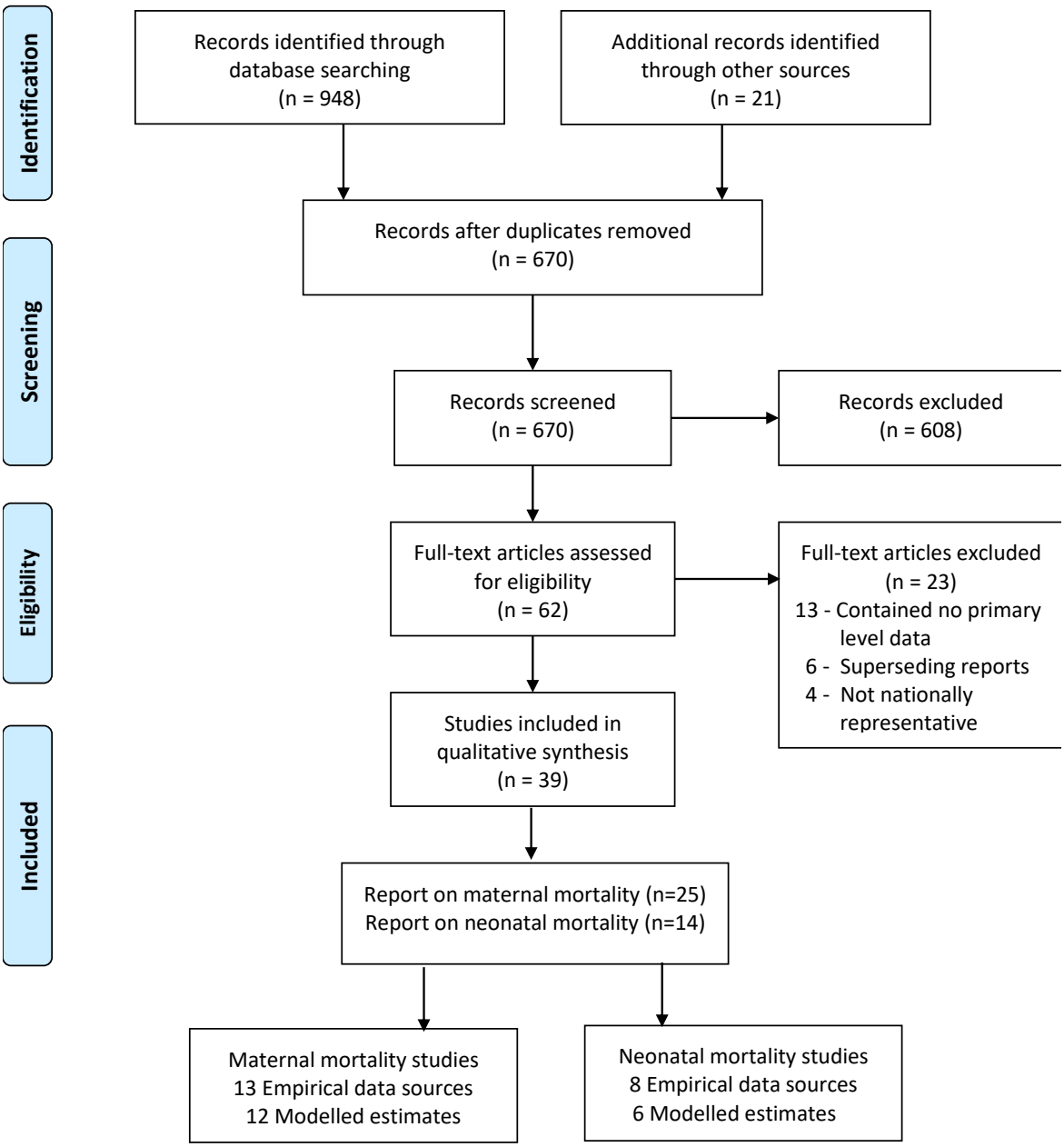


Figure 3: Flow diagram of study selection for inclusion in the qualitative synthesis

Characteristics of the included studies

Maternal mortality

Table 7 depicts the characteristics of studies reporting maternal mortality data. All studies were nationally representative presenting national level data covering a period between 1990 and 2015. Twelve studies estimated MMR though modelling[36,37,66,67,38–45] while 13 studies estimated MMR empirically[28,46,55,56,68,47–54]. Regarding the data sources and methods, 12 studies based on modelling[36,37,67,38–45], 7 active surveillance[28,49,50,52,54,55,68], 3 vital registration[46–48], 2 population-based household[56,66] and a census[53]. The most common definitive data source was the Confidential Enquiry into Maternal Deaths (CEMD) (n=7)[28,49,50,52,54,55,68] followed by the WHO models (n=6)[36,37,40,41,43,45] and vital registration (n=3)[46–48].

Table 7: Study characteristics for maternal mortality data

No.	Author(year)	Duration covered	Approach	Estimation method/Data source	Grouped estimates source*
1.	Dorrington *et al*, 2016[46]	2008-2013	Direct estimation	Vital registration	Vital registration
2.	WHO, 2015[36]	1990-2015	Modelling	Bayesian maternal mortality estimation model	WHO
3.	MDG/Stats SA, 2015[47]	1998-2013	Direct estimation	Vital registration	Vital registration
4.	Dorrington *et al*, 2015[48]	2011-2014	Direct estimation	Vital registration	Vital registration
5.	WHO, 2014[37]	1990-2013	Modelling	Multilevel-regression model	WHO
6.	Kassebaum *et al*, 2014[38]	1990-2013	Modelling	Cause of Death Ensemble model (CODEm)	IHME
7.	Department of Health, 2014[49]	2005-2014	Direct estimation	Active Surveillance: Empirical data	CEMD
8.	NCCEMD, 2014[50]	2011-2013	Direct estimation	Active Surveillance: Empirical data	CEMD

9.	Udjo, 2014[39]	2001-2007	Modelling	Growth Balance method + Relation Gompertz model	Model A
10.	Pattinson et al, 2013[68]	1999-2012	Direct estimation	Active Surveillance: Empirical data	CEMD
11.	WHO, 2012[40]	1990-2010	Modelling	Multilevel-regression model	WHO
12.	NCCEMD, 2012[28]	2008-2010	Direct estimation	Active Surveillance: Empirical data	CEMD
13.	Garenne, 2011[66]	2007	Modelling	Linear logistic model	Model B
14.	WHO, 2011[41]	1990-2009	Modelling	Bayesian maternal mortality estimation model	WHO
15.	Lozano *et al*, 2011[42]	1990-2011	Modelling	Cause of Death Ensemble model (CODEm)	IHME
16.	Stats SA, 2011[51]	2008-2009	Direct estimation	Vital registration/census	Vital registration
17.	WHO, 2010[43]	1990-2008	Modelling	Multilevel-regression model	WHO
18.	Hogan *et al*, 2010[44]	1980-2008	Modelling	Generalised negative binomial regression	IHME
19.	NCCEMD, 2008[52]	2005-2007	Direct estimation	Active Surveillance: Empirical data	CEMD
20.	Garenne *et al*, 2008[53]	2001	Direct estimation	Census	Census
21.	Moodley, 2003[54]	1999-2001	Direct estimation	Active Surveillance: Empirical data	CEMD
22.	AbouZahr *et al*, 2001[45]	2000	Modelling	Robust regression	WHO
23.	Hill *et al*, 2001[67]	1995	Modelling	Robust regression	Model C
24.	Moodley, 2000[55]	1999	Direct estimation	Active Surveillance: Empirical data	CEMD
25.	SADHS, 1998[56]	1992-1998	Direct sisterhood	population-based survey	DHS

***CEMD** - Confidential Enquiry to Maternal Deaths; **IHME**: Institute for Health Metrics and Evaluation; **DHS** – Demographic and Health Survey; and **WHO** - World Health Organization; *Reported in the graphs*

Neonatal mortality

The study characteristics for neonatal mortality data are presented in Table 8 below. Eight studies used empirical data sources[23,46,56–61] whereas 6 studies were based on modelling[14,41,62–65].

Regarding the data sources and methods, six studies were modelling/systematic analysis[14,62,63,65], three vital registration[23,46,58], three population surveys[56,57,61] and two active surveillance[59,60]. The most dominant definitive data sources for neonatal mortality estimates were vital registration[23,46,58] and population surveys[56,57,61] respectively.

Table 8: Study characteristics for neonatal mortality data

No.	Author(year)	Duration covered	Approach	Estimation method/ Data sources	Grouped estimation sources*
1.	SADHS, 2016[57]	2011-2016	Direct estimation	Population-based survey	DHS
2.	Dorrington *et al*, 2016[46]	2012-2015	Direct estimation	Vital registration	Vital registration
3.	UNICEF, 2015[14]	1990-2015	Modelling	Bayesian Hierarchical Splines Regression	UNICEF
4.	UNICEF, 2014[62]	1990-2013	Modelling	Bayesian Hierarchical Splines Regression	UNICEF
5.	Dorrington *et al*, 2014[58]	2009-2013	Direct estimation	Vital registration	Vital registration
6.	Pattinson *et al*, 2014[23]	2012-2013	Direct estimation	Vital registration	Vital registration
7.	NaPeMMCO, 2014[59]	2010-2013	Direct estimation	Active Surveillance	PPIP
8.	WHO, 2011[41]	1990-2009	Modelling	Bayesian B-splines bias-adjusted model	WHO
9.	Oestergaard *et al*, 2011[63]	1990-2009	Modelling	Multilevel-regression model	Model A
10.	NaPeMMCO, 2011[60]	1997-2008	Direct estimation	Active Surveillance	PPIP
11.	Rajaratnam *et al*, 2010[64]	1970–2010	Modelling	Gaussian process regression	Model B
12.	SADHS, 2007[61]	1998-2003	Direct estimation	Population-based survey	DHS
13.	Hyder *et al*, 2003[65]	1995	Modelling	UN projections	Model C

14.	SADHS, 1998[56]	1988-1998	Direct estimation	Population-based survey	DHS

Reported in the graphs; DHS – Demographic and Health Survey; **WHO - World Health Organization; **PPIP** - Perinatal Problem Identification Program*

Risk of bias within studies

Table 9 below presents the assessment of risk of bias of the individual studies. A total of 11 studies reporting maternal mortality data[28,39,68,50,52–56,66,67] and six studies reporting neonatal mortality data[23,56,57,61,63,65] had overall high risk of bias. Among studies reporting maternal mortality data, three studies didn't use the ICD-10 definition of maternal death[53,66,67]; eight studies used data which were not population-representative[28,39,49,50,54–56,68], one study used sisterhood estimation methods[56] and the sampling technique was unclear in one study[54]. Of seven studies reporting neonatal mortality data having an overall high risk of bias, four were not population-representative[23,63,65] and 3 were surveys based on the recall of neonatal deaths more than 6 months previously[56,57,61].

Table 9: Risk of bias assessment in individual studies

No.	Author(year)	Definition	Ascertainment of deaths/livebirths	Sampling technique/design	Data quality	Overall risk of bias
Maternal mortality						
1.	Dorrington *et al*, 2016[46]	Low	Low	Low	Low	Low
2.	WHO, 2015[36]	Low	Low	Low	Low	Low
3.	MDG/Stats SA, 2015[47]	Low	Low	Low	Low	Low
4.	Dorrington *et al*, 2015[48]	Low	Low	Low	Low	Low
5.	WHO, 2014[37]	Low	Low	Low	Low	Low
6.	Kassebaum *et al*, 2014[38]	Low	Low	Low	Low	Low

7.	Department of Health, 2014[49]	Low	High	Low	Low	High
8.	NCCEMD, 2014[50]	Low	High	Low	Low	High
9.	Udjo, 2014[39]	Low	Low	Low	Low	Low
10.	Pattinson et al, 2013[68]	Low	High	Low	Low	High
11.	WHO, 2012[40]	Low	Low	Low	Low	Low
12.	NCCEMD, 2012[28]	Low	High	Low	Low	High
13.	Garenne, 2011[66]	High	Low	Low	Low	High
14.	WHO, 2011[41]	Low	Low	Low	Low	Low
15.	Lozano *et al*, 2011[42]	Low	Low	Low	Low	Low
16.	Stats SA, 2011[51]	Low	Low	Low	Low	Low
17.	WHO, 2010[43]	Low	Low	Low	Low	Low
18.	Hogan *et al*, 2010[44]	Low	Low	Low	Low	Low
19.	NCCEMD, 2008[52]	Low	High	Low	Low	High
20.	Garenne *et al*, 2008[53]	High	Low	Low	Low	High
21.	Moodley, 2003[54]	Low	High	High	Low	High
22.	AbouZahr *et al*, 2001[45]	Low	Low	Low	Low	Low
23.	Hill *et al*, 2001[67]	High	Low	Low	Low	High
24.	Moodley, 2000[55]	Low	High	Low	Low	High
25.	SADHS, 1998[56]	Low	High	Low	Low	High
Neonatal mortality						
1.	SADHS, 2016[57]	Low	High	Low	Low	High
2.	Dorrington *et al*, 2016[46]	Low	Low	Low	Low	Low
3.	UNICEF, 2015[14]	Low	Low	Low	Low	Low
4.	UNICEF, 2014[62]	Low	Low	Low	Low	Low
5.	Dorrington *et al*, 2014[58]	Low	Low	Low	Low	Low
6.	Pattinson *et al*, 2014[23]	Low	High	Low	Low	High
7.	NaPeMMCO, 2014[59]	Low	Low	Low	Low	Low

8.	WHO, 2011[41]	Low	High	Low	Low	High
9.	Oestergaard *et al*, 2011[63]	Low	High	Low	Low	High
10.	NaPeMMCO, 2011[60]	Low	Low	Low	Low	Low
11.	Rajaratnam *et al*, 2010[64]	Low	Low	Low	Low	Low
12.	SADHS, 2007[61]	Low	High	Low	Low	High
13.	Hyder *et al*, 2003[65]	Low	High	Low	Low	High
14.	SADHS, 1998[56]	Low	High	Low	Low	High

Results of individual studies and data synthesis

Maternal mortality

Figure 4 depicts the trend of maternal mortality from 1990 to 2015. Estimates of MMR from most reports indicate an upward trend over time, at least until 2006 or 2009, thereafter a downward trend until 2015. Notably, four studies that ascertained maternal mortality using empirical[53,55] and modelling[39,66] approaches reported extreme estimates of MMR compared to those of other sources. Nevertheless, all recent empirical estimates appeared to converge over time.

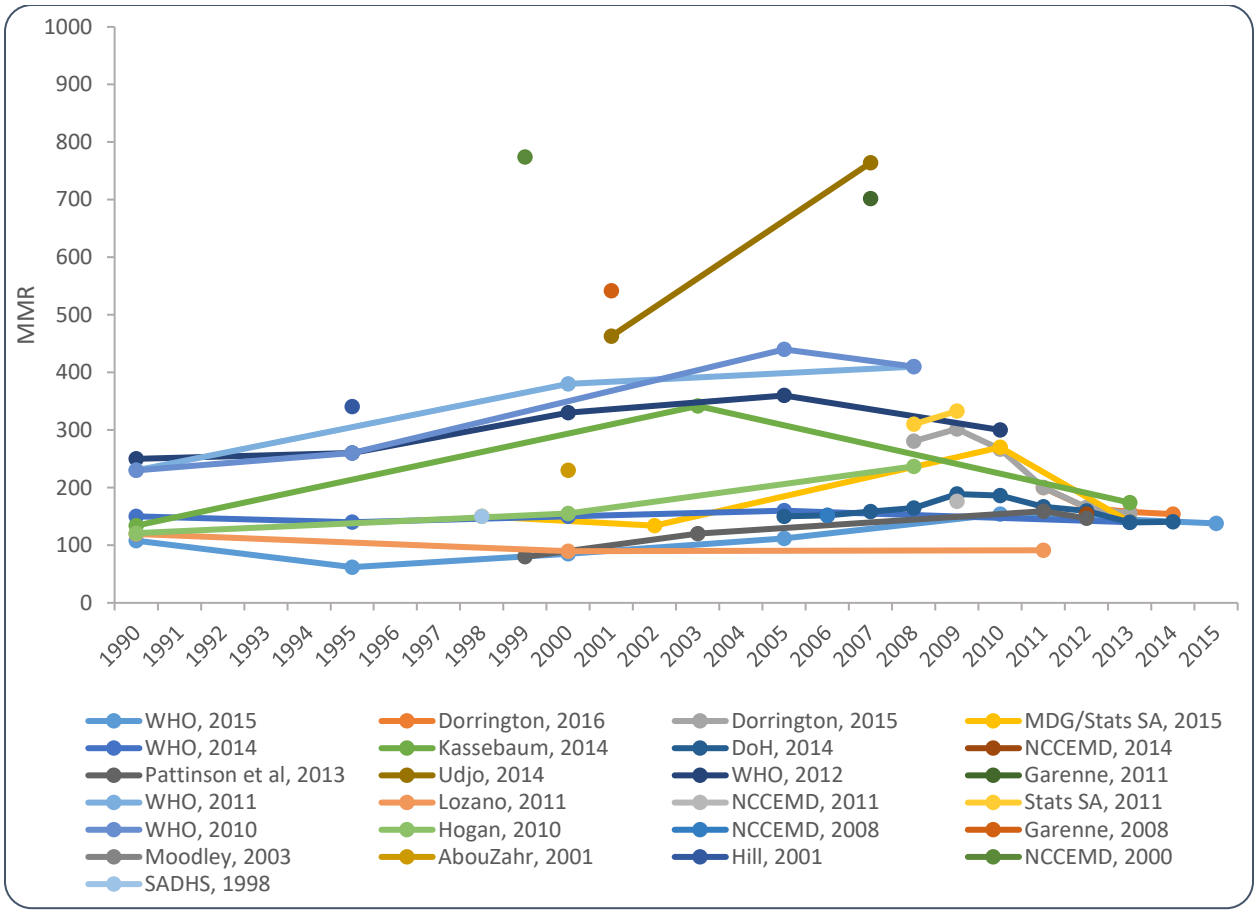
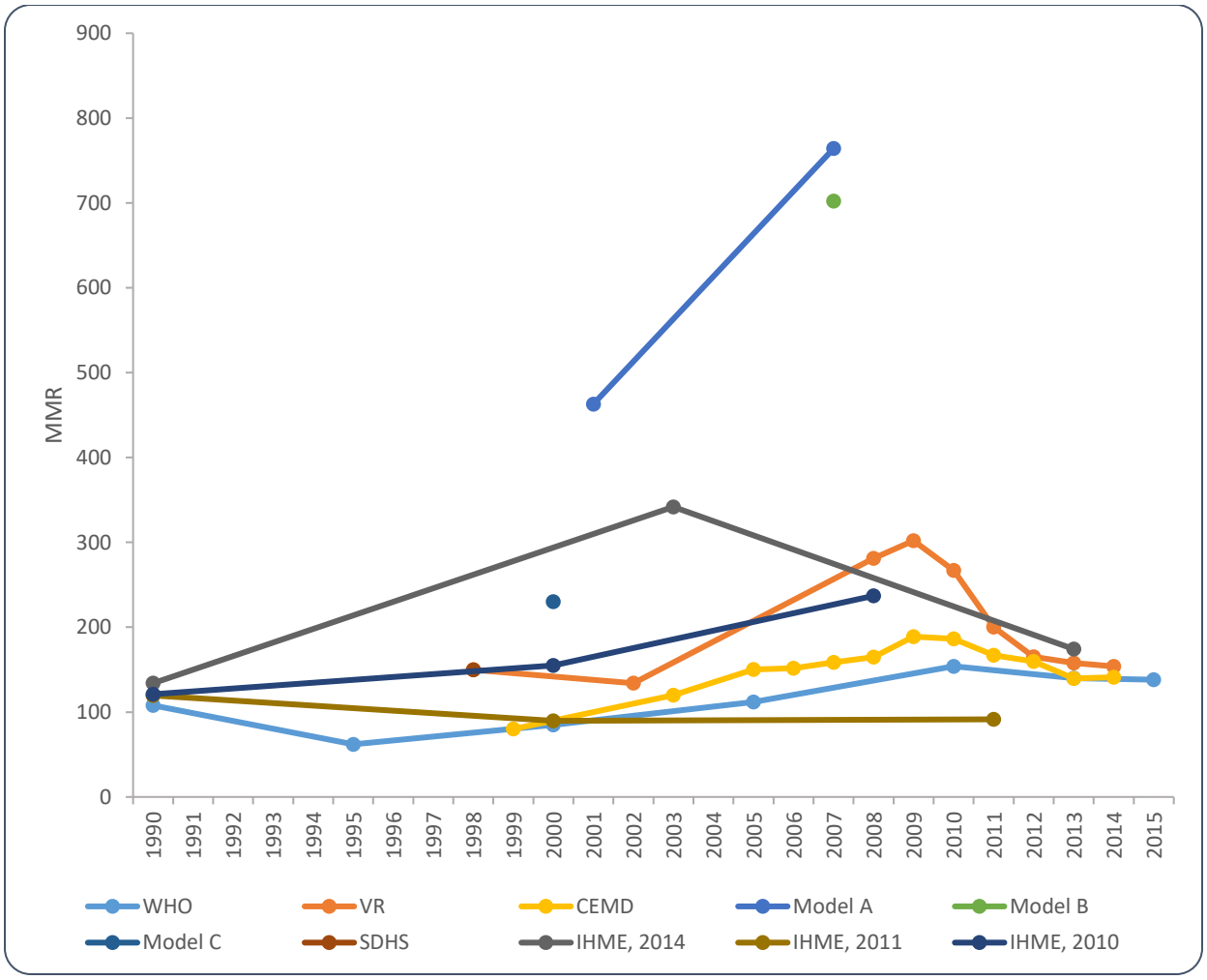


Figure 4: Trends in maternal mortality from 1990-2015 in South Africa

Figure 5 shows the trends in maternal mortality according to data source and estimation method, with the most up-to-date estimates superseding all previously published reports. Most estimates of MMR reported by the global agencies (WHO)[69] were divergent from institutional reports (IHME)[42,70,71], and most modelled estimates (Model A, B and C)[39,66,72] were widely divergent from estimates obtained through empirical methods (VR, CEMD and SDHS) [28,46,48–50,52,53,56,68]. Trends in MMR based on estimates from confidential enquiry (CEMD)[28,49,50,52,54,55,68] and vital registration (VR)[46–48,53] showed an increase until 2009 followed by a drop in 2010. These estimates appeared to converge over time.



***CEMD** - Confidential Enquiry to Maternal Deaths; **IHME**: Institute for Health Metrics and Evaluation; **SDHS** – South Africa Demographic and Health Survey; **VR** - Vital Registration; **WHO** - World Health Organization.*

Figure 5: Trends in maternal mortality according to data source and estimation method

Neonatal mortality

Figure 6 depicts the trends of neonatal mortality from 1990 to 2015. Estimates of NMR from two sources indicated a slightly upward trend over time until 2004, followed by a steady decrease until 2013. Two single-year studies deriving their estimates using empirical[57] and modelling[65] approaches respectively, reported substantially higher neonatal mortality rates than the others.

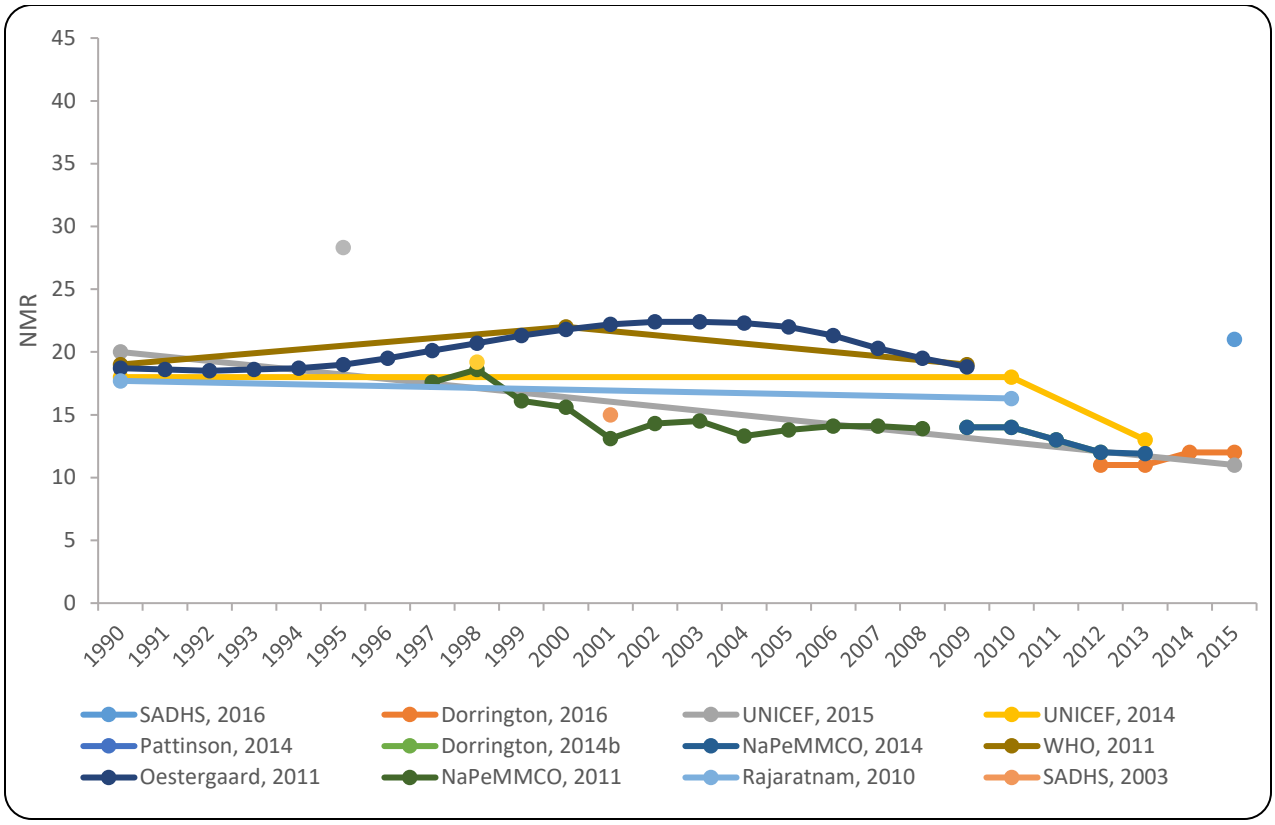


Figure 6: Trends in neonatal mortality from 1990-2015 in South Africa

Furthermore, estimates of NMR from global agencies (WHO and UNICEF)[14,41,62] were widely and periodically divergent from institutional reports (PPIP and VR)[23,46,58–60,73]. Modelled estimates (WHO, UNICEF, Model A and B)[14,41,62,65,74,75] were high and divergent from estimates obtained through empirical methods (VR, PPIP and SDHS)[23,46,56–61,73], with no clear pattern. Trends in NMR based on estimates from the PPIP and vital registration show a slight decline with periodic increase in neonatal mortality from 2000 to 2015[23,46,58–60,73].

Figure 7 shows trends in neonatal mortality according to data source and estimation method.

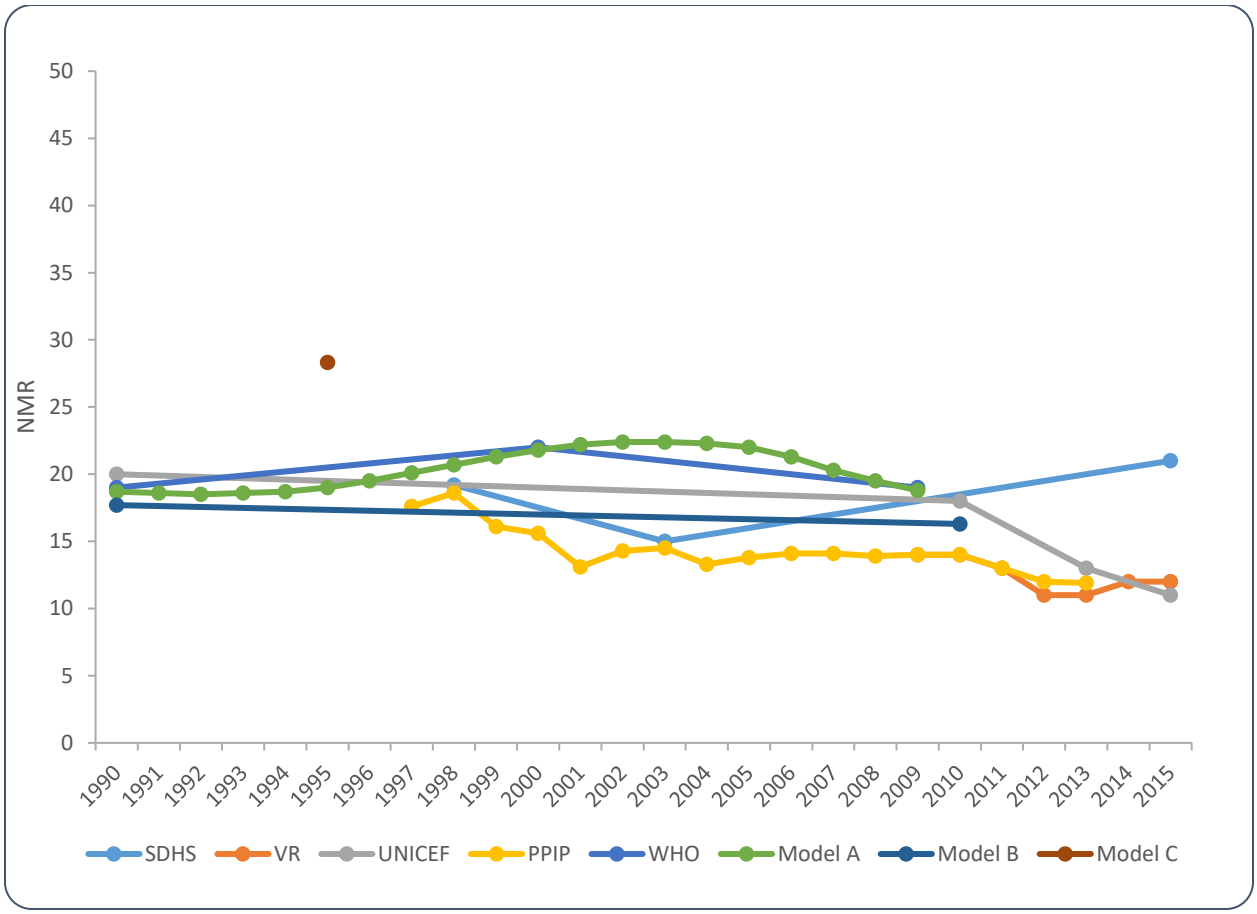


Figure 7: Trends in neonatal mortality according to data source and estimation method

3.5 Discussion

Summary of evidence

This systematic review aimed to provide an overview of maternal and neonatal mortality from 1990 to 2015 for monitoring purposes, tracking progress and to advocate resources and policy attention. In general, the estimates derived from all studies and reports indicated that South Africa did not achieve the MDG 4a and 5a goals of reducing under-five mortality by two-thirds and maternal mortality by three-quarters between 1990 and 2015 respectively. Despite the country struggling to achieve the MDG goals for maternal and neonatal mortality in the last two

decades, recent reports showed significant progress has been made in reducing maternal mortality, with a slight increase in neonatal mortality[46,69,76].

Broad trends

Although maternal and neonatal mortality rates are highly researched by both local and international authors or institutions in South Africa, there is considerable uncertainty around these estimates in the country. The possible reason for this might be, high reliance on only a few data sources and limited empirical work. Despite the uncertainties about the actual levels of MMR and NMR in the country, estimates from both the in-country institutional reports and global agencies indicated an upward trend in MMR until around 2006 and 2009. However, the increase in MMRs between 2001 and 2006 might specifically be explained by a consistent increase in HIV prevalence among pregnant women in the same period[77].

In addition, the downward trends in MMRs and NMRs from 2009 can be linked with the massive uptake of HIV treatment and an increased coverage of essential interventions, in particular the Prevention of Mother-to-Child Transmissions (PMTCT) of HIV which currently stands at over 90%[29,30]. Nonetheless, all recent estimates are much more closely grouped indicating a convergence over time. This might be contributed to by the massive increase in birth registration, availability of multiple data sources in the recent years (used as inputs for modelled estimates) as well as the significant investments and improvements in maternal and neonatal vital registration, and the relative importance of mortality statistics in clinical management, healthcare administrations and epidemiologic analyses[78].

Challenges in measuring maternal and neonatal mortality

The study suggests empirical methods i.e., vital registration, household surveys and censuses are subject to misclassification and under-reporting of maternal deaths, thus leading to divergent estimates of these outcomes. Furthermore, estimating neonatal mortality from census and household surveys in high HIV prevalence settings is known to provide biased estimates of child mortality (overestimating the actual level) due to the correlation between HIV deaths in mothers and their children[72].

Highly variable estimates

Large margins of uncertainty associated with the estimated MMR and NMR highlight the need for interpreting these estimates with caution as well as not using them for monitoring trends over a short duration. The reasons for variations in the estimates of maternal and neonatal mortality remain poorly researched over the past two decades. In this review, we have observed a substantial discrepancy in the consistency of definitions used in estimation of these outcomes, such as differentiating maternal deaths from pregnancy related deaths[53,66,67]. Thus, uncertainties in estimates of MMR might be partly explained by differences in definitions used. Different estimation techniques used to obtain MMR and NMR have necessitated the use of different data inputs i.e., empirical data versus modelled estimates, which likely contributed to the divergent estimates of these outcomes.

For these reasons, cross-country comparisons; comparisons based on data from different sources; and assessments of the overall burden become difficult. These comparisons should in many cases be interpreted with considerable caution due to different strategies being employed to derive such estimates. Evidence from recent studies focusing on estimating child mortality have revealed that methodological differences bias and compromise international comparisons of perinatal mortality[79–81]. Moreover, divergent estimates of MMR and NMR by different sources compromise the interpretation of trends over time.

Improving estimates

Over the past three decades, efforts have been made to improve the quality of maternal and neonatal mortality data due to incompleteness of vital registration systems as well as the lack of reliable population surveys collecting detailed information on birth histories in the country. This includes the introduction of modules about sibling history in national household surveys (e.g. Demographic and Health Survey (DHS)), including questions in censuses about whether a woman's deaths was related to pregnancy, and the use of mixed methods cross-referencing facility records to determine the extent of under-registration of maternal deaths in vital registration system[4,7,82,83].

However, these improvements could also have contributed to the increasing mortality estimates over time. Despite improvements in the completeness of death registrations in the last decade, the completeness of death registration has been reported to be lower in children, when compared to adults, and in rural areas than in urban areas[83]. This might potentially explain some of the variability in the estimates of both maternal and neonatal mortality in the country.

Generally, this review has revealed divergent estimates of MMR and NMR obtained from vital registration, household surveys, censuses and modelling over time. To obtain more accurate estimates, there is a need for applying additional estimation techniques which utilise multiple data sources when available, to correct for underreporting of these outcomes. Capture-recapture methods are one possibility, for resolving uncertainties in estimating conditions that have diverse estimates, by operationalising statistically overlapping information from multiple data sources[84–88]

3.6 Conclusion

Estimates from the global agencies and in-country institutional reporting, although widely divergent, indicate South Africa has not achieved the MDG targets for maternal and neonatal mortality but it has made significant progress in reducing maternal mortality in the later years under review. Discrepancies in data sources and quality from which these estimates were obtained, and highly variable estimates, highlight the uncertainties in the true estimates of maternal and child mortality in South Africa.

In order to track progress and monitor the Sustainable Development Goals (SDGs) and the goal for Health care for all by 2030; the country needs accurate, reliable, continuous and timely mortality statistics from the vital registration system, a clear understanding of any under-ascertainment of maternal or neonatal mortality, and consistent approaches to accounting for these. It would be ideal if global agencies worked closely with local researchers to agree on optimal calibration of South African estimates in multi-country models.

3.7 References

1. United Nations. Handbook on Training in Civil Registration and Vital Statistics Systems. *United Nations* 281 (2002).

2. Atrash, H. K., Alexander, S. & Berg, C. J. Maternal mortality in developed countries: not just a concern of the past. *Obstet. Gynecol.* **86**, 700–5 (1995).

3. Hill, K. *et al.* How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches. *Bull. World Health Organ.* **84**, 173–80 (2006).

4. Stanton, C. *et al.* Every death counts: measurement of maternal mortality via a census. *Bull. World Health Organ.* **79**, 657–64 (2001).

5. WHO. *WHO guidance for measuring maternal mortality from a census*. *World Health Organisation* (2008).

6. Leitao, J. *et al.* Revising the WHO verbal autopsy instrument to facilitate routine cause-of-death monitoring. *Glob. Health Action* **6**, 21518 (2013).

7. Graham, W. The sisterhood method for estimating maternal mortality. *Mothers Child.* **8**, 1–2 (1989).

8. Merdad, L., Hill, K. & Graham, W. Improving the Measurement of Maternal Mortality: The Sisterhood Method Revisited. *PLoS One* **8**, (2013).

9. WHO, UNICEF, UNFPA, World Bank & United Nations Population Division. Trends in maternal mortality 1990 to 2015. (2015) doi:10.

10. Graham, W. J. Now or never: The case for measuring maternal mortality. *Lancet* **359**, 701–704 (2002).

11. Berg, C. J., Danel, I. & Mora, G. Guidelines for Maternal Mortality Epidemiological Surveillance. *Pan Am. Heal. Organ.* 50 (1996).

12. Walker, N., Hill, K. & Zhao, F. Child Mortality Estimation: Methods Used to Adjust for Bias due to AIDS in Estimating Trends in Under-Five Mortality. *PLoS Medicine* vol. 9 (2012).

13. Silva, R. Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories. *PLoS Med.* **9**, e1001296 (2012).

14. UNICEF, WHO, World Bank group & United Nations. *Levels & Trends in Child Mortality - Report 2015: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. *UNICEF* http://www.childmortality.org/files_v20/download/igme report 2015 child mortality final.pdf (2015) doi:10.1371/journal.pone.0144443.

15. Moodley, J. *Saving Mothers 2011-2013: Sixth report on confidential enquiries into maternal deaths in South Africa, Short report*. *Pretoria: Department of Health* http://www.kznhealth.gov.za/mcwh/Maternal/Saving-Mothers-2011-2013-short-report.pdf (2014).

16. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* vol. 384 980–1004 (2014).

17. Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 957–979 (2014).

18. Saleem, S. *et al.* A prospective study of maternal, fetal and neonatal deaths in low- and middle-income countries. TT -. *Bull. World Health Organ.* **92**, 605–612 (2014).

19. Bradshaw, D. *et al.* Every death counts: use of mortality audit data for decision making to save the lives of mothers, babies, and children in South Africa. *Lancet* **371**, 1294–1304 (2008).

20. Moodley, J. *et al.* The confidential enquiry into maternal deaths in South Africa: a case study. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 53–60 (2014).

21. Chopra, M., Daviaud, E., Pattinson, R., Fonn, S. & Lawn, J. E. Saving the lives of South Africa's mothers, babies, and children: can the health system deliver? *Lancet* **374**, 835–846 (2009).

22. Souza, J. P. *et al.* Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *The Lancet* vol. 381 1747–1755 (2013).

23. Pattinson, R. & Rhoda, N. *Saving Babies 2012-2013: Ninth report on perinatal care in South Africa*. *RC Pattinson* http://www.ppip.co.za/wp-content/uploads/Saving-Babies-2012-2013.pdf (2014).

24. Domingo-Salvany, a *et al.* Analytical considerations in the use of capture-recapture to estimate prevalence: case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *Am. J. Epidemiol.* **148**, 732–740 (1998).

25. AbouZahr, C. New estimates of maternal mortality and how to interpret them: choice or confusion? *Reprod. Health Matters* **19**, 117–128 (2011).

26. Brazil, N. & Wilmoth, J. *A Comparison of UN and IHME Maternal Mortality Estimates*. http://paa2013.princeton.edu/papers/132416 (2013).

27. Republic of South Africa. *Termination of Pregnancy Act*. http://www.saflii.org/za/legis/num_act/cotopa1996325/ (1996).

28. National Committee for Confidential Enquiry into Maternal Deaths. *Saving Mothers 2008-2010: Fifth Comprehensive Report on Confidential Enquiries into Maternal Deaths in South Africa*. *Department of Health* vol. 5 (2012).

29. UNAIDS. *On the Fast-Track to an AIDS-free generation*. http://www.unaids.org/sites/default/files/media_asset/GlobalPlan2016_en.pdf (2016).

30. UNAIDS. *Global AIDS update*. http://www.who.int/hiv/pub/arv/global-AIDS-update-2016_en.pdf?ua=1 (2016).

31. Damian, D. J., Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review protocol. *Syst. Rev.* **6**, 165 (2017).

32. Shamseer, L. *et al.* Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *Br. Med. J.* **349**, (2015).

33. Liberati, A. *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. in *Journal of clinical epidemiology* vol. 62 e1-34 (2009).

34. WHO. *International Classification of Diseases (ICD): Eleventh revision instruction manual*. http://www.who.int/classifications/icd/en/ (2017).

35. WHO. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. http://apps.who.int/classifications/icd10/browse/2016/en (2016).

36. WHO. *Trends in Maternal Mortality : 1990 to 2015 Estimates by WHO,UNICEF,UNFPA,The World Bank and the United Nations Population Division*. *Organization* http://whqlibdoc.who.int/publications/2010/9789241500265_eng.pdf%5Cnhttp://apps.who.int/iris/bitstream/10665/112697/1/WHO_RHR_14.13_eng.pdf?ua=1 (2015) doi:ISBN 978 92 4 156514 1.

37. World Health Organization. Trends in Maternal Mortality : 1990 to 2013 Executive Summary. (2014).

38. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, (2014).

39. Udjo, E. O. & Lalthapersad-Pillay, P. Estimating maternal mortality and causes in South Africa: National and provincial levels. *Midwifery* **30**, 512–518 (2014).

40. WHO, UNICEF, UNFPA & The World Bank Estimates. *Trends in maternal mortality: 1990 to 2010. WHO, UNICEF, UNFPA and The World Bank estimates*. http://whqlibdoc.who.int/publications/2010/9789241500265_eng.pdf (2012) doi:10.1016/j.geoforum.2013.12.013.

41. World Health Organization. *World health statisitics 2011*. *World health statistics* (2011) doi:10.1002/(SICI)1096-987X(199802)19:3<259::AID-JCC1>3.0.CO;2-S.

42. Lozano, R. *et al.* Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: An updated systematic analysis. *Lancet* **378**, 1139–1165 (2011).

43. WHO, UNICEF, UNFPA & Bank, T. W. *Trends in Maternal Mortality : 1990 to 2008. Estimates developed by WHO, UNICEF, UNFPA and The World Bank*. *World Health Organization* vol. 32 http://whqlibdoc.who.int/publications/2010/9789241500265_eng.pdf (2010).

44. Hogan, M. C. *et al.* Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. *Lancet* **375**, 1609–1623 (2010).

45. Abouzahr, C. & Wardlaw, T. Maternal mortality at the end of a decade : signs of progress ? **79**, 561–573 (2001).

46. Dorrington, R., Bradshaw, D., Laubscher, R. & Nannan, N. *Rapid Mortality Surveillance Report 2015*. www.mrc.ac.za/bod/bod.htm (2016) doi:ISBN: 978-1-920618-00-1.

47. Statistics South Africa. *Millennium Development Goals 5: Improve maternal health 2015 / Statistics South Africa*. (2015).

48. Dorrington, R., Bradshaw, D., Laubscher, R. & Nannan, N. *Rapid Mortality Surveillance Report 2014*. http://www.mrc.ac.za/bod/RapidMortalitySurveillanceReport2014.pdf (2015).

49. Department of Health. *2014 Saving Mothers: Annual report and detailed analysis of maternal deaths due to non-pregnancy related infections*. (2015).

50. National Committee for Confidential Enquiries into Maternal Deaths. *Saving Mothers 2011-2013: Sixth report on confidential enquiries into maternal deaths in South Africa Short report*. http://www.doh.gov.za/docs/reports/2007/savingmothers.pdf (2014).

51. Statistics South Africa. *Census 2011: Estimation of Mortality in South Africa*. (2011).

52. NCCEMD. *Saving Mothers 2005-2007 : Fourth Report on Confidential Enquiries into Maternal Deaths in South Africa*. *Health (San Francisco)* http://www.doh.gov.za/docs/reports/2007/savingmothers.pdf (2008).

53. Garenne, M., McCaa, R. & Nacro, K. Maternal mortality in South Africa in 2001: From demographic census to epidemiological investigation. *Popul. Health Metr.* **6**, 4 (2008).

54. Moodley, J. Saving mothers: 1999 - 2001. *SAMJ South African Med. J.* **93**, (2003).

55. Moodley, J. Saving mothers in South Africa. *South African Journal of Obstetrics and*

Gynaecology vol. 6 13–15 (2000).

56. Department of Health, Medical Research Council & Measure DHS. *South Africa Demographic and Health Survey: Final Report*. http://dhsprogram.com/pubs/pdf/FR131/FR131.pdf (1998).

57. National Department of Health (NDoH), Statistics South Africa Council (Stats SA), South African Medical Research (SAMRC) & ICF. *South Africa Demographic and Health Survey 2016: Key Indicators.* (2017).

58. Dorrington, R., Bradshaw, D. & Laubscher, R. *Rapid Mortality Surveillance Report 2012*. www.mrc.ac.za/bod/bod.htm (2014) doi:ISBN: 978-1-920618-00-1.

59. National perinatal morbidity and mortality committee. *National Perinatal Morbidity and Mortality Committee Report 2014*. http://www.ppip.co.za/saving-babies/ (2014).

60. NAPEMMCO. *National Perinatal Mortality and Morbidity Committee (NaPeMMCo) Triennial Report (2008-2010)*. *National Perinatal Mortality and Morbidity Committee (NaPeMMCo) Triennial Report (2008-2010)* http://www.doh.gov.za/docs/reports/2011/perireport.pdf (2011).

61. Department of Health, Medical Research Council & OrcMacro. *South African Demographic and Health Survey - 2003: Full Report*. https://dhsprogram.com/pubs/pdf/FR206/FR206.pdf (2007) doi:10.1017/CBO9781107415324.004.

62. UNICEF, WHO, World Bank, U. *Levels & Trends in Child Mortality in 2014- Report: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. http://www.unicef.org/videoaudio/PDFs/UNICEF_2012_child_mortality_for_web_0904.pdf (2014).

63. Oestergaard, M. Z. *et al.* Neonatal mortality levels for 193 countries in 2009 with trends since 1990: A systematic analysis of progress, projections, and priorities. *PLoS Medicine* vol. 8 e1001080 (2011).

64. Rajaratnam, J. K. *et al.* Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. *Lancet (London, England)* **375**, 1988–2008 (2010).

65. Hyder, A. A., Wali, S. A. & McGuckin, J. The burden of disease from neonatal mortality: A review of South Asia and Sub-Saharan Africa. *BJOG An Int. J. Obstet. Gynaecol.* **110**, 894–901 (2003).

66. Garenne, M., McCaa, R. & Nacro, K. Maternal mortality in South Africa: An update from the 2007 Community Survey. *J. Popul. Res.* **28**, 89–101 (2011).

67. Hill, K., AbouZahr, C. & Wardlaw, T. Estimates of maternal mortality for 1995. *Bull. World*

Health Organ. **79**, 182–193 (2001).

68. Pattinson, R. C., Fawcus, S. & Moodley, J. *Tenth interim report on Confidential Enquiries into Maternal Deaths in South Africa*. (2013).

69. WHO, UNICEF, UNFPA, World Bank Group & The United Nations Population Division. *Trends in Maternal Mortality: 1990 to 2015. Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*. (2015) doi:10.

70. Kassebaum, N. J. *et al.* Global, regional, and national levels of maternal mortality, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, (2016).

71. Hogan, M. C. *et al.* Maternal mortality for 181 countries, 1980-2008: a systematic analysis of progress towards Millennium Development Goal 5. *Lancet* **375**, 1609–1623 (2010).

72. AbouZahra, C. & Wardlawb, T. *MATERNAL MORTALITY IN 2000: Estimates Developed by WHO, UNICEF and UNFPA*. http://apps.who.int/iris/bitstream/10665/68382/1/a81531.pdf (2002) doi:10.1016/S0968-8080(07)30326-1.

73. Nannan, N. *et al. Under-5 mortality statistics in South Africa: Shedding some light on the trends and causes 1997-2007.* http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:UNDER-5+MORTALITY+STATISTICS+IN+SOUTH+AFRICA+:+Shedding+some+light+on+the+trend+and+causes+1997-2007#0 (2012).

74. Oestergaard, M. Z. *et al.* Neonatal mortality levels for 193 countries in 2009 with trends since 1990: A systematic analysis of progress, projections, and priorities. *PLoS Medicine* vol. 8 (2011).

75. Rajaratnam, J. K. *et al.* Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards Millennium Development Goal 4. *Lancet* **375**, 1988–2008 (2010).

76. UNICEF, WHO, World Bank, U.-D. P. D. *Levels and trends in child mortality 2015*. *WHO* http://www.who.int/maternal_child_adolescent/documents/levels_trends_child_mortality_2015/en/ (2017).

77. Statistics South Africa. *Millennium Development Goals 6 : Combat HIV/AIDS, malaria and other diseases*. http://www.statssa.gov.za/MDG/MDG_Goal6_report_2015_.pdf (2015).

78. Mathers, C. D., Fat, D. M., Inoue, M., Rao, C. & Lopez, A. D. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull. World Health Organ.* **83**, 171–7 (2005).

79. Su, B.-H. *et al.* Neonatal Outcomes of Extremely Preterm Infants from Taiwan: Comparison with Canada, Japan, and the USA. *Pediatr. Neonatol.* **56**, 46–52 (2015).

80. Hossain, S. *et al.* Outborns or Inborns: Where Are the Differences? A Comparison Study of Very Preterm Neonatal Intensive Care Unit Infants Cared for in Australia and New Zealand and in Canada. *Neonatology* **109**, 76–84 (2016).

81. Joseph, K. S., Razaz, N., Muraca, G. M. & Lisonkova, S. Methodological Challenges in International Comparisons of Perinatal Mortality. *Curr. Epidemiol. Reports* **4**, 73–82 (2017).

82. Stanton, C., Abderrahim, N. & Hill, K. An assessment of DHS maternal mortality indicators. *Stud. Fam. Plann.* **31**, (2000).

83. Nannan, N. *et al.* UNDER-5 MORTALITY STATISTICS IN SOUTH AFRICA: Shedding some light on the trend and causes. (1997).

84. Forecasting, I. W. G. for D. M. and. Capture-recapture and multiple-record systems estimation I: History and theoretical development. International Working Group for Disease Monitoring and Forecasting. *Am. J. Epidemiol.* **142**, 1047–58 (1995).

85. Sekar, C. C. & Deming, W. E. On a Method of Estimating Birth and Death Rates and the Extent of Registration (Excerpt). *Am. Stat.* **58**, 13–15 (2004).

86. Wittes, J. & Sidel, V. W. A generalization of the simple capture-recapture model with applications to epidemiological research. *J. Chronic Dis.* **21**, 287–301 (1968).

87. Fienberg, S. E. The Multiple Recapture Census for Closed Populations and Incomplete 2 k Contingency Tables. *Biometrika* **59**, 591 (1972).

88. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20**, 3123–3157 (2001).

CHAPTER 4: Validating the Provincial Deterministic Data Linkage approach through a fully Probabilistic Record Linkage method applied to Mortality and Pregnancy data, Western Cape, South Africa.

4.1 Abstract

Background: The WC province has linked individuals across datasets in the Provincial Health Data Centre (PHDC) using a decision-rule based approach, which has not been formally evaluated. Probabilistic Record Linkage (PRL) is known to perform well in health datasets but had not yet been tested in the PHDC. The chapter sought to test the implementation of PRL on a relevant use-case, and if feasible, use this to cross-validate the pre-existing linkage approach.

Methods: Mortality and pregnancy datasets from the WC provincial health data centre from 2010 to 2013 were linked for potential matches using the PRL method. Linkage fields included first name, surname, date of birth, South African ID, folder number and postal address. Probability weights and defined thresholds (0.9) were used to classify potential matched pairs. Standard diagnostic agreement measures were used to compare the performance of PRL and fuzzy matching. Sensitivity analyses were conducted to examine the effect of varying linkage parameters.

Results: In total, 301,918 and 10,437 records from the pregnancy and mortality datasets respectively, were linked for potential matches. A total of 370 matched pairs were identified by the exact matching, 571 by the decision-rule based linkage approach used by the province (PHDC linkage) and 569 by the PRL after manual review of the possible links. Compared with PRL, the provincial (PHDC) linkage approach had 99.94% accuracy, with sensitivity 99.65% (95%

Confidence Interval [CI]: 98.74, 99.96); specificity 99.96% (95% CI: 99.90, 99.99); positive predictive value 99.30% (95% CI: 98.22, 99.81); negative predictive value 99.98% (95% CI: 99.93, 100.0); and 0.99 F-score. PRL performed better when including names, date of birth, South African ID, folder number and postal code information, as well as when using a threshold composite weight score of 0.9 for identifying matched pairs.

Conclusion: The PRL and the decision-rule based approach (deterministic) used by the province provided similar performance. The provided datasets reliant on the pre-existing matching approach ought to be of sufficient fidelity to enable the remaining planned thesis analyses.

Keywords: Data linkage, probabilistic record linkage, fuzzy matching, decision-rule based linkage, provincial linkage, maternal mortality, neonatal mortality, Western Cape, South Africa

4.2 Introduction

Globally, there is a dramatic increase in the availability of digitised routinely collected data which provides new opportunities for research studies[1,2]. These include population registries, electronic medical records, disease monitoring systems, birth and death records, administrative claims data, and other databases which have primarily been developed for clinical and administrative purposes with no specific research questions in mind. The increasing use of routinely-collected data for research purposes presents opportunities for linking records from multiple sources[3,4]. This in-turn increases demand for robust data linkage techniques to improve ascertainment of key research outcomes.

Data linkage, a complex and extensive process which involves merging records from disparate sources belonging to the same entity, can be viewed as both an art and a science[5,6]. Generally, there are numerous data (record) linkage methods although the two most commonly applied methods are deterministic matching and probabilistic record linkage[3,4,7,8]. Deterministic record linkage refers to the use of a unique identifier or exact matching of defined identifiers (such as names, dates of birth, addresses etc.) to link records from different datasets. Using the deterministic linkage approach, records are matched, provided the matching fields agree; otherwise they are unmatched[9]. However, the unique identifiers in most routinely-collected data are prone to registration errors and sometimes they are not reliably collected[8,10,11]. Lack of univocal identifiers in datasets complicates the deterministic linkage procedure, requiring the application of more complex and sophisticated methods which link records using combinations

of various identification variables, or transformation of variables to increase comparability (e.g., the Soundex algorithm)[12–14].

Probabilistic Record Linkage (PRL) is defined as the process of linking records from different data sources that have a high probability of belonging to the same individual when a unique identifier or identifier combinations are unavailable, unreliable or of insufficient quality[3,8,13,15]. In most cases, the probabilistic record linkage is preferred rather than the deterministic linkage due to its superior accuracy and high precision in matching health records[16–22]. However, the accuracy of probabilistic record linkage can be compromised by factors such as variations in data quality and discriminating power of variables[16,23,24].

Monitoring the Sustainable Development Goals 3.1 and 3.2 requires valid, reliable and internationally comparable estimates of maternal and child mortality countrywide. Comparing to other African countries; South Africa has widely divergent estimates of maternal and neonatal mortality from institutional reporting and global metrics, and among global metrics[25–30]. The country and WC province in particular provide unique opportunities to link records by having multiple electronic data sources developed for healthcare administration, clinical management and administrative purposes[31–37]. Due to varying estimates of maternal and neonatal mortality reported by different sources in the country and in the Western Cape (WC) province, there is value in utilising the available multiple data sources to correct for underreporting of these outcomes.

In South Africa, the Department of Health (DoH) is primarily responsible for managing and storing most health data. Comprehensive details of data sources were discussed in chapter 2 (Data sources and Common methods section). Briefly, estimation of child mortality in the country is often based on vital registration although sometimes it is known to under-record neonatal mortality since the neonates are not always on the population register[38]. In addition, the Child Healthcare Problem Identification Program (CHIP) and Perinatal Problem Identification Program (PPIP) are used for clinical audit purposes, which record under-five deaths and, stillbirths and neonatal deaths respectively[32,39]. Similarly, the estimation of maternal mortality is necessarily based on the National Confidential Enquiry into Maternal Deaths (NCEMD) which records maternal deaths[32,39–41]. However, there are additional data sources in the country and particularly in the WC province, which capture mortality information i.e., patient-level mortality data from the clinical system and death certificates with reported cause-of-death from the Department of Home Affairs (DHA). Although these systems can be linked, they are not automatically interconnected.

The amendments to the South African Births and Deaths Registration Act in 2014, prevented access to identifiable cause-of-death data which are now exclusively managed by the national statistics office, Statistics South Africa[42]. This compromised cause-of-death data linkage with other systems for evaluation of health programmes and epidemiological research activities[43]. From 2010-2013 in the Western Cape, the unique patient identifier (also known as folder number or “Clinicom number” based on the name of the hospital information system which issues the number) and the South African identification number (SA ID) have been the common unique

identifiers across sources that can be used to link records in these databases. However, these identifiers are often prone to registration errors and sometimes they are not reliably collected[8,44,45]. For these reasons, the Western Cape Department of Health (WCDoH) data centre has to date merged these databases using a decision rule-based deterministic linkage approach referred to here as the PHDC/provincial linkage.

Deterministic/rule-based linkage methods are known to have limitations; in particular, rules are empirically created in most cases, have no statistical justifications, are difficult to replicate and require maintenance[46,47]. Moreover, they are known to have less discriminating power than probabilistic linkage methods although they are easy to implement[9,17,47]. In our setting, the performance of the decision-rule based approach used by the province which underpins subsequent analyses in this thesis, has never been tested. The quality of the linkage also affects planning, performance and efficiency monitoring (such as the tracking of maternal and neonatal mortality), business intelligence and programmes evaluations, creating thereby a further rationale to formally evaluate the performance of the current linkage approaches.

The primary aim of this chapter was to validate the performance of the provincial record-linkage approach which uses decision-rule based fuzzy linkage by comparison to an independent probabilistic record-linkage exercise. A secondary aim was to determine the feasibility of a fully probabilistic approach to record linkage on the available data sets.

4.3 Methods

Design and Setting

This was a validation study, comparing the decision-rule based approach used by the province to a fully probabilistic method, using observational routinely collected data for clinical management and administrative purposes in the Western Cape Province, South Africa. According to the recent statistics, the province has an estimated population of 6.5 (≈11%) million people[48]. Of them, about three-quarters are estimated to be using public-sectors services[49].

Data sources and types

This chapter used pregnancy and mortality data from the PHDC database. The comprehensive details of the PHDC database are discussed in Chapter 2. The pregnancy data were extracted from the episodes table and the mortality data from the encounters table in the Clinical database. The pregnancy data from 2010 to 2013 and mortality data in 2013 formed the two datasets for this analysis**.** A single-year of mortality data was chosen for the validation exercise to limit the size of the database to enable completion using desktop computing resources. The fully probabilistic linked dataset was compared with the provincial matched data of the same period. The analyses involved implementing the PRL approach and then identifying the overlaps in linkage and non-linkage. The provincial linkage algorithm comprised a series of match criteria that were executed in a particular order so as to link records. The methodology of linking PHDC data used in this comparison is provided in Appendix 2. For illustrative purposes, a linkage approach relying only on exact matching was also evaluated.

Data processing and analysis

Data processing, linkage and analysis were conducted using Stata software version 14.2 (Stata Corporation, College Station, Texas, USA). Probabilistic linkage procedures involved four steps namely data cleaning and standardisation, blocking, records linkage and clerical review of the matched pairs. Selected linkage variables included first name, surname, date of birth, South African ID, folder number and postal address. Sensitivity analyses were conducted to determine the robustness of the linkage criteria. The approaches for implementing the Probabilistic Record Linkage and for comparing outputs of both methods involved: -

1. Implementing the Probabilistic Record Linkage

Data cleaning and standardisation

Preparing datasets before they can be linked requires significant effort and time in order to ensure quality of the linked dataset[50]. Due to differences in coding between the two datasets, the linkage variables were pre-processed across datasets. This involved examining the datasets and carrying out data cleaning and standardisations in order to identify, correct and remove any errors or inconsistencies in the data so that the data items are comparable in each data file[52].

The process involved deleting and trimming the blank spaces, removing special characters, unifying date formats, converting to lower cases all string variables, modulus 10 validation of the folder numbers, checking the validation of the SA ID number and de-duplication.

Blocking

When linking pregnancy and mortality datasets, the number of possible comparisons to be made are equal to the product of the number of records in the two data sets i.e., 301,918 x

10,437=3,151,118,166 respectively. To avoid computational problems when applying probabilistic linkage approaches on large datasets, it is advisable to divide/partition datasets into groups (blocking) outside of which valid matches would not be expected, so as to reduce the number of comparisons needed to be made to find a pair of records to be linked[8,52,53]. To improve computational efficiency, we tried to keep the blocks as small as possible by reducing the comparison space. Nevertheless, using small blocks can exclude valid matched pairs because of inconsistency in data quality of the blocking variable. To overcome this, we decided to include multiple blocking strategies (block passes) to allow records to find the true match pairs using consecutive blocking variables. In our analysis, the blocking criteria included the output of the Soundex algorithm applied to the surname, and date of birth (in combination and individually: day, month, and year).

Linkage

The probabilistic record linkage presented here was carried out using the Stata™ user-written programme "reclink2", a more generalised version of "reclink" by Blasnik (2010, Statistical Software Components S456876, Department of Economics, Boston College). In addition, the record linkage process was also coded manually to ensure a full understanding of the automated procedure. The process of matching records using the probabilistic data linkage involved comparing record pairs on each field/linkage variable (i.e., first name, surname, date of birth, SA ID, folder number and postal address) and an associated match criterion (i.e., exact match on SA ID and/or folder number). This was followed by calculating field probabilities (probability of being linked - m_i or not linked - u_i if the match criteria are met), field weights (the contribution of this

probability to the overall record linkage probability) and determining links based on the threshold/cut-off points for each record pair (the summative probability threshold above which a link will be accepted). The input values (m_i and u_i) were estimated based on the sample frequencies with known links and non-links using mortality data for the year 2012. The optimal parameters were therefore obtained using expectation maximization (EM) approach and string comparators. To determine the cut-off in classifying record pairs as links (matches), the linkage algorithm is assumed to have a single threshold parameter (T_{max})[8]. Setting these threshold values (cut-off) is known to be a difficult and a critical step in PRL which affects the classifications of links and non-links[8]. The threshold cut-off points were defined in advance such that: upper threshold (T_{max}) $= 0.8$ and Lower threshold (T_{min}) $= 0.6$ and then optimised in selecting the next best value based on the maximum f-score (f-measure) i.e., the trade-off between positive predictive value (PPV) and sensitivity[8]. The probabilistic linkage procedure is further summarised as follows[3,15,52,54]: -

Field probabilities

- ***Computing agreement probabilities***

  $m_i = P(r_i(a,b)/(a,b) \in L)$ →→ Probability [ith field] a given pair agrees, given that it is a true match/link.

- ***Computing disagreement probabilities***

  $u_i = P(r_i(a,b)/(a,b) \in N)$ →→ Probability [ith field] a given pair agrees, given that it is a true non-link.

Field weights

- ***Computing match weights for linked records and non-links***

  $w_{i1} = \log_2\left(\frac{m_i}{u_i}\right)$ →→ match weight if ith field agrees between two records.

$w_{i2} = \log_2\left(\frac{1-m_i}{1-u_i}\right)$ →→ match weight if i[th] field disagrees between two records.

☞ ***Computing composite weights (w)***

$w = \sum_{ij} w_{ij}$

$w = \log_2\left(\frac{m_{firstname}}{u_{firstname}}\right) + \log_2\left(\frac{m_{surname}}{u_{surname}}\right) + \log_2\left(\frac{m_{sa\,id}}{u_{sa\,id}}\right) + \ldots\ldots\ldots + \log_2\left(\frac{1-m_{dob}}{1-u_{dob}}\right)$

Threshold weight scores

☞ ***Setting the threshold value (T) in order to assign status to pairs***

Different analyses were carried out by varying the threshold cut-off points of the calculated composite weights based on the composite weight score using the selected matching variables as stated in the linkage step. The cut-off point maximising (minimising) f-score/f-measure was set as the upper (lower) threshold such that: -

$w \geq T_{max}$	-	Definitive match/link
$T_{min} \leq w < T_{max}$	-	Possible match/possible link
$T_{min} < w$	-	Don't match/non-link

Where:

m_i = Probability the i[th] field agrees given that two records match

$(1 - m_i)$ = Probability the i[th] field disagrees given that two records match

u_i = Probability the i[th] field agrees given that two records do not match

$(1 - u_i)$ = Probability the i[th] field disagrees given that two records do not match

r_i = i[th] record

a & b = Dataset A and B

L = Link

N = Non-link

w_{i1} = Match weight given the i[th] field agrees between two records

w_{i2} = Match weight given the i[th] field disagrees between two records

Clerical review

This involved manually inspecting all variables available for the record pairs with possible matches/links (i.e., record pairs having composite weight between lower and upper cut-off points - $T_{min} \leq w < T_{max}$) and deciding whether the record pairs are links or non-links. For a record pair to be deemed a link, it must match on at least two variables/fields. Record pairs which were categorized as non-links in this step were taken back to step 2 (blocking) for another comparison until the process was exhausted.

Linked datasets

The final dataset with linked record pairs between the two data files was obtained once the above steps were completed and exhausted. The linked dataset was then used in the final analysis i.e., comparing the performance of the provincial fuzzy linkage over PRL. Figure 8 below shows key steps followed in probabilistic record linkage.

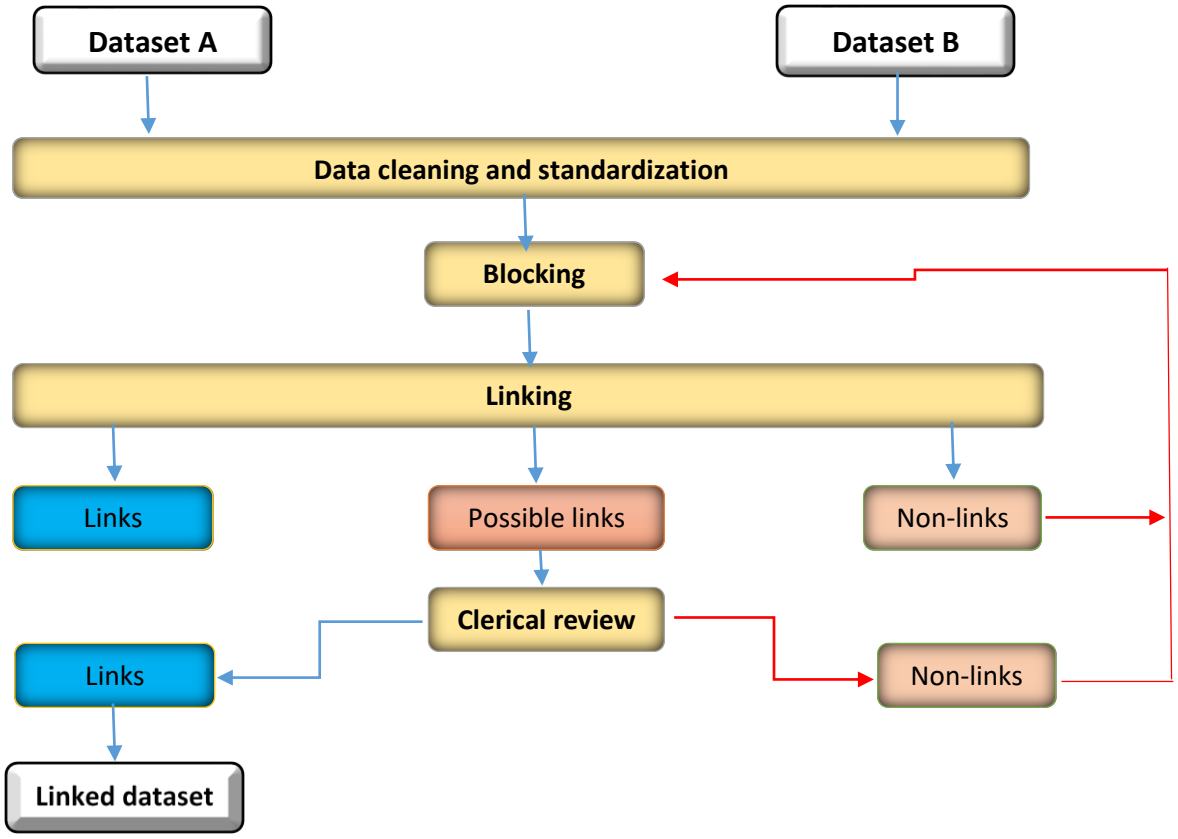


Figure 8: Probabilistic data linkage procedures

2. Evaluation of the Probabilistic Record linkage

Primary analyses

Standard diagnostic agreement measures were used to compare the performance of PRL and the decision-rule based provincial linkage. The independent PRL exercise was used as a reference to formally quantify the agreement of the existing linkage approach as a means of validating it for subsequent analyses. Assessing performance of the decision-rule based provincial linkage and exact matching over fully probabilistic record linkage approach involved the use of the following quality measures: sensitivity/recall/match rate ($\frac{TP}{TP+FN}$), specificity/true-negative rate ($\frac{TN}{TN+FP}$),

positive predictive values/precision/link accuracy ($\frac{TP}{TP+FP}$), negative predictive value ($\frac{TN}{TN+FN}$), accuracy ($\frac{TP+TN}{TP+FP+TN+FN}$), F-measures/F-score ($2x\left(\frac{PPV*Sensitivity}{PPV+Sensitivity}\right)$), false-negative rate ($\frac{FN}{TP+FN}$) and false-positive rate ($\frac{FP}{TN+FP}$). F-measure was considered "very good" when F-score ≥ 0.90, "relatively good" when 0.85 ≤ F-score < 0.90, fair when 0.80 ≤ F-score < 0.85, poor when 0.50 ≤ F-score < 0.85, and very poor when F-score ≤ 0.50 in determining linkage quality[55]. Using the binomial approximations, the 95% Confidence Intervals were estimated for the respective quality measures.

Sensitivity analyses

A composite dataset was constructed based on the combined linkages from both approaches, in which all discrepancies were manually reviewed and curated. Using this dataset, sensitivity analyses were conducted using five different sets of matching variables and different weight thresholds to examine the effects of varying linkage parameters in identifying record pairs. The first sensitivity analysis involved conducting probabilistic linkage using first name Soundex, surname Soundex, date of birth, SA ID, clinicom/folder number and the postcode. The second, involved conducting probabilistic linkage using first name, surname, date of birth, SA ID and the postcode. Thirdly, probabilistic linkage was performed using first name, surname Soundex, date of birth, SA ID and the postcode. The fourth, involved conducting probabilistic linkage using first name, surname, date of birth and SA ID. The fifth, probabilistic linkage was performed using first name, surname and date of birth. The above sensitivity analyses were conducted using a

threshold cut-off point of 0.9. Lastly, probabilistic linkage was performed by varying the threshold match score (T_{max}) from 0.6 to 1.0 without manual review.

4.4 Results

Description of the datasets

In total, 330,016 pregnancy records were extracted from the episodes data table and 204,923 death records from the encounters table in the Clinical database from 1st January 2010 to 30th December 2013. The mortality dataset included 51,203 deaths which occurred from 1st January to 31st December 2013. A total of 6,863 and 26,699 duplicate records in the mortality and pregnancy datasets were identified and deleted in the data cleaning stage. Additionally, restricting subjects to only "females" as well as ages ranging from 10 to 65 years, resulted to 33,880 records from mortality and 1,399 records from pregnancy dataset being deleted. The final probabilistic linkage dataset consisted of 301,918 and 10,437 records from the pregnancy and mortality dataset respectively. Figure 9 below presents the flow diagram of the datasets used for this study.

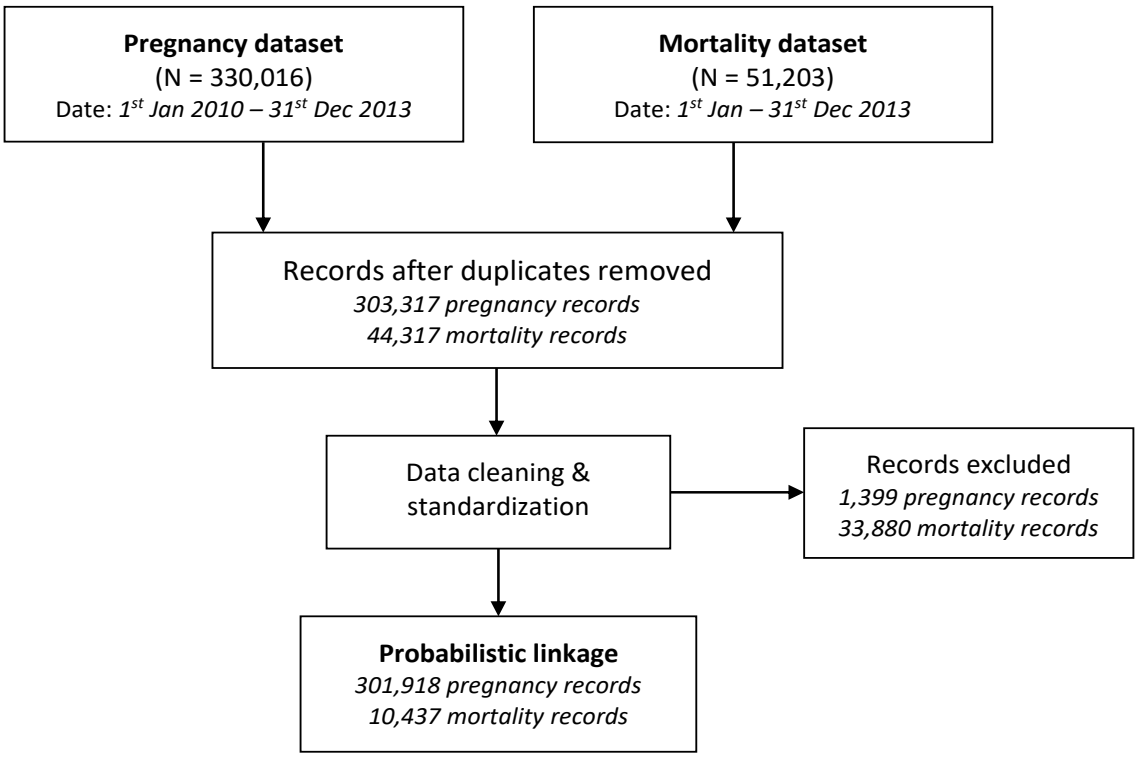


Figure 9: Flow diagram of the database used for the analysis.

Descriptive analyses of the linkage variables

Table 10 shows the descriptive analyses of the selected linkage variables according to data sources. Both pregnancy and mortality datasets had no record with missing information on surname and date of birth. Pregnancy dataset had one, 80,955 (36.6%) and 1,725 (0.6%) records missing first name, SA ID and postcode information respectively. A total of 626 (0.2%) records were missing both SA ID and postcode information; with no record missing information in more than two variables. The mortality dataset had 2,164 (26.2%), 406 (4.0%) and 277 (2.7%) records missing SA ID, postcode and folder number information respectively. Fourteen (0.1%) records had missing information in all three variables i.e., SA ID, postcode and folder number. 177 (1.7%)

records were missing postcode and folder number information and 100 (1.0%) records were missing both SA ID and postcode information.

Table 10: Descriptive analysis of the selected linkage variables according to data source

Linkage variables	Pregnancy dataset (N=301,918)		Mortality dataset (N=10,437)	
	Available records	Missing records	Available records	Missing records
First name	301,917	1	10,437	0
Surname	301,918	0	10,437	0
SA ID	220,963	80,955 (36.6%)	8,273	2,164 (26.2%)
Date of birth	301,918	0	10,437	0
Postcode	300,193	1,725 (0.6%)	10,031	406 (4.0%)
Folder number	301,918	0	10,160	277 (2.7%)

Record linkage

Exact matching of the pregnancy and mortality database using the same linkage variables as PRL (i.e., first name, surname, date of birth, SA ID, Postal code and folder number) identified 370 matched pairs. In total, 10,068 (3.2%) records in the mortality dataset and 301,552 (96.7%) records from the pregnancy dataset were unique.

Probabilistic records linkage of the pregnancy and mortality datasets identified 541 linked pairs. Following manual review of the 1,542 “possible linked” pairs, an additional 28 pairs were marked as links. In total, 571 matches were identified by the PHDC linkage (decision-rule based linkage approach) whereas 569 matches were identified by the probabilistic record linkage after manual review of the possible match pairs respectively.

Figure 10 shows the distribution of composite weight scores ≥ 0.7 before conducting manual review of the matched pairs. The optimal threshold used (0.9) excluded a large number of non-matches/non-links.

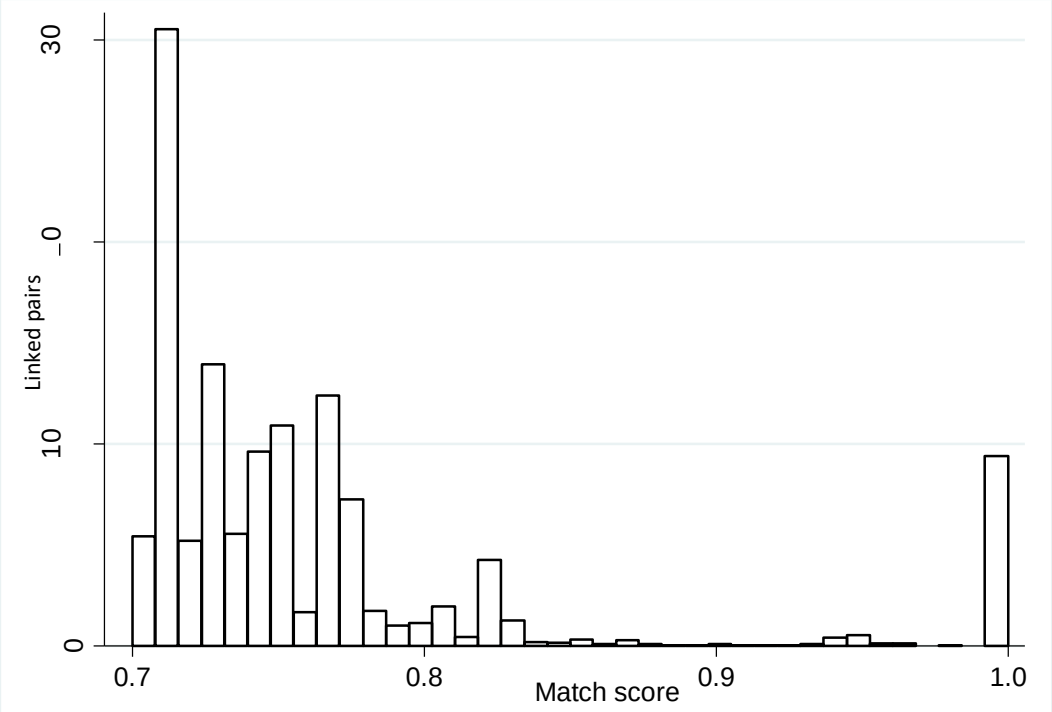


Figure 10: Distribution of agreement probabilities before manual review of record pairs

Comparison between decision-rule based linkage and PRL in identifying matches without manual review

Table 11 shows the comparison between the decision-rule based linkage approach used by the province and PRL method in identifying matches. Compared to the PRL without manual review of the possible links, PHDC linkage approach has classified 539 matches that are true matches and 9864 non-matches that are true non-matches. The decision-rule based linkage used by the

province has classified 32 matches that were non-matches and only 2 non-matches that were matches in PRL.

Table 11: Comparison of decision-rule based linkage method and PRL in identifying matches

Decision-rule based provincial linkage	**Probabilistic record linkage**		
	Links	Non-links	**Total**
Links	539	32	**571**
Non-links	2	9,864	**9,896**
Total	**541**	**9,866**	**10,437**

Comparison between decision-rule based linkage and PRL in identifying matches with manual review

Compared to PRL approach after manual reviews of all possible links, the decision-rule based linkage approach used by the province has classified 567 matches that are true matches and 9864 non-matches that are true non-matches. It has classified four matches that were non-matches and only two non-matches that were matches in the PRL approach, which in fact were true non-matches when interrogated further. Table 12 depicts the record pairs identified by the two linkage approaches as links and non-links.

Table 12: Comparison of decision-rule based linkage method and PRL in identifying matches

Decision-rule based provincial linkage	**Probabilistic record linkage**		
	Links	Non-links	**Total**
Links	567	4	**571**
Non-links	2	9,864	**9,866**
Total	**569**	**9,868**	**10,437**

Feasibility of the probabilistic approach

Table 13 presents the quality measure results of the compared linkage approaches, if the existing decision-rule-based approach is used as the reference. Although *a priori* it was decided to

compare the existing PHDC approach to an independent probabilistic data linkage output, the first comparison presented here shows the agreement of the PRL approach with the historical approach, to establish coherence of the independent approach. Agreement across all measures, compared to the decision-rule-based linkage approach used by the province, was predictably higher for the PRL than the exact matches, in particular when including manual review of uncertain matches.

Table 13: Quality results of PRL and exact matching as compared to the provincial linkage algorithm

Measure	**Exact Matching***	**PRL** (without manual review)*	**PRL** (with manual review)*	**Provincial linkage****
Accuracy	97.87%	99.67%	99.94%	99.94%
Sensitivity (Recall/match rate)	62.87%	94.40%	99.30%	99.65%
Specificity (true-negative rate)	99.90%	99.98%	99.98%	99.96%
Precision (positive predictive value)	97.29%	99.63%	99.65%	99.30%
Negative predictive value	97.90%	99.67%	99.96%	99.98%
False-negative rate	37.13%	5.60%	0.70%	0.04%
False-positive rate	0.10%	0.02%	0.02%	0.35%
F-measure (F-score)	76.0%	97.0%	99.47%	99.50%

**Compared to provincial linkage approach with manual review (PHDC linkage-Reference)*
***Compared to PRL with manual review (PRL- Reference)*

Sensitivity analyses

Robustness of PRL when excluding the unique identifier

When excluding folder numbers in the PRL linkage variables, and without manual review of the possible links, there were 547 concordant matches and 9801 concordant non-matches. Provincial linkage classified 24 matches that were PRL non-matches and 65 non-matches that were PRL

matches. In general, provincial linkage had a lower sensitivity and precision when PRL excluded folder numbers in the linkage variables compared to its inclusion i.e., 89.38% vs. 99.65% sensitivity and 95.78% vs 99.30% precision respectively. Table 14 depicts comparison between decision-rule based linkage approach and the PRL in identifying matches when excluding folder number without a manual review of the possible linked pairs.

Table 14: Comparison of decision-rule based linkage method over PRL in identifying matches when excluding the folder number

Decision-rule based provincial linkage	**Probabilistic record linkage n (%)**		**Total**
	Links	Non-links	
Links	547	24	**571**
Non-links	65	9801	**9866**
Total	**612**	**9825**	**10437**

Sensitivity analyses based on PRL variable inclusion

Table 15 shows number of matched pairs identified by different linkage variables using a subset of data confirmed to be links/matches. Including many linkage variables (five variables) resulted in a higher match rate (>94%) and precision (>87%) than including fewer (three) linkage variables (93.52% match rate and 81.53% precision respectively). When excluding both SA ID, postcode and folder number information in the selected linkage variables, the identified number of linked pairs increased slightly but not as much as when excluding postcode and folder number i.e., 655 vs. 760 matches respectively.

Table 15: Sensitivity analyses when using different linkage variables without manual review

Linking variables	**PRL matches**	**Sensitivity**	**Specificity**	**Precision**
Sensitivity analysis 1	540	94.05%	99.97%	99.44%
Sensitivity analysis 2	612	95.80%	99.34%	89.38%

Sensitivity analysis 3	623	95.45%	99.21%	87.48%
Sensitivity analysis 4	760	94.75%	97.78%	71.18%
Sensitivity analysis 5	655	93.52%	98.77%	81.53%

Sensitivity analysis 1 – First name Soundex, last name Soundex, date of birth, SA ID, postcode & folder number.
Sensitivity analysis 2 – First name, last name, date of birth, SA ID & postcode.
Sensitivity analysis 3 – First name, last name Soundex, date of birth, SA ID and postcode.
Sensitivity analysis 4 – First name, last name, date of birth and SA ID.
Sensitivity analysis 5 – First name, last name and date of birth.

Varying the threshold match score (T_{max}) from 0.9 to 1.0 (exact match) resulted in a reduced number of matches, an increased number of true negatives and false positives as well as decreased number of true positives and false negatives. Lowering the threshold parameter (T_{max}) from 0.8 to 0.6 resulted in a marked increase in the number of matches, reduction of true negatives and false positives as well as an increased number of true positives and false negatives. Table 16 shows the sensitivity analyses when varying the threshold parameter.

Table 16: Sensitivity analyses when varying agreement probability threshold without manual review

Threshold weight score	PRL matches	Sensitivity	Specificity	Precision
1.0	353	61.82%	100.0%	100.0%
0.9	541	94.40%	99.98%	99.63%
0.8	628	98.25%	93.98%	89.33%
0.7	2065	99.30%	84.82%	27.46%
0.6	5,877	99.65%	46.20%	9.68%

Finally, the probabilities calculated by the *Reclink* routine were verified through manual coding of the probabilities (data not shown).

4.5 Discussion

Recent studies have highlighted the need for comparative evaluations of data linkage methods[8,56]. This study determined the feasibility of implementing PRL and validated the

performance of a fuzzy linkage approach used by the WC province in linking provincial health data as compared to a fully probabilistic data linkage. The PRL when including manual review of possible matches was shown to be feasible through a high level of agreement with the historical method.

Using the independent PRL linkage as a reference, few discrepancies were observed with the decision-rule based provincial linkage approach. Both approaches (provincial linkage and PRL) presented high and relatively equal performance in identifying record pairs based upon key identifiers i.e., names, date of birth, SA ID, postal code and folder number. Based on review of previous linkage studies using both real-life and simulated datasets, PRL has demonstrated a better performance than deterministic linkage method/exact matching in most cases[16–22]. In this study, PRL has proven feasible in identifying linked pairs as demonstrated by high levels of agreement with the current method when including first name, last name, date of birth, SA ID, postal code and folder number information; and using a threshold match score of 0.9. Relatively high accuracy and precision have been observed even after excluding postcode and folder number information from the set of linkage variables.

The appropriateness of PRL to the data context, which frequently has missing values for key identifiers

Folder number and SA ID are important unique identifiers and perhaps key variables for data linkage in the exact matching procedures. However, they are not reliably or always collected as demonstrated by a considerable number of cases with missing SA ID number in our datasets (i.e., 36.6% in pregnancy and 26.2% in mortality datasets) and missing folder numbers (i.e., 2.7% in

mortality dataset). Previous studies have reported the rates of missing data and errors in key linkage variables ranging from 0 to 10% and 4 to 15% respectively[57–61]. High proportions of missing data have further been linked with high false negative links which compromise sensitivity in a deterministic linkage thus necessitating the applications of PRL[3,62]. Despite relatively high proportions of missing SA ID information in both datasets, the accuracy and precision of PRL remained reasonably high when SA ID is included in the linkage variables. This can be viewed as a result of imperfect matches due to errors or missing data resulting in false negative links in the exact matching being captured as links by the PRL. This observation highlight the strength of PRL over exact matching when there is high proportions of error or missing information in linkage variables as also reported elsewhere[3,18,62].

Trade-off of shifting the thresholds in terms of how many records require manual review and loss of performance i.e., tuning/optimisation of the PRL method.

It is reported that, having large number of true negative links biases the sensitivity, specificity and false negative rates since the number of true negatives dominate in their respective formulae[8,63]. As a consequence, positive predictive value has been claimed to be the most adequate and recommended quality measure over sensitivity, specificity and false negative rates in assessing the quality of the data linkage approach[8]. In this regard, PRL has still demonstrated high ability of classifying matches that are true matches in our study by having notably high positive predictive value (99.7%). This study supports evidence from previous studies that PRL has sufficient accuracy and high precision in matching health datasets[17,20–22,63–65].

Although, optimising quality measures is ideal, some authors have suggested maintaining high PPV over the sensitivity since false positive links - as compared to false negative links, tends to bias risk ratio and risk difference to the null[60,66,67]. Nevertheless, there is a trade-off between sensitivity and PPV such that, high PPV can only be attained at the expense of low sensitivity and vice versa[8,18]. For this reason, finding the compromise between sensitivity and PPV is advised, particularly using the maximum f-measure (f-score)[8]. This is the only quality measure which captures the trade-off between sensitivity and PPV[8,55]. Our study has obtained remarkably high f-measure/f-score i.e., 0.99%. According to Ferrante and Boyd, f-measure ≥0.90 shows a very good linkage quality[55]. However, there is no suggested cut-off values for these quality measures from previous linkage studies[8,20,59,68].

Fidelity of existing automated linkage in the Western Cape PHDC

The existing linkage algorithm has provided high levels of agreement with the independent PRL approach in matching health data in our setting. These findings suggest that the provincial linkage approach can be used to generate datasets which can be used to estimate burden of diseases for a better understanding of the trends, as well as assessments of the successes or failures of various interventions in the WC province and the country at large.

Strengths and limitations

The key strengths of this study include large size of matching datasets, availability of multiple unique identifiers across datasets, combination of both strings and numeric linkage variables, availability of multiple linkage variables and presence of missing values which complicated the linkage procedures and proved its robustness. Furthermore, the author's ability to review

manually the large number of “possible links” in this exercise has improved the linkage quality of the dataset used for assessment, by curating additional correct links.

The generalisation of performance of the decision-rule based provincial linkage approach might be limited by the fact that only one linkage example was explored, and the characteristics of other datasets might well differ.

4.6 Conclusion

The decision-rule based provincial linkage and PRL approaches have both been demonstrated to be feasible on a sample of datasets on which this thesis is based, and both have demonstrated high levels of agreement in matching health records. This provides reassurance on the performance of the record-linkage on which subsequent analyses rely.

Health officers and researchers can confidently continue using the linked (using the decision-rule based provincial linkage approach) WC provincial health data for healthcare administration, business intelligence and epidemiological research activities due to robustness of the linkage approach used.

4.7 References

1. Nicholls, S. G. *et al.* The REporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) Statement: Methods for Arriving at Consensus and Developing Reporting Guidelines. *PLoS One* **10**, e0125620 (2015).

2. Dean, B. B. *et al.* Review: Use of Electronic Medical Records for Health Outcomes Research. *Med. Care Res. Rev.* **66**, 611–638 (2009).

3. Fellegi, I. P. & Sunter, A. B. A Theory for Record Linkage. *Source J. Am. Stat. Assoc.* **64**, 1183–1210 (1969).

4. Gill, L., Goldacre, M., Simmons, H., Bettley, G. & Griffith, M. Computerised linking of medical records: methodological guidelines. *J. Epidemiol. Community Health* **47**, 316–9 (1993).

5. Levenshtein, V. I. & I., V. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Sov. Phys. Dokl. Vol. 10, p.707* **10**, 707 (1966).

6. Roos, L. L., Wajda, A. & Nicol, J. P. The art and science of record linkage: Methods that work with few identifiers. *Comput. Biol. Med.* **16**, 45–57 (1986).

7. Melaku, Y. A. *et al.* Poor linkages in maternal health care services ? evidence on antenatal care and institutional delivery from a community-based longitudinal study in Tigray region , Ethiopia. 1–13 (2014). doi:10.1186/s12884-014-0418-7

8. Christen, P. & Goiser, K. *Quality and complexity measures for data linkage and deduplication*. *Quality Measures in Data Mining* (2007).

9. Grannis, S. J., Overhage, J. M. & McDonald, C. J. Analysis of identifier performance using a deterministic linkage algorithm. *Proceedings. AMIA Symp.* 305–9 (2002).

10. Dimick & Chris. Exposing Double Identity at Patient Registration. *J. AHIMA* **80**, web extra (2009).

11. Hakimzada, A. F., Green, R. A., Sayan, O. R., Zhang, J. & Patel, V. L. The nature and occurrence of registration errors in the emergency department. *Int. J. Med. Inform.* **77**, 169–75 (2008).

12. Christen, P. *Data matching : concepts and techniques for record linkage, entity resolution, and duplicate detection*. (Springer, 2012).

13. Herzog, T. N., Scheuren, F. J. & Winkler, W. E. *Data Quality and Record Linkage Techniques*. (Springer Science & Business Media, 2007).

14. Sayers, A., Ben-Shlomo, Y., Blom, A. W. & Steele, F. Probabilistic record linkage. *Int. J. Epidemiol.* **45**, 954–964 (2016).

15. Chevrette, A. & Wenzowski, M. G-Link: A probabilistic record linkage system. *Pers. Valid. Entity Resolut. Conf.* (2011).

16. da Silveira, D. P. & Artmann, E. Accuracy of probabilistic record linkage applied to health databases: Systematic review. *Rev. Saude Publica* **43**, 875–882 (2009).

17. Grannis, S. J., Overhage, J. M., Hui, S. & McDonald, C. J. Analysis of a probabilistic record linkage technique without human review. *AMIA ... Annu. Symp. proceedings. AMIA Symp.* 259–63 (2003).

18. Ying Zhu, Matsuyama, Y., Yasuo Ohashi & Setoguchi, S. When to conduct probabilistic linkage vs deterministic linkage? A simulation study. *J. Biomed. Inform.* **56**, 80–8 (2015).

19. Tromp, M., Ravelli, A. C., Bonsel, G. J., Hasman, A. & Reitsma, J. B. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J. Clin. Epidemiol.* **64**, 565–572 (2011).

20. Méray, N., Reitsma, J. B., Ravelli, A. C. J. & Bonsel, G. J. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J. Clin. Epidemiol.* **60**, 883–91 (2007).

21. Corbell, C. *et al.* Records linkage of electronic databases for the assessment of adverse effects of antiretroviral therapy in sub-Saharan Africa. (2011). doi:10.1002/pds

22. Aldridge, R. W., Shaji, K., Hayward, A. C. & Abubakar, I. Accuracy of probabilistic linkage using the enhanced matching system for public health and epidemiological studies. *PLoS One* **10**, 1–15 (2015).

23. Quantin, C. *et al.* Which are the best identifiers for record linkage? *Med. Inform. Internet Med.* **29**, 221–227 (2004).

24. Mason, C. A. & Tu, S. Data linkage using probabilistic decision rules: A primer. *Birth Defects Res. Part A Clin. Mol. Teratol.* **82**, 812–821 (2008).

25. UNICEF, WHO, World Bank group & United Nations. *Levels & Trends in Child Mortality - Report 2015: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. *UNICEF* (2015). doi:10.1371/journal.pone.0144443

26. WHO, UNICEF, UNFPA, World Bank & United Nations Population Division. Trends in maternal mortality 1990 to 2015. (2015). doi:10

27. Moodley, J. *Saving Mothers 2011-2013: Sixth report on confidential enquiries into maternal deaths in South Africa, Short report*. *Pretoria: Department of Health* (2014).

28. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384**, 980–1004 (2014).

29. Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 957–979 (2014).

30. Saleem, S. *et al.* A prospective study of maternal, fetal and neonatal deaths in low- and middle-income countries. TT -. *Bull. World Health Organ.* **92**, 605–612 (2014).

31. Pattinson, R. & Rhoda, N. *Saving Babies 2012-2013: Ninth report on perinatal care in South Africa*. *RC Pattinson* (2014).

32. Chopra, M., Daviaud, E., Pattinson, R., Fonn, S. & Lawn, J. E. Saving the lives of South Africa's mothers, babies, and children: can the health system deliver? *Lancet* **374**, 835–846 (2009).

33. United Nations. Handbook on Training in Civil Registration and Vital Statistics Systems. *United Nations* 281 (2002).

34. Atrash, H. K., Alexander, S. & Berg, C. J. Maternal mortality in developed countries: not just a concern of the past. *Obstet. Gynecol.* **86**, 700–5 (1995).

35. Hill, K. *et al.* How should we measure maternal mortality in the developing world? A comparison of household deaths and sibling history approaches. *Bull. World Health Organ.* **84**, 173–80 (2006).

36. Stanton, C. *et al.* Every death counts: measurement of maternal mortality via a census. *Bull. World Health Organ.* **79**, 657–64 (2001).

37. WHO. *WHO guidance for measuring maternal mortality from a census*. *World Health Organisation* (2008).

38. Dorrington, R., Bradshaw, D. & Laubscher, R. *Rapid Mortality Surveillance Report 2012*. (2014). doi:ISBN: 978-1-920618-00-1

39. Bradshaw, D. *et al.* Every death counts: use of mortality audit data for decision making to save the lives of mothers, babies, and children in South Africa. *Lancet* **371**, 1294–1304 (2008).

40. Moodley, J. *et al.* The confidential enquiry into maternal deaths in South Africa: a case study. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 53–60 (2014).

41. Souza, J. P. *et al.* Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *The Lancet* **381**, 1747–1755 (2013).

42. Republic of South Africa. Births and Deaths Registration Act No. 51 of 1992. Regulations on the Registration of Births and Deaths. *Government Gazette* 74 (2014).

43. Groenewald, P. *et al.* The importance of identified cause-of-death information being available for public health surveillance, actions and research. *South African Med. J.* **105**, 528–530 (2015).

44. Dimick & Chris. Exposing Double Identity at Patient Registration. *J. AHIMA* **80**, web extra (2009).

45. Hakimzada, A. F., Green, R. A., Sayan, O. R., Zhang, J. & Patel, V. L. The nature and occurrence of registration errors in the emergency department. *Int. J. Med. Inform.* **77**, 169–75 (2008).

46. Jaro, M. A. Probabilistic linkage of large public health data files. *Stat. Med.* **14**, 491–498 (1995).

47. Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P. & Fienberg, S. Adaptive name matching in information integration. *IEEE Intell. Syst.* **18**, 16–23 (2003).

48. Statistics South Africa. *Statistical release P0302: Mid - year population estimates 2017*. (2017).

49. Western Cape Provincial Centre Health Data Centre. *Strategic approach to patient - level health data harmonisation and integration Western Cape Department of Health International Classification of Diseases*. (2013).

50. Gill, L. *Methods for automatic record matching and linkage and their use in national statistics / Leicester Gill. - Version details - Trove*. (Office for National Statistics, 2001).

51. Guiver, T. *Sampling-based clerical review in probabilistic linking. Australian Bureau of Statistics* (2011).

52. Sariyar, M. & Borg, A. The RecordLinkage Package: Detecting Errors in Data. *R J.* **2**, 61–67 (2010).

53. Michelson, M. & Knoblock, C. A. Learning blocking schemes for record linkage. 440–445 (2006).

54. Bigelow, W., Karlson, T. & D, P. Using Probabilistic Linkage to Merge Multiple Data Sources for Monitoring Population Health. *Syst. Res.* (1999).

55. Ferrante, A. & Boyd, J. A transparent and transportable methodology for evaluating Data Linkage software. *J. Biomed. Inform.* **45**, 165–172 (2012).

56. Clark, D. E. Practical introduction to record linkage for injury research. *Inj. Prev.* **10**, 186–91 (2004).

57. Setoguchi, S., Zhu, Y., Jalbert, J. J., Williams, L. A. & Chen, C.-Y. Validity of deterministic record linkage using multiple indirect personal identifiers: linking a large registry to

claims data. *Circ. Cardiovasc. Qual. Outcomes* **7**, 475–80 (2014).

58. Bell, R. M., Keesey, J. & Richards, T. The Urge to Merge: Linking Vital Statistics Records and Medicaid Claims. *Med. Care* **32**, 1004–1018 (1994).

59. Muse, A. G., Mikl, J. & Smith, P. F. Evaluating the quality of anonymous record linkage using deterministic procedures with the New York state aids registry and a hospital discharge file. *Stat. Med.* **14**, 499–509 (1995).

60. Howe, G. R. Use of Computerized Record Linkage in Cohort Studies. *Epidemiol. Rev.* **20**, 112–121 (1998).

61. Harron, K., Hagger-Johnson, G., Gilbert, R. & Goldstein, H. Utilising identifier error variation in linkage of large administrative data sources. (2017). doi:10.1186/s12874-017-0306-8

62. Newcombe, H. B., Kennedy, J. M., Axford, S. J. & James, A. P. Automatic Linkage of Vital Records: Computers can be used to extract "follow-up" statistics of families from files of routine records. *Science (80-.).* **130**, 954–959 (1959).

63. Zhu, Y., Matsuyama, Y., Ohashi, Y. & Setoguchi, S. When to conduct probabilistic linkage vs. deterministic linkage? A simulation study. *J. Biomed. Inform.* **56**, 80–86 (2015).

64. Silveira, D. P. da & Artmann, E. Accuracy of probabilistic record linkage applied to health databases: systematic review. *Rev. Saude Publica* **43**, 875–82 (2009).

65. Tromp, M., Ravelli, A. C., Bonsel, G. J., Hasman, A. & Reitsma, J. B. Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage. *J. Clin. Epidemiol.* **64**, 565–572 (2011).

66. Blakely, T. & Salmond, C. Probabilistic record linkage and a method to calculate the positive predictive value. *Int. J. Epidemiol.* **31**, 1246–1252 (2002).

67. Copeland, K. T., Checkoway, H., McMichael, A. J. & Holbrook, R. H. Bias due to misclassification in the estimation of relative risk. *Am. J. Epidemiol.* **105**, 488–95 (1977).

68. Gomatam, S., Carter, R., Ariet, M. & Mitchell, G. An empirical comparison of record linkage procedures. *Stat. Med.* **21**, 1485–1496 (2002).

CHAPTER 5: A three-source capture-recapture estimate of maternal deaths under-reporting in the Western Cape Province, South Africa

5.1 Abstract

Background: Maternal mortality is an important health challenge and insufficient progress has been made in addressing it. South Africa and particularly the Western Cape province has multiple data sources with wide coverage which offer unique opportunities to estimate these outcomes empirically or analytically. Capture-recapture is a useful approach to resolve uncertainties in estimating population sizes, including those with specific health conditions or outcomes, when there has been underreporting. However, there has been no attempt to estimate maternal mortality under-ascertainment in the Western Cape province using the capture-recapture method.

Methods: Maternal deaths reported by the Department of Home Affairs (DHA), institutional reporting (CEMD), and hospital death reporting system (PHDC), from 2010 to 2013, were linked for potential duplicates, before applying a three-source capture-recapture method to estimate maternal deaths' under-reporting. Evaluation of dependence between data sources was determined using log-linear models and the sample coverage estimator. Testing the goodness-of-fit of the log-linear models involved the use of likelihood ratio statistics (G^2), the Pearson goodness-of-fit statistics (χ^2), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The estimates of maternal deaths were presented annually, and combined from 2010 to 2013, with corresponding 95% Confidence Intervals (95%CI).

Results: In total, 238 maternal deaths were registered by the institutional reporting (CEMD), 108 by the DHA and 205 hospital deaths from the reporting system (PHDC). Of these, 391 (70.1%) were unique cases. Only 22 maternal deaths were registered by all three sources. The capture-recapture method estimated 719 (95% CI: 605, 879) maternal deaths between 2010 and 2013, with an estimated 328 (45.6%) maternal deaths not registered by any of the three sources. By calendar year, 150 (95% CI: 117, 207) maternal deaths were estimated in 2010; 108 (95% CI: 94, 131) in 2011, 313 (95% CI: 185, 678) in 2012; and 153 (95% CI: 124,202) in 2013. The completeness of registrations of the three sources combined was 54.38%, with CEMD having the highest level of completeness (33.1%).

Conclusion: The capture-recapture approach applied to three available data sources estimated substantial under-reporting of maternal mortality in the Western Cape between 2010 and 2013. Relying on estimates from a single source could mask the actual burden of the problem thus affecting the evaluation of interventions focusing on reducing maternal mortality, understanding trends, assessing progress against global commitments and evidence-based decision making.

Keywords: Capture-recapture, estimate, maternal, neonatal, mortality, log-linear, Western Cape, South Africa.

5.2 Introduction

Maternal mortality is an important health challenge and insufficient progress has been made in addressing it. Estimation of this outcome in most developing countries is challenging, particularly due to a lack of accurate, valid and reliable data[1–4]. Nonetheless, monitoring progress in global goals (Millennium Development Goals (MGD) 5 and Sustainable Development Goal (SDG) 3.1) requires reliable and accurate estimates of maternal mortality at a national and sub-national level in a given country.

Compared to other African countries, and as illustrated in chapter one, global estimates of South African maternal mortality often conflict with in-country institutional reporting, and both have wide uncertainty intervals[5–11]. Estimation of maternal mortality in the country is primarily based on the vital registration as well as the Confidential Enquiry into Maternal Deaths (CEMD)[12–15]. Nevertheless, South Africa is unique in the region in that, the country and the Western Cape province in particular provide opportunities to estimate this outcome empirically or analytically by having multiple data sources with wide coverage. Additionally, the country is unusual among developing countries in that, the national facility-based mortality audits for maternal, child and perinatal deaths are effectively carried out[12,14]. The availability of multiple data sources in the country and in the WC province facilitates the application of capture-recapture methods[5,6,8–10,12–16].

As introduced in chapter two, capture-recapture, also known as *dual-system estimation* or *dual-record system* when having two data sources and *multiple-system approach* or *multiple-record system* when having more than two sources, is a common method used by ecologists and

biologists to estimate wildlife population size[17,18]. Since it is not possible to count every animal in the specified setting, and detection is imperfect, closed capture-recapture models were developed to obtain an accurate estimate of the population size. This method is useful in resolving uncertainties in estimating population sizes (including of those with specific health conditions or outcomes) that have diverse estimates, since it doesn't rely on complete case ascertainment but rather it corrects underreporting by estimating the number of missed cases in the population.

Despite originally being developed for ascertaining the size of animal populations in wildlife biology and ecology, these methods are now applied to human populations particularly in human epidemiology, demography and medicine[17–21]. The emerging application of capture-recapture methods in estimating maternal mortality, corrects for under-reporting of this outcome by estimating the overall number of maternal deaths using multiple data sources available in a particular area[22–25]. Having independent data sources that measure the same outcome, the capture-recapture method allows one to estimate the number that are not on any list and hence to estimate the total population size. In this context, data sources which are also considered as samples might be hospital records, birth and death registries, and other related databases[21,26,27]. Having dependence between data sources can bias results[21]. However, it is not possible to estimate dependence in a *dual system* but it is possible with the *multiple system estimation*. The effect of dependencies can be controlled statistically through the application of ecological models, log-linear models and sample coverage approaches. These methods have been extensively discussed elsewhere and are proposed to handle dependencies among data

sources[20,28–36]. The methodology and assumptions of capture-recapture methods were discussed in further detail in Chapter two.

Due to varying estimates of maternal mortality provided by different and limited sources in South Africa, there is value in applying additional estimation techniques which utilise multiple data sources to correct for potential underreporting of this outcome. Despite its limitations when applied to human epidemiology, capture-recapture methods have been identified as a simple, quick and cheap alternative method for ascertaining reliable estimates of population size when estimating conditions with diverse estimates[17,33,37]. The available multiple data sources in South Africa and particularly the Western Cape province, provides unique opportunities for estimating maternal mortality by facilitating the application of the capture-recapture method.

The primary aim of this chapter was to estimate maternal mortality under-reporting; a secondary aim was assessing the completeness of data sources that register maternal deaths, individually and combined (case-ascertainment).

5.3 Methods

Design and Setting

This study used routinely collected cross-sectional clinical and administrative data to estimate the underlying burden of maternal mortality in the Western Cape Province, South Africa. The province has an estimated population size of 6.5 million people, with the 72% of population using public health sector facilities[38,39].

Case definition and analysis period

In this chapter, maternal death was defined as the death of a woman while pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes. We restricted our analyses to maternal deaths which occurred between 1st January 2010 to 31st December 2013, due to data availability and consistency in reporting mortality data across sources in the province. As the study period concluded long before this analysis, there is reasonable assurance of completeness of death registration across sources.

Data sources

Data used in this chapter came from the Western Cape Department of Health (WCDOH) and the South African Department of Home Affairs (DHA). As broadly discussed in chapter 2; in the Western Cape province, the main data sources registering maternal death records are hospital death reporting irrespective of cause in patients known to be pregnant or recently pregnant, clinical audit tools specific to maternal mortality (CEMD/institutional reporting) and death certificates identifying maternal causes of death. For this analysis, data from hospital deaths in patients with known pregnancies, irrespective of this being mentioned on the death certificate, with pregnancy ascertained from multiple sources (e.g., admitted in maternity ward, having laboratory results specific to pregnancy or delivery), were extracted from the PHDC SQL database and used as one of the three data sources.

The data from the CEMD served as the second data source for maternal mortality ascertainment and estimation. Due to limited access to cause-of-death data as stipulated in chapter 2, our

analysis used DHA data from 2010 to 2013 as a third source for capture-recapture analysis. Due to data availability and consistency in reporting maternal deaths across three sources, we limited our analyses to data covering the period 1st January 2010 to 31st December 2013. The data from the clinical system were linked to data from the DHA for the period 2010 to 2013 through the common unique identifiers (PMI) in order to identify duplicates before applying the capture-recapture method (Figure 2). The detailed descriptions of data sources providing datasets for our analyses are presented in Chapter 2.

Data processing and analysis

Data processing was undertaken using Stata software version 14.2 (Stata Corporation, College Station, Texas, USA). This involved data cleaning, standardisation and record linkage. Capture histories were obtained through record linkage using the provincial linkage algorithm (since the PHDC linkage approach was found to be comparable to an independent probabilistic linkage exercise in the previous chapter). Multiple-record-per-person data were converted to single record data with capture histories reflected by indicator variables. The three-source capture-recapture analyses were carried out using a Stata user-written programme "recap"[40] (Stata software version 15.1, Stata Corporation, College Station, Texas, USA) and "Rcapture" package in R software version 3.5.0 (R Foundation for Statistical Computing, Vienna, Austria). Additional analyses employing a sample coverage approach were carried out using the "CARE1" package in S-plus. Graphical data exploration was used to check whether the homogeneity of catchability and independence assumptions were met to facilitate model selection. Figure 11 below depicts data management and analytical procedures involved in this chapter.

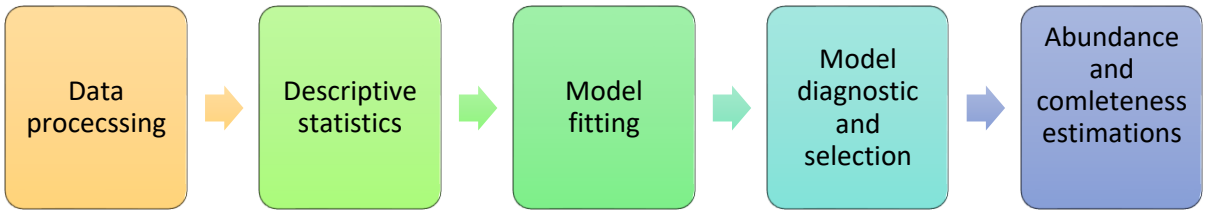


Figure 11: A flow chart indicating analytical procedures for a closed population CRC modelling

Capture-recapture analysis

We fitted three-source capture-recapture models to estimate maternal mortality under-reporting in the Western Cape province. Except for independence, other underlying assumptions for applying capture-recapture methods were 'satisfactorily' met.

In this study, we fixed death registrations in a defined time (2010 - 2013) and in a specified area (Western Cape province). Using personal identifiers and mainly the PMI, it was possible to link deaths registered by all three sources when these referred to the same woman. Due to the nature of the data sources used, the probability of registration is unlikely to depend on individual characteristics; although perhaps the probability of registration in PHDC might have been associated with the type of facility a woman was attending (whether public or private). Nevertheless, the probability of a death being registered in PHDC and CEMD as well as PHDC and DHA for a particular woman, was expected to be positively corelated since PHDC registers deaths from public sector facilities which usually require the providers providing death notifications to the NCCEMD and DHA through the respective forms. Lack of independence, if not controlled in the analysis stage, is therefore be expected to underestimate the number of maternal deaths.

Evaluation of dependence between sources was done by fitting three classes of models as described in Chapter 2 i.e., the log-linear models, ecological models and sample coverage approach[20,28–36]. The model giving the best fit to the data for the combined data (full four-years) was selected and fitted to the yearly data in order to provide yearly estimates.

Maternal mortality estimation

Estimations of maternal deaths were conducted yearly and combined following the applications of the ecological and log-linear capture-recapture models. Estimates were presented as maternal deaths with their corresponding 95% Confidence Intervals (CIs). Estimates obtained from the different model classes were compared for the two scenarios i.e., dependence and independence assumption.

Estimation of completeness

Completeness was estimated for each individual data source and combined, as a proportion of the estimated maternal deaths. Completeness of a source was obtained by dividing the number of maternal deaths registered by that source over the total number of deaths estimated by the capture-recapture method. The overall completeness level was also obtained by dividing the number of unique maternal deaths registered by the three sources over the estimated maternal deaths by the log-linear capture recapture model.

Sensitivity analyses

Three sensitivity analyses were conducted using 75%, 50% and 25% of non-overlapping PHDC cases (i.e., these cases were only ascertained by the PHDC) to see the effects of potential over-ascertainment of pregnancy by the PHDC on the final maternal mortality estimates using the combined data (four years of data). Although data obtained from the PHDC were limited to a high confidence score i.e., high probability that pregnancy and death records belong to a same person (as presented in Chapter 2 on PHDC data linkage), there is still a possibility that a certain proportion of PHDC maternal deaths were cases, in which pregnancy was incorrectly ascertained or incidental.

A possible explanation is that, confidence is largely driven by birth records, and pregnancies resulting in death are unlikely to have birth records. Due to the nature of PHDC data (as discussed in Chapter 2); even if data were perfect, we expect a certain proportion of the deaths not to be pregnancy-related but rather incidental. These highlight the importance of conducting sensitivity analyses to validate the obtained estimate of maternal mortality.

5.4 Results

Distribution of maternal deaths registration by calendar years and data sources

Between 1st January 2010 and 31st December 2013, a total of 205 maternal deaths were registered by the hospital death reporting system (PHDC), 238 by the institutional reporting/clinical audits tool (CEMD) and 108 through the death certificates by the DHA. The hospital death reporting system (PHDC) and institutional reporting (CEMD) have registered

almost twice the number of maternal deaths reported by DHA over calendar years and combined. Figure 12 below depicts the trends of maternal deaths according to data sources in the Western Cape province, South Africa.

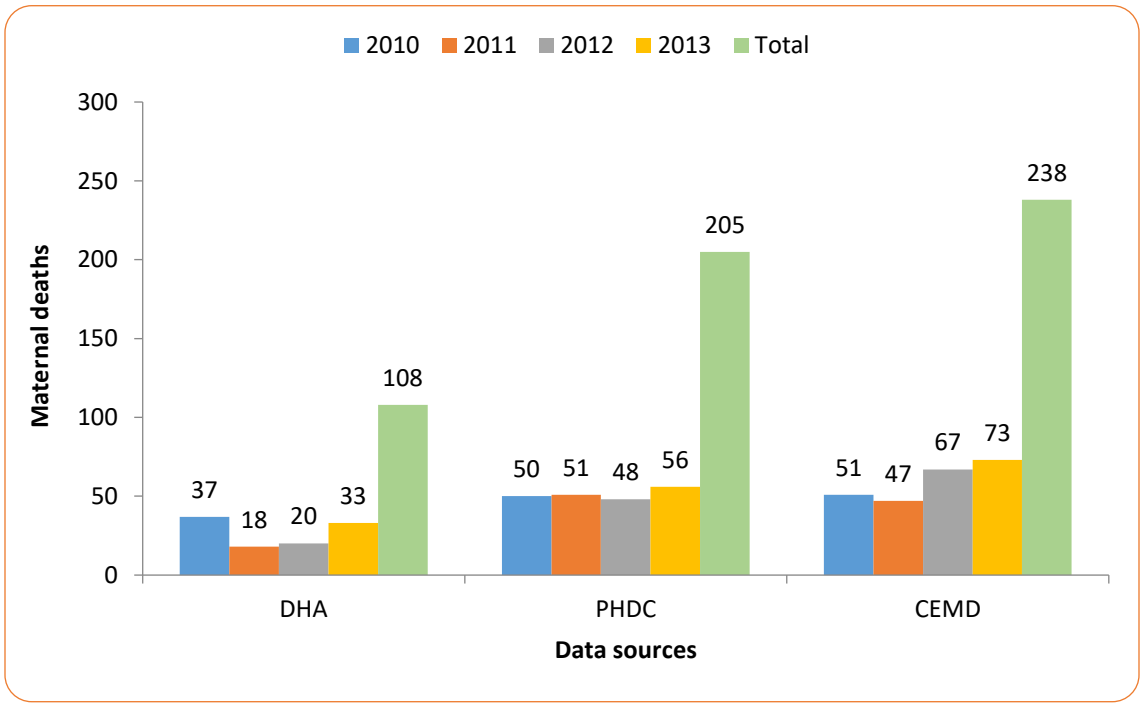


Figure 12: Distribution of maternal deaths by year of registration and data sources

The proportional Venn diagram below presents the distribution of maternal deaths registered uniquely and across (overlap between sources) each source (Figure 13). For the 4-year period, only 22 maternal deaths were registered by all three sources. Nine maternal deaths were registered by DHA and PHDC; 19 by CEMD and DHA; and 88 by PHDC and CEMD. A total of 391 (71.0%) unique maternal deaths were registered by the three sources during the analysis period, of which; DHA registered 58, PHDC 86 and CEMD 109 unique maternal deaths.

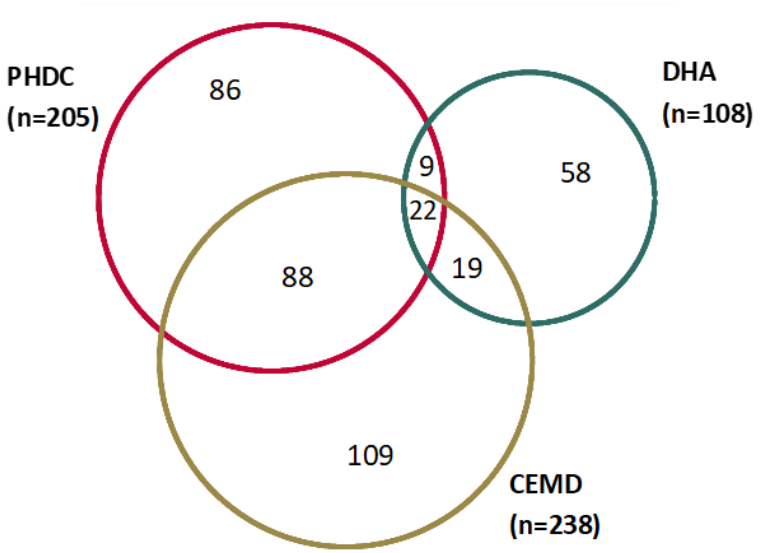


Figure 13: Distribution of maternal deaths registered by the three sources (2010 - 2013).

Ecological models

Results of the closed-population ecological models fitted to data covering the period 2010 to 2013 are presented in table 17 below. Based on the exploratory heterogeneity graph (fi plot), there were no apparent heterogeneous capture probabilities in our data as indicated by a straight line (Figure 14). Due to lack of heterogeneity, no heterogeneity structures were included in the log-linear models.

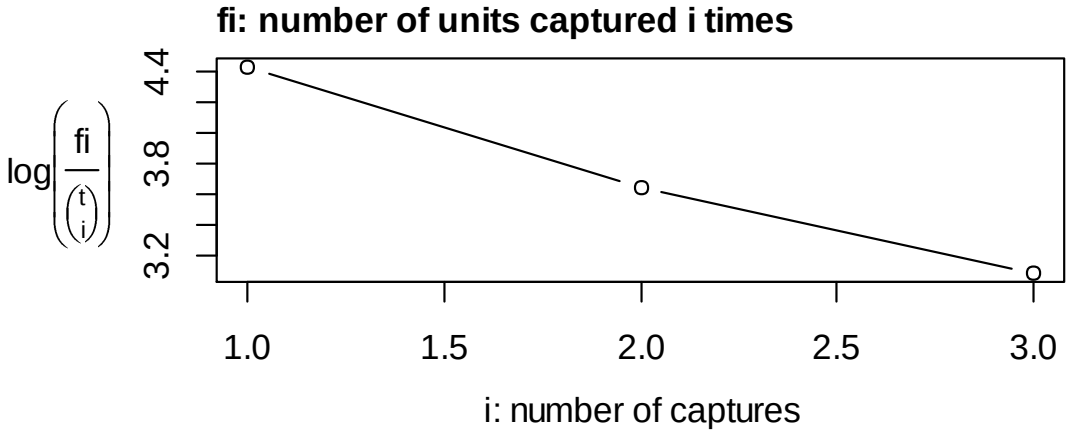


Figure 14: Exploratory heterogeneity graph of the capture histories by source (2010 - 2013).

Based on the Pearson residuals of the fitted models, ignoring models with behaviour characteristics ("M_b" and "M_{bh}"); since these are considered inappropriate for epidemiological applications, models "M_t" and "M_{th}" have the tightest residuals. Due to homogeneity in capture probabilities as depicted above and an insufficient number of data sources ($n<5$), the model "M_t" is preferred over "M_{th}". However, the distribution of residuals in model "M_t" is not centred around zero. Boxplots of the Pearson residuals of the ecological models are presented in Figure 15 below.

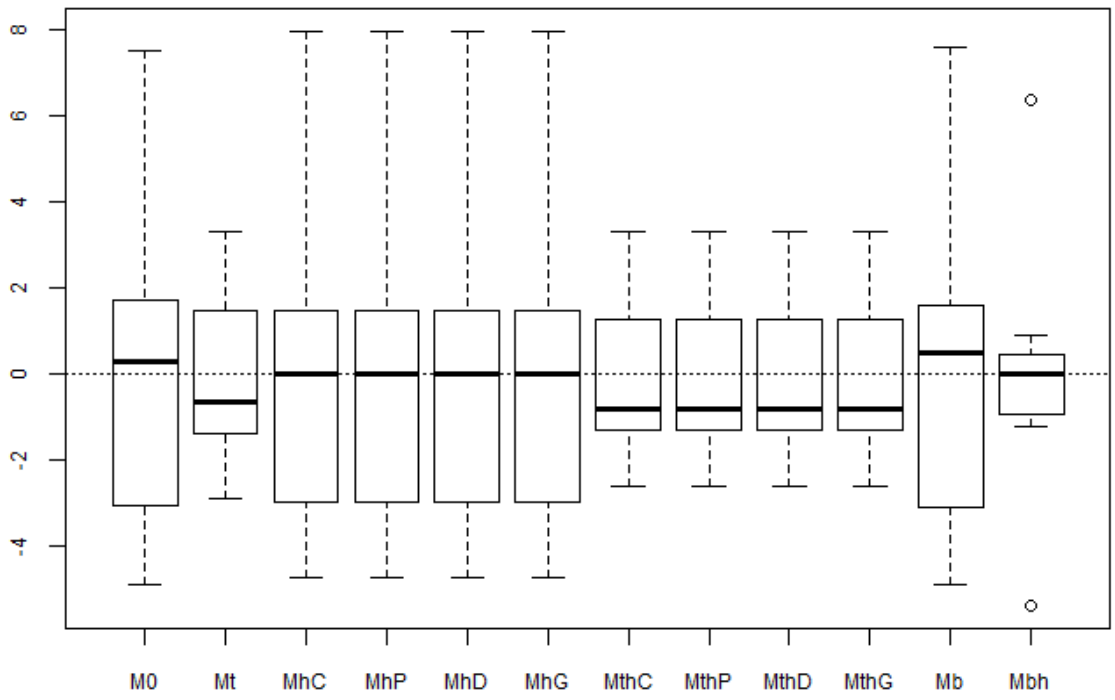


Figure 15: Boxplots of Pearson residuals of the model fit for the complete dataset (2010 – 2013).

Table 17 shows estimates from ecological models assuming various capture probabilities. Based on AIC, BIC and deviance statistics, model "M_t" gives the best fit to the data. This model estimated number of maternal deaths at 542 (95% CI: 500, 595). The results obtained from a model giving best fit have been presented and compared with results from other estimation approaches in table 20.

Table 17: Closed population capture-recapture ecological models fitted with R (2010 – 2013).

Model	Abundance	Std Error	Deviance	DF	AIC	BIC
M_0	563.82	26.33	107.92	5	150.74	158.68
M_t	541.97	24.13	30.01	3	76.84	92.71
M_h Chao (LB)	574.93	31.79	107.37	4	152.20	164.10
M_h Poisson2	595.90	56.27	107.37	4	152.20	164.10
M_h Darroch	619.25	92.15	107.37	4	152.20	164.10

M_h Gamma3.5	645.62	138.18	107.37	4	152.20	164.10
M_{th} Chao (LB)	559.33	29.49	28.45	2	77.28	97.12
M_{th} Poisson2	593.60	55.71	28.45	2	77.28	97.12
M_{th} Darroch	634.85	98.59	28.45	2	77.28	97.12
M_{th} Gamma3.5	685.20	160.13	28.45	2	77.28	97.12
M_b	577.34	55.48	107.83	4	152.65	164.56
M_{bh}	19.72	112.40	77.91	3	124.74	140.61

Log-linear models

Table 18 below shows estimates from log-linear models assuming various capture probabilities. The log-linear model assuming independence ("M_{000}") estimated the number of maternal deaths at 542 (95% CI: 500, 595). This result is equivalent to that of ecological model ("M_t"). Nevertheless, this was expected since model "M_t" does not account for any dependence and the two approaches fit the same model. The model controlling for dependence between PHDC and CEMD data sources fitted the data best, estimating a total of 719 (95% CI: 605, 879) maternal deaths to have occurred between 2010 and 2013 in the WC province. Estimates obtained with the alternative package i.e., Stata™ were almost identical (Appendix 3).

Table 18: Closed population capture-recapture log-linear models (2010 - 2013).

	Estimated maternal deaths		**Goodness-of-fit statistics**			
Model	$\widehat{N}$	**95% CI**	**Deviance**	**DF**	**AIC**	***p* value**
Independent	542	500, 595	30.01	3	76.84	<0.001
$M_{23/1}$	520	480, 569	24.30	4	73.12	<0.001
$M_{13/2}$	719	605, 879	4.55	4	53.38	0.985
$M_{12/3}$	525	483, 578	27.64	4	76.46	<0.001

Key:

Sources: **S_1** – PHDC; **S_2** – DHA; **S_3** – CEMD

M_{000}: Independent model i.e., model with no interaction

$M_{23/1}$: Model with one-way interaction i.e., $S_2 * S_3$

$M_{13/2}$: Model with one-way interaction i.e., $S_1 * S_3$

$M_{12/3}$: Model with one-way interaction i.e., $S_1 * S_2$

Sample coverage approach

Table 19 below shows estimates from the sample coverage approach. The sample coverage approach, assuming that the sources are independent "$\widehat{N}_0$" estimated the number of maternal deaths at 580 (95% CI: 532, 645). The estimated number using this approach is similar to that from the corresponding ecological and log-linear models i.e., "Mt" and "M_{000}". However, this was expected since all three approaches assumes independence. Due to low overlapping information/estimated sample coverage ($\hat{C} = 52.8\%$), the one-step population size estimate ($\widehat{N}_1$) is recommended over the "$\widehat{N}_0$". Model $\widehat{N}_1$ estimated the lower bound of maternal deaths in the province at 662 (95% CI: 585, 769). The results from this model ($\widehat{N}_1$) indicated that there is a weak negative dependence between PHDC and DHA, moderate positive dependence between PHDC and CEMD, and a weak positive dependence between DHA and CEMD.

Table 19: Estimated maternal deaths using sample coverage approach (2010 - 2013).

Model	M	D	$\widehat{C}$	Est	SE	95% CI
$\widehat{N}_0$	391	306.667	0.528	580	29	532, 645
$\widehat{N}$	391	306.667	0.528	932	271	606,1757
$\widehat{N}_1$	391	306.667	0.528	662	46	585,769

Parameter estimates:

Model	u1	u2	u3	r12	r13	r23
$\widehat{N}_0$	0.35	0.19	0.41	-0.19	0.31	-0.07
$\widehat{N}$	0.22	0.12	0.26	0.31	1.1	0.49
$\widehat{N}_1$	0.31	0.16	0.36	-0.07	0.49	0.06

Key:

M: Number of individuals ascertained in at least one list;

D: The average of the number of individuals listed in the combination of any two lists omitting the other one;

$\hat{C}$: Sample coverage estimate;

Est: Population size estimate;

SE: Estimated standard error of the population size estimation based on bootstrap replications;

95% CI: 95% confidence interval for $\boldsymbol{Est}$;

$\widehat{N}$: Population size estimate for sufficiently high sample coverage cases;

$\widehat{N}_1$: One-step population size estimate for low sample coverage cases. This estimator is recommended for use when the estimated sample coverage (overlapping information) is too low such that $\widehat{N}$ becomes unstable i.e., the estimated SE of $\widehat{N}$ is relatively large.

u1, u2, u3: estimated mean probabilities depending on the estimate of N for PHDC, DHA and CEMD respectively.

r12, r13 and r23: estimated coefficient of covariation (CCV) depending on the estimate of N for PHDC & DHA, PHDC & CEMD and DHA & CEMD respectively.

Capture-recapture estimate of maternal deaths assuming independence

The models from the three approaches assuming independence among lists have as expected obtained similar estimates of maternal mortality over the full 4-year period with overlapping 95% Confidence Intervals. The sample coverage approach has estimated maternal deaths (n=580) slightly higher than that obtained from the ecological and log-linear models (n≈543 respectively). These results are depicted in table 20 below. Appendices 2 and 3 presents broadly the mortality estimates and completeness levels when assuming independence.

Table 20: Selected model results depicting estimated maternal mortality assuming independence between sources (2010 - 2013).

Model	Description	Estimate (95% CI)
M_t	Ecological model	542 (95% CI: 500, 595)
M_{000}	Log-linear model	543 (95% CI: 501, 597)
$\widehat{N}_0$	Sample coverage	580 (95% CI: 532, 645)

M_t, M_{000} and $\widehat{N}_0$ – Models assuming homogeneity of capture and independence

Based on results from the log linear model, Table 21 shows the distribution of maternal deaths according to data source and year of death when assuming independence among sources. A total of 542 maternal deaths were estimated to have occurred between 2010 and 2013 in the Western Cape province. Of them, 151 maternal deaths were not registered by any source during the same period of analysis. In 2012, more than half (83/108) of the maternal deaths reported by the three sources were not captured.

Table 21: Estimates and distribution of maternal deaths by source and year of death assuming independence between the sources.

Year of death	Registered deaths			Total unique captured	Not captured	$\widehat{N}$(95% CI)
	PHDC	DHA	CEMD			
2010	50	37	51	92	27	119 (105, 143)
2011	51	18	47	85	36	121 (102, 154)
2012	48	20	67	108	83	191 (153, 258)
2013	56	33	73	106	24	130 (117, 151)
2010-2013*	**205**	**108**	**238**	**391**	**151**	**542 (500, 595)**

**Values are not the summation of yearly estimates but rather independent estimate*

Capture-recapture estimates of maternal deaths accounting for source dependencies

Different estimation techniques (log-linear and sample coverage approach) accounting for pairwise dependence between sources have obtained comparable/similar estimates of maternal mortality with slightly wider confidence intervals. The log linear model accounting for dependence between PHDC and CEMD (i.e., S_1*S_3) has obtained 719 maternal deaths for the period 2010 to 2013. The sample coverage approach using the one-step population size estimate ($\widehat{N}_1$) has provided a lower bound of the maternal deaths consistent with estimates from the log linear model i.e., 662 (95% CI: 585, 769) versus 719 (95% CI: 605, 879) respectively. These results are depicted in Table 22 below.

Table 22: Selected model results depicting estimated maternal deaths accounting for source dependencies (2010 - 2013).

Model	Description	Estimate (95% CI)
$M_{13/2}$	Log-linear model	719 (95% CI: 605, 879)
$\widehat{N}_1$	Sample coverage	662 (95% CI: 585, 769)

*$M_{13/2}$: Model with one interaction (S_1*S_3) i.e., u12 = u23 = 0.*

Table 23 shows the distribution of maternal deaths according to data source and year of death after controlling for source dependencies (PHDC and CEMD). A total of 719 maternal deaths were estimated to have occurred between 2010 and 2013 in the Western Cape province. Of them, 328 maternal deaths were not registered by any source during the same analysis period. In 2012, almost twice the number of maternal deaths reported by the three sources were not captured.

Table 23: Estimates and distribution of maternal deaths by source and year of death controlling for source dependencies.

Year of death	Registered deaths			Total unique captured	Not captured	$\widehat{N}$*(95% CI)
	PHDC	DHA	CEMD			
2010	50	37	51	92	58	150 (117, 207)
2011	51	18	47	85	23	108 (94, 131)
2012	48	20	67	108	205	313 (185, 678)
2013	56	33	73	106	47	153 (124, 202)
2010-2013**	**205**	**108**	**238**	**391**	**328**	**719 (605, 879)**

**Estimates obtained from the best fitting model (model controlling for PHDC and CEMD dependencies).*
***Values are not the summation of yearly estimates but rather independent estimate*

Sensitivity analyses

Using the combined data (full-four years data), reducing the non-overlapping PHDC cases by 75%, 50% and 25% (assuming those are not pregnancy-related deaths) resulted in decreasing maternal mortality estimates by a lower margin when fitting an independent model than when controlling for pairwise dependencies. Table 24 below shows these results.

Table 24: Sensitivity analyses to validate the obtained estimate of maternal mortality.

SN	Sensitivity analyses	Estimated maternal deaths (95% CI)	
		Independence[¥]	Dependence[*]
1	Including 25% of PHDC non-overlaps	410 (383, 446)	581 (492, 712)
2	Including 50% of PHDC non-overlaps	453 (421, 495)	626 (530, 769)
3	Including 75% of PHDC non-overlaps	499 (461, 547)	673 (569, 828)
4	All 391 cases	542 (500, 595)	719 (607, 885)

[¥] Assuming independence between sources; [*]Accounting for dependencies between sources.

Completeness of maternal death registrations according to sources

Table 25 presents estimates of completeness of maternal deaths registration by source, assuming independence between sources. Overall completeness of maternal deaths registration by the three sources was 72.14%. Higher level completeness was observed in CEMD (43.91%) followed by PHDC (37.82%). Relative to other years, the completeness of deaths registrations was highest in 2013 and lowest in 2012 i.e., 81.54% and 56.54% respectively. In addition, CEMD and PHDC recorded the highest completeness of death registrations in 2013 (56.15%) and (43.08%) respectively; and DHA in 2010 (31.09%).

Table 25: Estimates of completeness in maternal deaths registration by source assuming independence.

Year of death	$\widehat{N}$	PHDC		DHA		CEMD		Total	
		Reg	%Comp	Reg	%Comp	Reg	%Comp	Reg	%Comp
2010	119	50	42.02	37	31.09	51	42.86	92	77.31
2011	121	51	42.15	18	14.88	47	38.84	85	70.25
2012	191	48	25.13	20	10.47	67	35.08	108	56.54
2013	130	56	43.08	33	25.38	73	56.15	106	81.54
2010-2013	**542**	**205**	**37.82**	**108**	**19.93**	**238**	**43.91**	**391**	**72.14**

***Reg** - Number of maternal deaths registered; **Comp** - Completeness level.*

Table 26 presents estimates of completeness of maternal deaths registration by source, accounting for pairwise dependencies. Overall completeness of maternal deaths registration by

the three sources was 54.38%. CEMD had a slightly higher level of completeness (33.10%) compared to PHDC (28.51%) and DHA (15.02%). Relative to other years, the completeness of deaths registrations was highest in 2011 and lowest in 2012 i.e., 78.7% and 34.5% respectively. In addition, CEMD recorded the highest completeness of death registrations in 2013 (47.7%); PHDC in 2011 (47.2%); and DHA in 2010 (24.7%).

Table 26: Estimates of completeness in maternal deaths registration by source accounting for pairwise dependencies.

Year of death¥	$\widehat{N}$	PHDC		DHA		CEMD		Total	
		Reg	%Comp	Reg	%Comp	Reg	%Comp	Reg	%Comp
2010	150	50	33.33	37	24.67	51	34.00	92	61.33
2011	108	51	47.22	18	16.67	47	43.52	85	78.70
2012	313	48	15.34	20	6.39	67	21.41	108	34.50
2013	153	56	36.60	33	21.57	73	47.71	106	69.28
2010-2013	**719**	**205**	**28.51**	**108**	**15.02**	**238**	**33.10**	**391**	**54.38**

Reg *- Number of maternal deaths registered;* ***Comp*** *- Completeness level.*
¥Yearly estimates were derived from the best fitting model (model controlling for PHDC and CEMD dependencies).

Adjustment factors for under-registration

Table 27 presents different adjustment/correction factors to correct for under-registration of maternal mortality in the WC province when considering various criteria. When assuming dependence, estimation of maternal mortality in the Western Cape province for the year 2010-2013 requires adjustment by 1.84. For CEMD data (the most commonly reported official data), the adjustment factor varied from 1.64 when assuming no missing ascertainment, only counting deaths reported on one system, to 2.63 when assuming pairwise dependences.

Table 27: Adjustment factors to correct for under-registration in the WC province (2010-2013)

SN	Criteria	CEMD		All 3 sources	
		$\widehat{N}/n$	Adjustment factor	$\widehat{N}/n$	Adjustment factor
1	Assuming no missing ascertainment	*391/238*	1.64	*391/391*	1.0
2	Assuming independence	*542/238*	2.28	*542/391*	1.39
3	Assuming independence and including 50% of PHDC non-overlaps*	*453/238*	1.90	*453/391*	1.16
4	Assuming dependence	*719/238*	3.02	*719/391*	1.84
5	Assuming dependence and including 50% of PHDC non-overlaps*	*626/238*	2.63	*626/391*	1.60

**Assuming 50% of non-overlapping cases in PHDC are not pregnancy-related deaths*

5.5 Discussion

This study has used capture-recapture, a novel estimation technique in the setting, to ascertain maternal mortality in the Western Cape province, South Africa. Estimation of maternal mortality under-reporting involved statistically operationalising the overlapping information from the three data sources rather than simply matching, deduplication and adding the registered cases from each source. In this analysis, three classes of capture-recapture models (log-linear, ecological and sample coverage approach) were applied in three distinct data sources registering maternal deaths in the province. The results of capture-recapture analyses indicated the burden of maternal mortality in the province could be substantially larger than previously estimated by both institutional reporting and in global metrics[41–50].

Overlap between sources

According to the observed data, there was a considerable lack of overlap between the three data sources (6%) which suggests maternal mortality under-reporting in the province. Estimates from

the model with a time (list) effect and pairwise dependencies between PHDC and CEMD indicated only 54.3% of the estimated maternal deaths were identified by the three sources combined, with completeness levels of the two sources (DHA and PHDC) below 30%. The possible reason for a slightly higher completeness level in the CEMD over the other two data sources lies with its active notification process and efficient reporting mechanism allowing prompt identification of maternal deaths[14]. The majority of maternal deaths are expected in the public sector. DHA was expected to possibly provide better completeness than the CEMD and PHDC since the latter are facility-based, and do not necessarily capture deaths at community level or in private facilities, therefore being potentially more prone to maternal death under-reporting.

It is possible that in DHA, pregnancy may not be mentioned due to poor coding of cause of death or for non-direct causes. Failure to record pregnancy could potentially explain the lack of capture by the CEMD of the same deaths. Additionally, DHA may also include deaths from the private health care facilities which lead to a lack of overlap with CEMD as well as PHDC, even though all maternal deaths are supposed to be reported to the health department. Despite a higher completeness level, CEMD might miss indirect maternal deaths especially in early pregnancy. On the other hand, due to reliance on fact rather than cause of death, PHDC is prone to potentially list deaths, in which pregnancy is incidental.

Performance of different capture-recapture analytical approaches

Based on goodness of fit, statistics and the heterogeneity exploratory graph, the log-linear model with time (list) effect and pairwise dependencies between PHDC and CEMD gave the best fit. In this analysis, dependence between PHDC and CEMD was expected, since these data sources

register facility-based maternal deaths predominantly. The enquiry and notification process of maternal deaths in CEMD usually begins from the respective facility, which in one way allows the facility to feed the death information in the hospital electronic-registration system which is directly captured by the PHDC. As one of the assumptions of the classical capture-recapture method[21,51,52], failing to control for dependence between data sources is associated with obtaining a biased population estimate[21]. Positive dependence is expected to result in underestimation and estimates from the models which assumed that independence did in fact result in lower estimates. However, in multiple estimations, capture-recapture methods are known to handle the pairwise dependencies between lists (data sources) analytically[20,28–36]. Due to interdependencies between PHDC and CEMD (moderate positive dependence), the application of log-linear methods provided the most robust and unbiased estimate of maternal mortality in the province (N=719; 95% CI: 605, 879). The existence of a moderate positive dependence between PHDC and CEMD highlights the importance of fitting more than two data sources to avoid potential underestimation of maternal deaths when the respective sources are used.

Over decades several packages for capture-recapture analysis have been written, accommodating epidemiological applications as well as allowing both closed and open population models[53]. In this analysis we have used various packages from three key statistical software (Stata, R and S plus) for comparison purposes, all potentially accessible to applied epidemiologists[31,40,54]. The sample coverage approach assuming independence has provided slightly higher estimates of maternal mortality compared to the ecological and log-linear models.

In our data, the estimated sample coverage (overlapping information) was as low as 53%.thereby accounting for a positive dependence between PHDC and CEMD as explained above, the estimate ('one-step' estimator) of maternal mortality obtained from the sample coverage approach should be treated as a lower bound. However, the lower bound estimate assuming dependence was slightly higher than the lower bound of the confidence interval from the log-linear models when assuming dependence. Broadly, the outputs are aligned.

Feasibility of using CR for estimation of maternal mortality in the South African setting

Due to divergent estimates of maternal mortality in South Africa and the Western Cape province in particular, our study demonstrates that; reliable local and recent estimates of these outcomes can be obtained relatively quickly and at a low-cost via the applications of capture-recapture methodologies. A considerable number of studies have revealed that, conducting a capture-recapture analysis is a relatively cheap, easy to use and appropriate estimation method, allowing self-calibrations using two or more available data sources to retrospectively identify cases when estimating conditions with diverse estimates[17,22,33,37]. Lack of accurate estimates of maternal mortality in the province and the country in general is expected to affect assessment and evaluation of programmes to reduce maternal mortality, monitoring progress in global goals and understanding trends and local needs in the population.

Although the capture-recapture method requires sufficiently high overlapping information in order to model dependence between sources and to provide reliable estimates of population size, ignoring underreporting gives biased population estimates. Consequently, this study has

proposed a plausible range for maternal mortality estimates in the Western Cape province. In this study, the estimated maternal mortality estimates are much higher than current reporting and the face validity of such a dramatic estimate of under-reporting might be of concern. For this reason, the study explored the extent to which misclassification could have inflated this estimate.

Corrections to officially reported maternal mortality in the Western Cape

The current study estimates that ascertainment of maternal deaths, assuming no misclassification and combining all sources, in the Western Cape between 2010-2013, was 54%. These data suggest that, estimation of maternal mortality in South Africa, when combining data from the same sources as used in the Western Cape, could require adjustment by a factor of 1.84 to account for under-registration of this outcome, and higher when adjusting the CEMD mortality rate.

Despite the lack of a published adjustment factor for recent data, both global agencies and in-country institutional reporting have used correction factors derived from local specialised studies[42,48,55]. For example, adjustment factors used to account for misclassification of maternal deaths in civil registration and vital statistics between the period 1999 and 2001 were 0.98 and 1.16 between 2002 and 2004[42,55]. These correction factors are much lower than the one estimated in the current study. The reasons for the varying ascertainment of maternal mortality are well explained in Chapter 3 (systematic review). However, this will be discussed in greater detail in the synthesis.

Strengths and limitations

The strengths of this study include application of various statistical packages, all potentially accessible to applied epidemiologists, for comparison purposes. The most important limitation lies in the fact that this analysis used secondary data, without a clinical review/audit of accuracy. This might have resulted in an inconsistent number of maternal deaths i.e., in 2012. In addition, existence of potential incidental overlaps of pregnancy status (misclassification) mainly in CEMD data.

5.6 Conclusion

The application of capture-recapture techniques show promise as an approach to resolve problems of estimating maternal mortality in the Western Cape province and similar settings. The current study has demonstrated that substantial adjustments might need to be made to officially reported maternal mortality estimates. The estimates of the extent of this adjustment varied widely based on the modelled dependence structure, and when allowance was made for potential misclassification of cases in the most likely source in which there might be misclassification. The variability also compromised the ability to explore trends in the four-year period under review. Although it was beyond the scope of the secondary data analysis, there is a need for clinical review of the source accuracy of each of the data sources which would help to refine any future attempts at implementing this method.

5.7 References

1. WHO, UNICEF, UNFPA, The World Bank & The United Nations Population Division. *Trends in Maternal Mortality: 1990-2013. Estimates by WHO,UNICEF, UNFPA, The World Bank and the United Nations Population Division*. *WHO, UNICEF, UNFPA, The World Bank The United Nations Population Division* (2014). doi:WHO/RHR/14.13

2. Merdad, L., Hill, K. & Graham, W. Improving the Measurement of Maternal Mortality: The Sisterhood Method Revisited. *PLoS One* **8**, (2013).

3. Graham, W. J. Now or never: The case for measuring maternal mortality. *Lancet* **359**, 701–704 (2002).

4. Berg, C. J., Danel, I. & Mora, G. Guidelines for Maternal Mortality Epidemiological Surveillance. *Pan Am. Heal. Organ.* 50 (1996).

5. UNICEF, WHO, World Bank group & United Nations. *Levels & Trends in Child Mortality - Report 2015: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. *UNICEF* (2015). doi:10.1371/journal.pone.0144443

6. WHO, UNICEF, UNFPA, World Bank & United Nations Population Division. Trends in maternal mortality 1990 to 2015. (2015). doi:10

7. Moodley, J. *Saving Mothers 2011-2013: Sixth report on confidential enquiries into maternal deaths in South Africa, Short report*. *Pretoria: Department of Health* (2014).

8. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384**, 980–1004 (2014).

9. Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 957–979 (2014).

10. Saleem, S. *et al.* A prospective study of maternal, fetal and neonatal deaths in low- and middle-income countries. TT -. *Bull. World Health Organ.* **92**, 605–612 (2014).

11. Dorrington, R. E., Bradshaw, D., Dorrington, R. E. & Bradshaw, D. Acknowledging uncertainty about maternal mortality estimates Maternal mortality estimates. *Bull World Heal. Organ* **94**, 155–156 (2016).

12. Bradshaw, D. *et al.* Every death counts: use of mortality audit data for decision making to save the lives of mothers, babies, and children in South Africa. *Lancet* **371**, 1294–1304 (2008).

13. Chopra, M., Daviaud, E., Pattinson, R., Fonn, S. & Lawn, J. E. Saving the lives of South Africa's mothers, babies, and children: can the health system deliver? *Lancet* **374**, 835–846 (2009).

14. Moodley, J. *et al.* The confidential enquiry into maternal deaths in South Africa: a case study. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 53–60 (2014).

15. Souza, J. P. *et al.* Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *The Lancet* **381**, 1747–1755 (2013).

16. UNICEF, WHO, World Bank, U. *Levels & Trends in Child Mortality in 2014- Report: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. (2014).

17. Forecasting, I. W. G. for D. M. and. Capture-recapture and multiple-record systems estimation I: History and theoretical development. International Working Group for Disease Monitoring and Forecasting. *Am. J. Epidemiol.* **142**, 1047–58 (1995).

18. Wittes, J. & Sidel, V. W. A generalization of the simple capture-recapture model with applications to epidemiological research. *J. Chronic Dis.* **21**, 287–301 (1968).

19. Sekar, C. C. & Deming, W. E. On a Method of Estimating Birth and Death Rates and the Extent of Registration (Excerpt). *Am. Stat.* **58**, 13–15 (2004).

20. Fienberg, S. E. The Multiple Recapture Census for Closed Populations and Incomplete 2 k Contingency Tables. *Biometrika* **59**, 591 (1972).

21. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20**, 3123–3157 (2001).

22. Van Hest, R., Grant, A. & Abubakar, I. Quality assessment of capture-recapture studies in resource-limited countries. *Tropical Medicine and International Health* **16**, 1019–1041 (2011).

23. Yazdizadeh, B. *et al.* Application of capture-recapture for fine-tuning uncertainties about national maternal mortality estimates. *Int. J. Prev. Med.* **5**, 624–631 (2014).

24. Checchi, F., Nyasulu, P., Chandramohan, D. & Roberts, B. Rates and causes of death in Chiradzulu District, Malawi, 2008: A key informant study. *Trop. Med. Int. Heal.* **16**, 375–378 (2011).

25. Qomariyah, S. N. *et al.* An option for measuring maternal mortality in developing countries: a survey using community informants. *BMC Pregnancy Childbirth* **10**, 74 (2010).

26. Domingo-Salvany, a *et al.* Analytical considerations in the use of capture-recapture to estimate prevalence: case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *Am. J. Epidemiol.* **148**, 732–740 (1998).

27. McCarty, D. J., Tull, E. S., Moy, C. S., Kwoh, C. K. & LaPorte, R. E. Ascertainment corrected rates: applications of capture-recapture methods. *Int. J. Epidemiol.* **22**, 559–65 (1993).

28. Oosterlee, A., Vink, R. M. & Smit, F. Prevalence of family violence in adults and children: estimates using the capture-recapture method. *Eur. J. Public Health* **19**, 586–91 (2009).

29. Nance, M. L., Carr, B. G., Kallan, M. J., Branas, C. C. & Wiebe, D. J. Variation in pediatric and adolescent firearm mortality rates in rural and urban US counties. *Pediatrics* **125**, 1112–8 (2010).

30. Lum, K., Price, M., Guberek, T. & Ball, P. Measuring Elusive Populations with Bayesian Model Averaging for Multiple Systems Estimation: A Case Study on Lethal Violations in Casanare, 1998-2007. *Stat. Polit. Policy* **1**, (2010).

31. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20**, 3123–57 (2001).

32. Cormack, R. M. Log-Linear Models for Capture-Recapture. *Biometrics* **45**, 395 (1989).

33. Yip, P. *et al.* Capture-recapture and multiple-record systems estimation I: History and theoretical development. (1995).

34. Cormack, R. . The statistics of capture-recapture methods. *Ocean. Mar Biol Ann Rev.* 455–506 (1968).

35. Coull, B. A. & Agresti, A. The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies. *Biometrics* **55**, 294–301 (1999).

36. Agresti, A. Simple Capture-Recapture Models Permitting Unequal Catchability and Variable Sampling Effort. *Biometrics* **50**, 494 (1994).

37. Hook, E. B. & Regal, R. R. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* **17**, 243–264 (1995).

38. Statistics South Africa. *Statistical release P0302: Mid - year population estimates 2017*. (2017).

39. Weterrn Cape Provincial Government. *Annual report 2016 / 2017*. (2017).

40. Heiden, M. an der. RECAP: Stata module to perform capture-recapture analysis for three sources with Goodness-of-Fit based confidence intervals. *Stat. Softw. Components* (2009).

41. Dorrington, R., Bradshaw, D., Laubscher, R. & Nannan, N. *Rapid Mortality Surveillance Report 2015*. (2016). doi:ISBN: 978-1-920618-00-1

42. WHO. *Trends in Maternal Mortality : 1990 to 2015 Estimates by WHO,UNICEF,UNFPA,The World Bank and the United Nations Population Division*. *Organization* (2015). doi:ISBN 978 92 4 156514 1

43. Dorrington, R., Bradshaw, D., Laubscher, R. & Nannan, N. *Rapid Mortality Surveillance Report 2014*. (2015).

44. Statistics South Africa. *Millennium Development Goals 5: Improve maternal health 2015 / Statistics South Africa*. (2015).

45. World Health Organization. Trends in Maternal Mortality : 1990 to 2013 Executive Summary. (2014).

46. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, (2014).

47. Department of Health. *2014 Saving Mothers: Annual report and detailed analysis of maternal deaths due to non-pregnancy related infections*. (2015).

48. National Committee for Confidential Enquiries into Maternal Deaths. *Saving Mothers 2011-2013: Sixth report on confidential enquiries into maternal deaths in South Africa Short report*. (2014).

49. World Health Organization. *World health statisitics 2011*. *World health statistics* (2011). doi:10.1002/(SICI)1096-987X(199802)19:3<259::AID-JCC1>3.0.CO;2-S

50. Lozano, R. *et al.* Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: An updated systematic analysis. *Lancet* **378**, 1139–1165 (2011).

51. Rossi, G., Pepe, P., Curzio, O. & March, M. Generalized Linear Models and Capture-Recapture Method in a Closed Population: Strengths and Weaknesses.

52. VALID International Ltd. *Notes on using capture-recapt ure techniques to assess the sensitivity of rapid case-finding methods*. (2006).

53. Chao, A. Capture-Recapture for Human Populations. *Wiley StatsRef Stat. Ref. Online* 1–16 (2015). doi:10.1002/9781118445112.stat04855.pub2

54. Baillargeon, S. & Rivest, L.-P. Rcapture : Loglinear Models for Capture-Recapture. *J. Stat. Softw.* **19**, 1–31 (2007).

55. National Committee for Confidential Enquiry into Maternal Deaths. *Saving Mothers 2008-2010: Fifth Comprehensive Report on Confidential Enquiries into Maternal Deaths in South Africa*. *Department of Health* **5**, (2012).

CHAPTER 6: Estimation of neonatal deaths under-reporting using the capture-recapture method in the Western Cape Province, South Africa

6.1 Abstract

Background: Tracking progress towards Sustainable Development Goal 3.2 requires valid and reliable estimate of child mortality in the country. Due to varying estimates of child mortality reported by different sources in the country and in the Western Cape (WC) province, there is value in applying additional estimation techniques, particularly capture-recapture methods, which utilise multiple data sources to correct for under-reporting of this outcome.

Methods: A three-source capture recapture analysis was conducted using mortality data from the Department of Home Affairs (DHA), institutional reporting (PPIP) and hospital death-reporting system (PHDC) from 2010 to 2013. The provincial data linkage algorithm was used to match records across sources for potential overlaps. Log-linear and sample coverage models were used to determine the unbiased estimates of neonatal deaths accounting for pairwise dependence. Testing the goodness-of-fit of the log-linear models involved the use of likelihood ratio statistics (G^2), the Pearson goodness-of-fit statistics (χ^2), the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The estimates of maternal deaths with the corresponding 95% Confidence Intervals (95%CI) were presented yearly and combined from 2010 to 2013.

Results: After record linkage, the three sources identified 8916 unique neonatal deaths from 2010 to 2013. Of these, 4,155 (46.6%) were registered by all three sources. Capture-recapture methods estimated 10,697 (95% Confidence Interval (CI): 10,508, 10,909) neonatal deaths between 2010 and 2013. By calendar year, neonatal deaths were estimated at 3,045 (95% CI:

2914, 3210) in 2010; 2,515 (95% CI: 2426, 2625) in 2011; 2,649 (95% CI: 2579, 2734) in 2012; and 2,522 (95% CI: 2452, 2615) in 2013. The completeness of registrations of the three sources combined was 84.0%, with PHDC having highest completeness (68.8%).

Conclusion: Neonatal deaths were moderately under-reported in the Western Cape Province in 2010 - 2013. Ascertainment of neonatal mortality using a single data source provide biased estimates of this outcome. Relying on crude neonatal mortality estimates could subsequently bias the evaluations of interventions focusing on reducing neonatal mortality, assessing progress against global commitments and evidence-based decision making.

Keywords: Capture-recapture, estimate, neonatal deaths, Western Cape, South Africa

6.2 Introduction

Neonatal mortality is a key health outcome that requires accurate documentation and ascertainment for clinical and public health value. Tracking progress towards population health targets, particularly the Sustainable Development Goal (SDG) 3.2[1], requires valid, reliable and internationally comparable estimates of neonatal mortality. Additionally, evaluating impacts of health interventions and the effective planning of health services requires accurate mortality estimates in the country. Globally, neonates have consistently been reported to account for an increasing proportion of under-five deaths despite the decreasing trends in neonatal mortality[2,3].

Global agencies and in-country institutional reporting, as illustrated in chapter 3, indicate South Africa has varying sub-national and national estimates of neonatal mortality. The country and particularly the Western Cape (WC) province is unique in the region in that it has multiple data sources registering maternal and new-born outcomes of institutional deliveries. Estimation of neonatal mortality in the country is primarily based on the vital registrations as well as the Perinatal Problem Identification Program (PPIP) which record stillbirths and neonatal deaths[4,5].

In spite of the availability of multiple data sources; ascertainment of new-born deaths is generally difficult owing to insufficient clinical documentation, poor vital registration, varying case definitions and absent/improper record linkage[6,7]. Official mortality reporting systems still do not capture all deaths despite the significant investments and improvements in neonatal vital registration, and the relative importance of mortality statistics in clinical management, healthcare administrations and epidemiologic analyses[8].

Although conducted 5 years prior to the study period, a global assessment of mortality data ranked South Africa as having 'low' quality death records. According to the findings, the country has remarkably high reporting completeness (88%) with low coverage levels (<50%) compared to other developing nations[8]. In this context, estimation of neonatal mortality without correction of underreporting of deaths can potentially bias the demographic estimates of the studied population, thereby affecting meaningful interpretation of the available data[9]. Additionally, crude estimates of neonatal mortality can be misleading if used for monitoring trends, evaluation purposes, health planning and evidence-based decision making.

Due to varying and limited estimates of neonatal mortality reported by different sources in the country and in the Western Cape (WC) province[10], there is value in applying additional estimation techniques which utilise multiple data sources to correct for underreporting of this outcome. Capture-recapture methods are known to resolve uncertainties in estimating conditions that have diverse estimates[4,5,11–18]. The background to capture-recapture methods has been broadly discussed in Chapters two and five.

Capture-recapture methods provide an additional tool for correcting and quantifying the under-reporting of neonatal deaths by utilising multiple data sources available in the respective area. These methods have been proven to be simple, quick and cheap in ascertaining unbiased estimates of conditions/populations known to have diverse estimates[19–21]. There has been limited application of capture-recapture models to estimate neonatal mortality underreporting in the Western Cape Province, South Africa.

The main purpose of this chapter is to estimate neonatal mortality under-reporting using the capture-recapture methods and assess the level of completeness of data sources registering neonatal deaths in the Western Cape province, South Africa.

6.3 Methods

Design and Setting

This analysis used routinely-collected cross-sectional clinical and administrative data from the Western Cape Department of Health (WCDOH) and the South African Department of Home Affairs (DHA) to estimate the underlying burden of neonatal mortality in the province.

Case definition and analysis period

Neonatal death has been defined as the death of a baby within the first 28 days of life. In this chapter, we restricted the analyses to neonatal deaths which occurred between 1st January 2010, though 31st December 2013, due to data availability and consistency in reporting mortality data across sources in the province[22]. Since the study period concluded long before this analysis, there is reasonable assurance of completeness of death registration across sources.

Data sources

As described in chapter 2, data from DoH and DHA were used. For this analysis, data from hospital deaths from 2010 to 2013, indicating a child within the first 28 days of life were extracted from the PHDC SQL database and served as the first data source for the capture-recapture modelling (S_1=PHDC). Neonatal deaths from PPIP from the same period, served as a second data source for neonatal mortality ascertainment and estimation (S_2=PPIP). Due to certain limitations on DHA

data (death certificates) as highlighted in chapter 2, we restricted our analysis to neonatal deaths occurred from 2010 to 2013 and formed a database which served as a third source for capture-recapture analyses (S_3=DHA).

Data processing and analysis

Data were extracted from the respective sources (PHDC, PPIP and DHA), cleaned and standardised using Stata software version 14.2 (Stata Corporation, College Station, Texas, USA). Using the provincial linkage approach (as presented in chapter 4), data from the PHDC were linked to data from the DHA and PPIP to identify overlaps[23]. Multiple-record-per-person data were converted to single record data with capture histories reflected by indicator variables. Descriptive and preliminary capture-recapture analyses assuming independence and homogeneity of capture probabilities were carried out using a Stata user-written programme "recap"[24] (Stata software version 15.1, Stata Corporation, College Station, Texas, USA), the R package "Rcapture"[25] (R software, version 3.5.0, R Foundation for Statistical Computing, Vienna, Austria) and CARE1 package.

In addition, the three-source capture-recapture modelling accounting for dependence and pairwise interactions were conducted using the above three software implementations for comparison purposes. An exploratory heterogeneity graph (fi plot) and boxplots of the Pearson residuals were used to graphically explore capture histories.

Capture-recapture analysis

In this analysis, three-source capture-recapture models (three classes of models) were fitted to estimate neonatal mortality underreporting and registration completeness. All underlying assumptions for applying capture-recapture methods were satisfactorily met, except for the assumption of independence. Due to the nature of these data (public facility-based), we anticipated positive dependence between PHDC and PPIP, i.e., the probability of a case being registered in PHDC and PPIP being high, since both sources register deaths from the public-sector facilities.

Neonatal mortality estimation

Estimations of neonatal deaths were conducted yearly and combined following the applications of the sample coverage approaches, ecological and log-linear capture-recapture models accounting for pairwise dependencies. Estimates were presented as neonatal deaths with their corresponding 95% Confidence Intervals (CIs). Estimates obtained from the different model classes were compared for the two scenarios i.e., dependence and independence assumption.

Estimation of completeness

Completeness of a source was obtained by dividing the number of neonatal deaths registered by the respective source to the total number of deaths estimated by the capture-recapture methods. The overall completeness level was obtained by dividing the number of unique neonatal deaths registered by the three sources combined, by the estimated neonatal deaths obtained from the capture-recapture analysis.

Sensitivity analyses

Unlike the maternal analysis, where there was uncertainty regarding the pregnancy and the pregnancy-associatedness of deaths, in this case we are certain that all deaths meet the case definition, and therefore there is no need for a sensitivity analysis in which the proportion of non-overlapping deaths is varied.

6.4 Results

Neonatal deaths registration by data sources and calendar years

In total, 14,938 neonatal deaths were registered by the three sources between 1st January 2010 and 31st December 2013. Of them, 7,362 (49.3%) neonatal deaths were registered by the hospital death reporting system (PHDC), 5,652 (37.8%) by the death certificates with reported causes of death (DHA) and 1,924 (12.9%) by the Perinatal Problem Identification Program (PPIP). Figure 16 below depicts the trends of neonatal deaths according to data sources in the Western Cape province, South Africa.

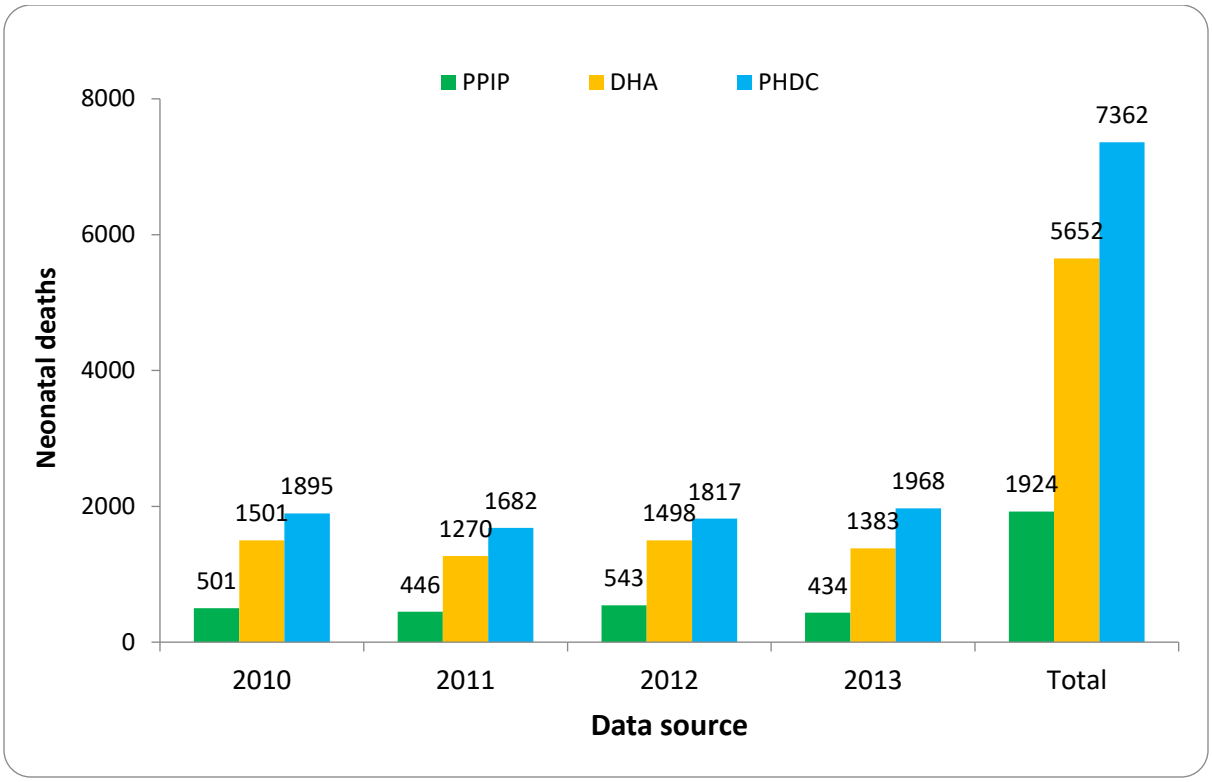


Figure 16: Distribution of neonatal deaths by year of registration and data sources

From 1st January 2010 to 31st December 2013, 1,097 out of 14,938 neonatal deaths were registered by the three sources combined. A total of 3,041 deaths were registered by PHDC and DHA, 674 by PPIP and PHDC and only 48 deaths by PPIP and DHA. A total of 8981 unique neonatal deaths were registered by the three sources and 4,121 out of 8981 deaths (45.8%) were registered by only a single source i.e., 2550 deaths by PHDC, 105 by PPIP and 1466 by DHA. Figure 17 below (proportional Venn diagram) depicts the overlaps between sources.

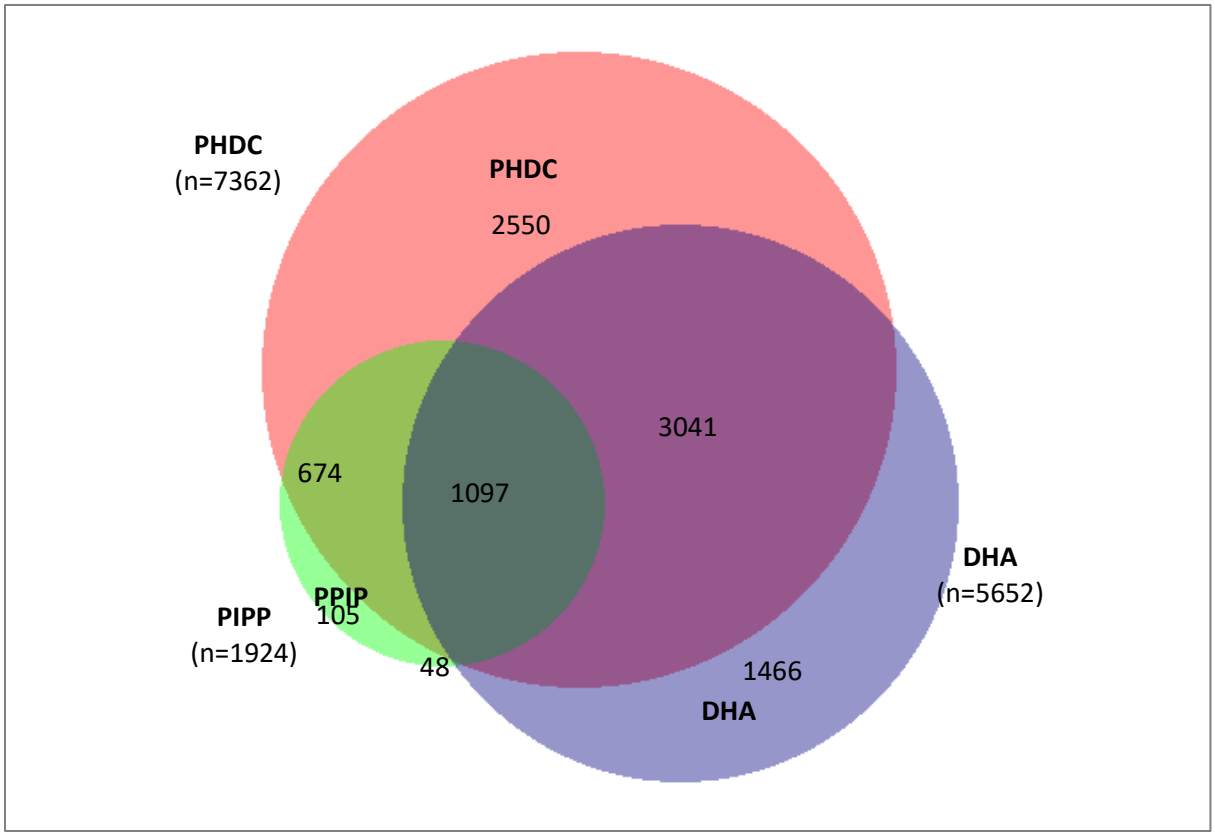


Figure 17: The proportional Venn diagram presenting overlaps between sources (2010-2013).

Ecological models

Figure 18 presents an exploratory heterogeneity graph of the capture probabilities. The exploratory heterogeneity graph (fi plot) depicts a slight heterogeneous capture probability in our data as indicated by a slightly non-linear trend. However, we could not fit heterogenous models using the ecological approach due to an insufficient (less than 5) number of data sources available as recommended. Details of the closed-population ecological models fitted to the combined data covering 2010 - 2013 are presented in Table 28 below.

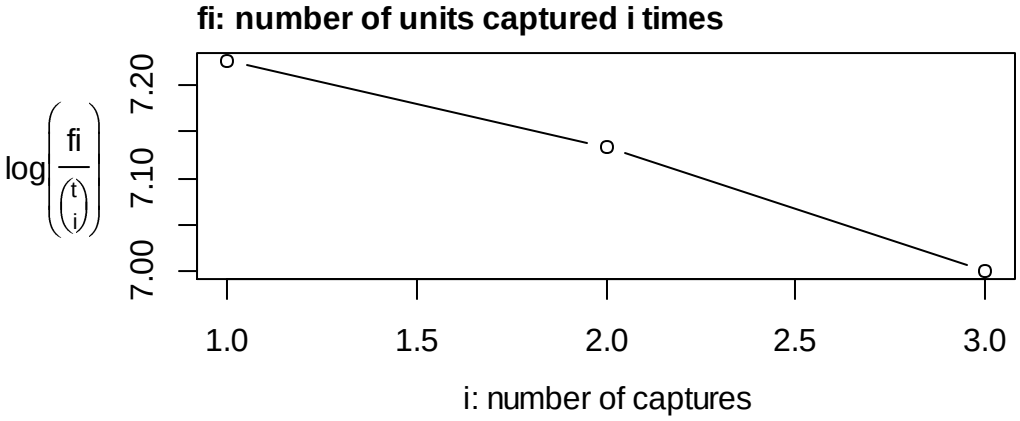


Figure 18: Exploratory heterogeneity graph of the capture histories by source

Figure 19 presents boxplots of the Pearson residuals for the different fitted ecological models. Ignoring models with behaviour characteristics ($M_b \ and \ M_{bh}$); since they are not relevant to this application, the Pearson residuals of the fitted models indicates that models "M_{th}" have the tightest residuals. As presented in Figure 18 above, the independent model (M_t) and model accounting for heterogeneous capture probabilities (M_{th}) are preferred over others. However, heterogeneous ecological models are only recommended, when the number of list/data sources are sufficiently large (at least five), as indicated in Chapter 2.

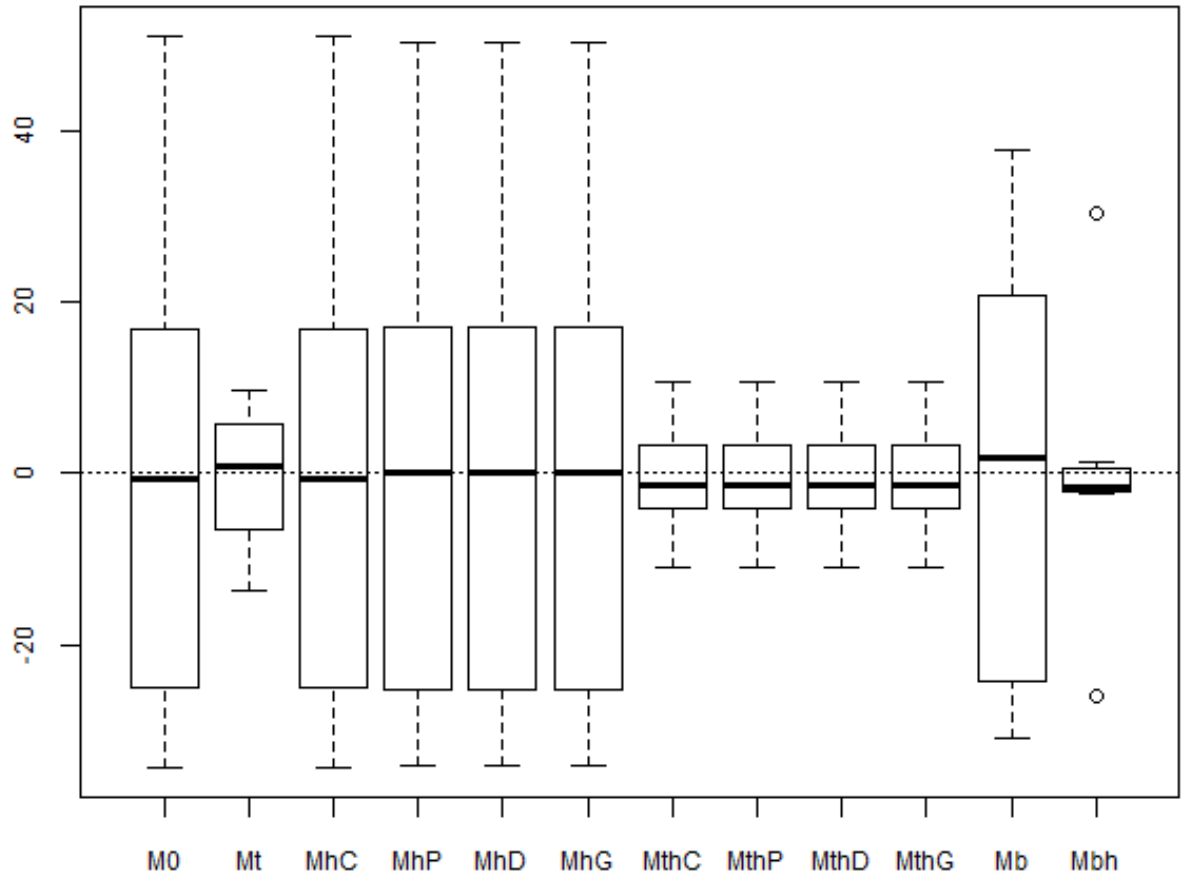


Figure 19: Boxplots of the Pearson residuals for the different fitted models

Table 28 below shows estimates from the CRC models controlling up to three sources of variations in capture probabilities. Based on model fit statistics i.e., AIC, BIC and deviance statistics, model M_{th} gives the best fit to the data. Since heterogeneity models are not being recommended, when having less than 5 samples/data sources, the second-best model (M_t) is selected, even though the fit is sub-optimal. This model estimated the abundance of neonatal deaths at 9,820 (95% CI: 9742, 9899).

Table 28: Closed population capture-recapture models (2010 - 2013).

Model	Neonatal deaths	Std Error	Deviance	DF	AIC	BIC
M_0	10,516	57.74	7042.09	5	7104.15	7118.35
Mt	9,820	40.08	561.00	3	627.07	655.48
Mh Chao (LB)	10,516	57.74	7042.09	5	7104.15	7118.35
Mh Poisson2	10,453	86.93	7041.25	4	7105.31	7126.62
Mh Darroch	10,422	113.28	7041.25	4	7105.31	7126.62
Mh Gamma3.5	10,391	141.26	7041.25	4	7105.31	7126.62
Mth Chao (LB)	10,057	51.44	358.89	2	426.95	462.46
Mth Poisson2	10,554	95.26	358.89	2	426.95	462.46
Mth Darroch	11,282	186.82	358.89	2	426.95	462.46
Mth Gamma3.5	12,363	356.35	358.89	2	426.95	462.46
Mb	9,226	20.77	5511.09	4	5575.15	5596.46
Mbh	7,344	5.22	1655.16	3	1721.23	1749.64

Log-linear models

The log-linear model assuming independence ("M_{000}") estimated the number of neonatal deaths at 9,820 (95% CI: 9743, 9901). This is equivalent to that of ecological model "M_t"; since the two approaches fit the same model. Based on the goodness-of-fit statistics, the model controlling for dependence between PPIP and DHA (S_2*S_3) gave the best fit. Adjusting for dependence between PPIP and DHA, there was an estimated 10,146 (95% CI: 10,051, 10,248) neonatal deaths between 2010 and 2013. Table 29 below shows estimates from log-linear models, assuming various capture probabilities.

Table 29: Closed population capture-recapture log-linear models accounting for pairwise dependencies (2010 - 2013)

Model	Estimated neonatal deaths		Goodness-of-fit statistics			
	$\widehat{N}$	95% CI	Std error	Deviance	DF	AIC
Independent	9,820	9742, 9899	40.1	561.00	3	627.10
$M_{23/1}$	10,146	10051, 10248	53.16	69.00	4	137.10
$M_{13/2}$	9,386	9303, 9481	45.85	485.30	4	553.40
$M_{12/3}$	9,838	9762, 9919	41.88	557.00	4	625.10

Key:

Sources: **S_1** – PHDC; **S_2** – PPIP; **S_3** – DHA

M_{000}: Independent model i.e., model with no interaction

$M_{23/1}$: Model with one-way interaction i.e., $S_2 * S_3$

$M_{13/2}$: Model with one-way interaction i.e., $S_1 * S_3$

$M_{12/3}$: Model with one-way interaction i.e., $S_1 * S_2$

Sample coverage approach

Table 30 shows estimates from the sample coverage approach. Assuming that the sources are independent "$\widehat{N}_0$", the sample coverage approach estimated the number of neonatal deaths at 9,754 (95% CI: 9686, 9829). Due to high estimated sample coverage ($\hat{C} = 78\%$), the precise estimate of neonatal mortality is expected. In this case, the population size estimate for sufficiently high sample coverage $\widehat{N}$ is preferred. Using 1000 bootstrap replications, this model estimated the neonatal deaths in the province at 10,697 (95% CI: 10508, 10909). Furthermore, this model $(\widehat{N})$ indicated the existence of a moderate positive dependence between PHDC and PPIP (r12=0.34), and weak positive dependence between PHDC and DHA (r13=0.06) and PPIP and DHA (r23=0.13) respectively.

Table 30: Estimated neonatal deaths using sample coverage approach (2010 - 2013).

Model	M	D	$\widehat{C}$	Est	SE	95% CI
$\widehat{N}_0$	8,981	7607	0.78	9,754	37	9686, 9829
$\widehat{N}$	8,981	7607	0.78	10,697	102	10508, 10909
$\widehat{N}_1$	8,981	7607	0.78	10,211	60	10099, 10334

Parameter estimates:

Model	u1	u2	u3	r12	r13	r23
$\widehat{N}_0$	0.75	0.20	0.58	0.22	-0.03	0.03
$\widehat{N}$	0.69	0.18	0.53	0.34	0.06	0.13
$\widehat{N}_1$	0.72	0.19	0.55	0.28	0.02	0.08

Key:

M: Number of individuals ascertained in at least one list;

D: The average of the number of individuals listed in the combination of any two lists omitting the other one;

$\hat{C}$: Sample coverage estimate;

Est: Population size estimate;

SE: Estimated standard error of the population size estimation based on bootstrap replications;

95% CI: 95% confidence interval for $\boldsymbol{Est}$;

$\widehat{N}$: Population size estimate for sufficiently high sample coverage cases;

$\widehat{N}_1$: One-step population size estimate for low sample coverage cases. This estimator is recommended for use when the estimated sample coverage/overlapping information is too low i.e., the estimated SE of $\widehat{N}$ is relatively large.

u1, u2, u3: estimated mean probabilities depending on the estimate of N.

r12, r13 and r23: estimated coefficient of covariation (CCV) depending on the estimate of N.

Capture-recapture estimate of neonatal deaths assuming independence

The models from ecological, log-linear and sample coverage approaches assuming independence among sources, fitted with different packages, have obtained similar estimates of neonatal mortality over the full 4-year period. The sample coverage approach has estimated neonatal deaths slightly lower (N=9,754; 95% CI: 9686, 9829) than that obtained from the ecological

(N=9,820; 95% CI: 9742, 9899) and log-linear (N=9,819; 95% CI: 9743, 9901) models respectively. These results are depicted in Table 31 below.

Table 31: Selected model results depicting estimated neonatal mortality, assuming independence between the sources (2010 - 2013)

Model	Description	Estimate (95% CI)
M_t	Ecological model	9,820 (95% CI: 9742, 9899)
M_{000}	Log-linear model	9,820 (95% CI: 9743, 9901)
$\widehat{N}_0$	Sample coverage	9,754 (95% CI: 9686, 9829)

M_t, M_{000} and $\widehat{N}_0$ – Models assuming homogeneity of capture and independence

Table 32 shows the distribution of neonatal deaths according to data sources and the year of death when assuming independence between sources based on estimates from the model M_t. A total of 9,820 neonatal deaths were estimated to have occurred between 2010 and 2013 in the Western Cape province. Of these, 839 neonatal deaths were not registered by any source during the same analysis period. The number of neonatal deaths not captured by any source were estimated to have decreased consistently from 2010 to 2013.

Table 32: Estimates and distribution of neonatal deaths by source and year of death assuming independence between sources.

Year of death	Registered deaths			Total unique captured	Not captured	$\widehat{N}$(95% CI)
	PHDC	DHA	CEMD			
2010	1895	501	1501	2412	398 (14.2%)	2810 (2747, 2879)
2011	1682	446	1270	2060	274 (11.7%)	2334 (2285, 2389)
2012	1817	543	1498	2259	258 (10.3%)	2517 (2471, 2568)
2013	1968	434	1383	2250	181 (7.4%)	2431 (2395, 2471)
2010-2013*	**7362**	**1924**	**5652**	**8981**	**839 (8.5%)**	**9,820 (9742, 9899)**

**Values are not the summation of yearly estimates but rather independent estimates*

Capture-recapture estimate of neonatal deaths accounting for source dependencies

Accounting for list effect between sources, the log-linear model, controlling for dependence between PPIP and DHA (S_2*S_3), has estimated neonatal deaths relatively lower than that obtained from the sample coverage approaches i.e., (N=10,146; 95% CI: 10,051, 10,248) *vs.* (N=10,697; 95% CI: 10,508, 10,909) respectively. The best fitting log-linear model that accounts for list dependence does not fit the data adequately. Based on these results, estimates from the sample coverage are used for estimating completeness of neonatal deaths registration over that of the log-linear model. These results are depicted in Table 33 below.

Table 33: Selected model results depicting estimated neonatal mortality accounting dependencies between sources (2010 - 2013)

Software	Model	Description	Estimate (95% CI)
R	$M_{12/3}$	Log linear model	10,146 (95% CI: 10051, 10248)
CARE1	$\widehat{N}$	Sample coverage	10,697 (95% CI: 10508, 10909)

*$M_{23/1}$: Model with one interaction (S_2*S_3) i.e., u12 = u13= 0.*
$\widehat{N}$: Population size estimate for sufficiently high sample coverage cases.

Table 34 shows the distribution of registered neonatal deaths according to source and year of death, as contrasted with the neonatal deaths estimated by the sample-coverage approach. In total, 10,697 neonatal deaths were estimated to have occurred between 2010 and 2013 in the Western Cape province. Of these, 1716 (16.0%) neonatal deaths were not registered by any source. The proportion of neonatal death under-reporting by the three sources was slightly higher in 2010 and 2011 than the overall average (i.e., 20.8% (n=633) and 18.1% (n=455) *vs*. 13.0% (n=1716) respectively) and decreasing over time.

Table 34: Estimates and distribution of neonatal deaths by source and year of death

Year of death	Registered deaths			Total unique captured	Not captured* N (%)	$\widehat{N}$*(95% CI)
	PHDC	PPIP	DHA			
2010	1895	501	1501	2412	633 (20.8)	3045 (2914, 3210)
2011	1682	446	1270	2060	455 (18.1)	2515 (2426, 2625)
2012	1817	543	1498	2259	390 (14.7)	2649 (2579, 2734)
2013	1968	434	1383	2250	272 (10.8)	2522 (2452, 2615)
2010-2013**	**7362**	**1924**	**5652**	**8981**	**1,716 (16.0)**	**10,697 (10508, 10909)**

** Estimated neonatal deaths obtained from model $M_{23/1}$*

***Values are not the summation of individual yearly' estimates but rather independent estimates*

Completeness of neonatal death registrations according to sources and year of death

Table 35 depicts estimates of completeness of neonatal death registration by sources and analysis period. Overall completeness of neonatal death registration was high (84.0%). PHDC had the highest level of registration completeness (68.8%) over the full 4-year period with PPIP recording the lowest registration completeness (18.0%). Relative to other years, the completeness of death registration was slightly lower in 2010 (79.2%), with an increasing trend over the analysis period.

Table 35: Estimates of completeness in neonatal deaths registration by source

Year of death	$\widehat{N}$	PHDC		PPIP		DHA		Total	
		Reg	%Comp	Reg	%Comp	Reg	%Comp	Reg	%Comp
2010	3045	1895	62.2	501	16.5	1501	49.3	2412	79.2
2011	2515	1682	66.9	446	17.7	1270	50.5	2060	81.9
2012	2649	1817	68.6	543	20.5	1498	56.5	2259	85.3
2013	2522	1968	78.0	434	17.2	1383	54.8	2250	89.2
2010-2013	**10697**	**7362**	**68.8**	**1924**	**18.0**	**5652**	**52.8**	**8981**	**84.0**

***Reg** - Number of neonatal deaths registered; **Comp** - Completeness level.*

Table 36 presents the adjustment factors to correct for under-registration of neonatal mortality in the WC province, when considering different criteria. Controlling for dependence between

PPIP and DHA, estimation of maternal mortality in the Western Cape province for the year 2010-2013 requires adjustment by 1.19, when all the sources are used. When focusing on DHA data only (vital registrations), the adjustment/correction factors ranged from 1.59 when assuming no missing ascertainment to 1.89, when assuming dependence.

Table 36: Adjustment factors to correct for under-registration of neonatal deaths in the WC province (2010-2013)

SN	Criteria	Vital registration (DHA)		All 3 sources combined	
		$\widehat{N}/n$	Adjustment factor	$\widehat{N}/n$	Adjustment factor
1	Assuming no missing ascertainment	*8981/5652*	1.59	*8981/8981*	1.0
2	Assuming independence	*9820/5652*	1.74	*9820/8981*	1.09
3	Assuming dependence*	*10697/5652*	1.89	*10697/8981*	1.19

6.5 Discussion

This study applied capture-recapture, a relatively novel statistical method in this context, to estimate under-ascertainment of neonatal deaths by operationalising statistically the overlapping information using three data sources in the Western Cape province, South Africa. This analysis observed a substantial overlap between sources, with fewer deaths recorded by PPIP. Neonatal mortality under-ascertainment from 2010-13 was as low as 16.0%; and decreasing with time. During the same period of reporting, neonatal mortality was estimated to be higher than that from institutional reports and global metrics.

The sample coverage estimator estimated the highest positive dependence between PHDC and PPIP; but the log-linear model accounted for dependence between PPIP and DHA. The analysis

did not yield new insights on the trend over time in the four-year period, when compared with the vital registration.

Feasibility of CRC for neonatal death estimation in the WC

The application of the CRC method has provided a reasonably robust estimate of the neonatal deaths in the Western Cape province for the period 2010 to 2013. Despite its limitations, estimates obtained from this method have been recognised to have reasonably high quality, specifically for neonatal mortality. This method is proven to be an essential tool for assessing and correcting mortality data for registration completeness and case ascertainment using multiple incomplete datasets[9,26–30].

Completeness of neonatal mortality ascertainment in the WC

Using information from the three data sources, our analysis shows that neonatal mortality under ascertainment is small (overall neonatal mortality under-registration=16.0%). By calendar years, the ascertainment of neonatal deaths by the three sources combined was consistently high (84%).

While there have been limited attempts to estimate registration completeness of mortality datasets in the Western Cape province and the country in general, our findings (even though they present sub-national estimates) are in accordance with the global assessment of mortality data in South Africa, as presented by the World Health Organization (WHO) in 2005[8]. Regarding the data sources used in this analysis, completeness of death registrations was lowest in PPIP, when contrasted with the PHDC and the DHA. This could be attributed to the under-reporting of the

late neonatal deaths, as documented in the "Saving babies" report, in addition to the overall incomplete registrations of PPIP, as demonstrated by the incomplete registration of institutional deliveries (ranged between 71% - 75% during the same period), when compared with the DHIS.[31,32] Nevertheless, this report has indicated possible under-notifications of children <1 year deaths by the DHA during the analysis period, making neonatal mortality estimates derived from this source alone (DHA) unreliable[33]. Variations in neonatal death registration highlights the potential to improve death registrations through the establishment of effective links across different data sources in the province.

Trends in neonatal mortality in the WC from 2010-2013

Although it is known to resolve uncertainties when estimating conditions with diverse estimates, the application of capture-recapture methods requires satisfying a set of assumptions about the population of interest[19–21,26,34–36]. Based on our data, the assumption requiring identifiability of individuals across sources to facilitate linkage was partly met. We couldn't identify and match more than 20% of the cases from PPIP and DHA with other records across sources, due to lack of potential identifiers.

The final dataset used for capture-recapture analysis excluded these cases. If the distribution of these cases was not random, we might have underestimated the true burden of neonatal deaths in this setting. Even though this method provides robust estimates of the respective population, it highly relies on and is known to be sensitive to accurate matching of overlapping cases across sources[34–36].

Dependence between sources

Initially, we anticipated the possibility of existing dependencies between PHDC and PPIP due to nature of these two data sources i.e., public-sector facility-based data. The results of sample coverage estimator confirmed this; and they indicated an apparent positive dependence between PHDC and PPIP. However, the application of log-linear models suggested dependence between PPIP and DHA. As indicated by the poor model fit, the observed dependence in log-linear models might be due to heterogeneity than to source/list dependence. The application of the three-source capture-recapture analysis, accounting for source dependence, resulted in an estimate of neonatal deaths (between 2010 and 2013) slightly higher than that obtained from the model assuming independence i.e., 10,697 *vs*. 10,057. Based on the best fitting model from the sample coverage approaches ($\hat{N}$) the estimate of neonatal deaths presented here is slightly higher than that of log-linear models i.e., 10,697 *vs.* 10,146.

Strengths and limitations

The key strengths of this study lie in the number of data sources used for neonatal mortality ascertainment, the application of a sophisticated analytical method, which is known to resolve uncertainties in conditions with diverse estimates, and its ability to control for source dependencies, which tend to bias true estimate of the population of interest. Using a study period, which was concluded some years ago, allowed for an effective estimation of the completeness of registrations.

Although it is advisable to use data with quality unique identifiers, sometimes it is not possible to identify and match all individuals from different sources due to lack of potential identifiers i.e.,

"tag-loss". In the current study, more than 20% of the cases from PPIP and DHA were excluded from the analysis since they were poorly identified and could not be linked with other records across sources.

6.6 Conclusion

Neonatal mortality was under-reported in the Western Cape Province between 2010 and 2013, by any of the three sources evaluated. Even if data from all sources had been combined, there would still have been under-reporting. The continuing use of crude neonatal mortality estimates from a single source (from which the MDG estimates are derived) is subsequently misleading, given the importance of this metric in monitoring trends, assessing progress in global goals (i.e., SDG 3.2),the evaluation of health interventions focusing on reducing neonatal mortality, health planning and evidence-based decision making. Capture-recapture is a viable method that may be used to estimate neonatal mortality rates as well as estimating the level of mortality under-registration by the 'active' and 'passive' mortality reporting systems in the WC province and in the country at large.

6.7 References

1. United Nations Sustainable Development. Sustainable Development Goals: 17 Goals to Transform Our World. *2016* Available at: https://www.un.org/sustainabledevelopment/. (Accessed: 23rd June 2018)

2. Wardlaw, T., You, D., Hug, L., Amouzou, A. & Newby, H. UNICEF Report: Enormous progress in child survival but greater focus on newborns urgently needed. *Reprod. Health* **11**, 82 (2014).

3. UNICEF, WHO, World Bank Group & United Nations. *Levels and Trends in Child Mortality: Report 2017*. (2017).

4. Bradshaw, D. *et al.* Every death counts: use of mortality audit data for decision-making to save the lives of mothers, babies, and children in South Africa. *Lancet* **371**, 1294–1304 (2008).

5. Chopra, M., Daviaud, E., Pattinson, R., Fonn, S. & Lawn, J. E. Saving the lives of South Africa's mothers, babies, and children: can the health system deliver? *Lancet* **374**, 835–846 (2009).

6. Barfield, W. D. & COMMITTEE ON FETUS AND NEWBORN. Standard Terminology for Fetal, Infant, and Perinatal Deaths. *Pediatrics* **137**, e20160551–e20160551 (2016).

7. Merrill, R. M. *Reproductive epidemiology : principles and methods*. (Jones and Bartlett Publishers, 2010).

8. Mathers, C. D., Fat, D. M., Inoue, M., Rao, C. & Lopez, A. D. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull. World Health Organ.* **83**, 171–7 (2005).

9. Carter, K. L. *et al.* Capture-recapture analysis of all-cause mortality data in Bohol, Philippines. *Popul. Health Metr.* **9**, 9 (2011).

10. Damian, D. J., Njau, B., Lisasi, E., Msuya, S. E. & Boulle, A. Trends in maternal and neonatal mortality in South Africa: a systematic review. *Syst. Rev.* (2019).

11. Moodley, J. *et al.* The confidential enquiry into maternal deaths in South Africa: a case study. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 53–60 (2014).

12. Souza, J. P. *et al.* Moving beyond essential interventions for reduction of maternal mortality (the WHO Multicountry Survey on Maternal and Newborn Health): A cross-sectional study. *The Lancet* **381**, 1747–1755 (2013).

13. UNICEF, WHO, World Bank group & United Nations. *Levels & Trends in Child Mortality - Report 2015: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. *UNICEF* (2015). doi:10.1371/journal.pone.0144443

14. WHO, UNICEF, UNFPA, World Bank & United Nations Population Division. Trends in maternal mortality 1990 to 2015. (2015). doi:10

15. Kassebaum, N. J. *et al.* Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **384**, 980–1004 (2014).

16. Wang, H. *et al.* Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* **384**, 957–979 (2014).

17. Saleem, S. *et al.* A prospective study of maternal, fetal and neonatal deaths in low- and middle-income countries. TT -. *Bull. World Health Organ.* **92**, 605–612 (2014).

18. UNICEF, WHO, World Bank, U. *Levels & Trends in Child Mortality in 2014- Report: Estimates Developed by the UN Inter-agency Group for Child Mortality Estimation*. (2014).

19. Forecasting, I. W. G. for D. M. and. Capture-recapture and multiple-record systems estimation I: History and theoretical development. International Working Group for Disease Monitoring and Forecasting. *Am. J. Epidemiol.* **142**, 1047–58 (1995).

20. Hook, E. B. & Regal, R. R. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* **17**, 243–264 (1995).

21. Yip, P. *et al.* Capture-recapture and multiple-record systems estimation I: History and theoretical development. (1995).

22. Republic of South Africa. Births and Deaths Registration Act No. 51 of 1992. Regulations on the Registration of Births and Deaths. *Government Gazette* 74 (2014).

23. Damian, D. J., Distiller, G. & Boulle, A. Performance of a decision-rule based linkage approach, which uses fuzzy linkage over a fully probabilistic record linkage in identifying mortality records across Provincial Health Data, Western Cape, South Africa. *BMC Med. Res. Methodol.* (2019).

24. Heiden, M. an der. RECAP: Stata module to perform capture-recapture analysis for three sources with Goodness-of-Fit based confidence intervals. *Stat. Softw. Components* (2009).

25. Baillargeon, S. & Rivest, L.-P. Rcapture : Loglinear Models for Capture-Recapture. *J. Stat. Softw.* **19**, 1–31 (2007).

26. Van Hest, R., Grant, A. & Abubakar, I. Quality assessment of capture-recapture studies in resource-limited countries. *Tropical Medicine and International Health* **16**, 1019–1041 (2011).

27. Yazdizadeh, B. *et al.* Application of capture-recapture for fine-tuning uncertainties about national maternal mortality estimates. *Int. J. Prev. Med.* **5**, 624–631 (2014).

28. Checchi, F., Nyasulu, P., Chandramohan, D. & Roberts, B. Rates and causes of death in Chiradzulu District, Malawi, 2008: A key informant study. *Trop. Med. Int. Heal.* **16**, 375–378 (2011).

29. Qomariyah, S. N. *et al.* An option for measuring maternal mortality in developing countries: a survey using community informants. *BMC Pregnancy Childbirth* **10**, 74 (2010).

30. Mony, P. K., Varghese, B. & Thomas, T. Estimation of perinatal mortality rate for institutional births in Rajasthan state, India, using capture-recapture technique. *BMJ Open* **5**, e005966 (2015).

31. Pattinson, R. & Rhoda, N. *Saving Babies 2012-2013: Ninth report on perinatal care in South Africa*. *RC Pattinson* (2014).

32. Pattinson, R. & Rhoda, N. *Saving Babies 2010-2011: Eighth report on perinatal care in South Africa*. *RC Pattinson* (2013).

33. Dorrington, R., Bradshaw, D., Laubscher, R. & Nannan, N. *Rapid Mortality Surveillance Report 2015*. (2016). doi:ISBN: 978-1-920618-00-1

34. Rossi, G., Pepe, P., Curzio, O. & March, M. Generalized Linear Models and Capture-Recapture Method in a Closed Population: Strengths and Weaknesses.

35. Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y. & Chao, D. Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* **20**, 3123–3157 (2001).

36. VALID International Ltd. *Notes on using capture-recapture techniques to assess the sensitivity of rapid case-finding methods*. (2006).

CHAPTER 7: Factors associated with maternal mortality in Western Cape province, South Africa

7.1 Abstract

Background: Timely identification of the factors associated with maternal mortality, and appropriate interventions to address these factors; this has the potential to improve maternal health and to reduce maternal mortality in the country. Whereas there is regular detailing in South Africa of the causes of maternal death from case reporting, there are few population-based studies assessing associations with maternal mortality. It is further unclear whether the inclusion of maternal deaths additional to those enumerated in the existing institutional reporting mechanisms, would alter associations with mortality.

Methods: Data of all pregnant women who attended public facilities from 1st January 2010 to 31st December 2013 in the Western Cape province were extracted from the Provincial Health Data Centre (PHDC) SQL database and used in this analysis. Maternal mortality information was ascertained from the Department of Home Affairs (DHA), institutional reporting i.e., Confidential Enquiry into Maternal Death (CEMD), and hospital death reporting system; and matched to the pregnancy data using the provincial linkage algorithm. Logistic regression analyses were used to determine factors independently associated with maternal mortality.

Results: In total, 294,986 women were pregnant from 1st January 2010 to 31st December 2013 in the Western Cape province. Over the full 4-year period, a total of 391 (0.1%) unique maternal deaths were registered by the CEMD (n=238), DHA (n=108) and PHDC (n=205). Factors independently associated with maternal mortality were being HIV positive and not being on ART (AOR=3.05; 95% CI: 2.30, 4.04; p<0.001); or enrolled on ART prior to pregnancy (AOR=1.61; 95%

CI: 1.05, 2.46; *p*<0.001); and being infected with tuberculosis (TB) during the pregnancy, especially if co-infected with HIV (AOR=9.78; 95% CI: 6.49, 14.75; *p*<0.001). In addition, chronic co-morbidities for which data were available (hypertension (AOR=2.28; 95% CI: 1.78, 2.92; *p*<0.001), diabetes (AOR=6.82; 95% CI: 5.32, 8.76; *p*<0.001) and mental health conditions (AOR=8.46; 95% CI: 6.70, 10.68; *p*<0.001)) were strongly associated with mortality.

Conclusion: High burden infectious co-morbidities of HIV and TB which were strongly associated with mortality, as were other measured chronic co-morbidities. The results underscore the importance of interventions focussing on preventing, diagnosing, and optimising treatment for HIV, TB, diabetes, and other high-burden medical co-morbidities in pregnancy.

Keywords: Maternal mortality, maternal deaths, factors, Western Cape, South Africa.

7.2 Introduction

Reducing the maternal mortality ratio (MMR) has been one of the major global health goals over the past 2 decades. In South Africa, MMR remains relatively high, but with room for improvement. The country recorded a 27.8% increase in the MMR with a 1% average annual increase in MMR between 1990 and 2015 (i.e., 108 deaths per 100,000 livebirths in 1990 to 138 deaths per 100,000 livebirths in 2015)[1]. Evidence suggests wide coverage and high-quality maternal healthcare services are among the essential interventions for reducing maternal mortality in countries[2,3].

The Western Cape (WC) remains the only province in South Africa to maintain a relatively low MMR during this period[4–11]. The provincial maternal mortality ratio currently stands at 55.1 deaths per 100,000 live births; a substantially lower ratio than the national average of 105.7 deaths per 100,000 live births[12]. Even though the WC province has a lower MMR than the rest of the country, it is anticipated that further reductions are possible.

Due to substantial investments in coverage and improvements in the quality of maternal healthcare services in the country, the Western Cape province has recorded a slow but steady reduction in maternal deaths resulting from both causes, which could be addressed through service interventions (quality of care) and the factors related to patient characteristics or community-related factors[11,13,14]. The province has achieved a more than 15% reduction in institutional maternal mortality, below the national average, due to non-pregnancy related infections, miscarriage, obstetric haemorrhage, complication of hypertension in pregnancy and

acute collapse[11]. The burden of patient/community-related factors remains poorly evaluated in the country and in the Western Cape province[14].

The causes of maternal deaths (direct and indirect) in the country have been consistently reported in the "Saving mother's reports" (as reported by the National Committee for Confidential Enquiry into Maternal Deaths (NCCEMD)[4–10]), with limited analyses of the factors associated with maternal mortality particularly in the Western Cape province. The latter is due to the fact that the CEMD only records deaths and does not have pregnancies which did not result in deaths, so does not routinely look at associations with mortality, but rather reviews factors associated with the deaths based on case reviews. These data are based on institutional reporting which is considered to underestimate MMR substantially as highlighted in Chapter 3 (systematic review) and in Chapter 5 (Estimation of MMR in the Western Cape province).

Understanding factors associated with maternal mortality is a key requirement for effective health policy and informed public health decision making to improve maternal health. Accessible literature from both developed and developing countries have described various factors associated with maternal mortality. Focusing only on the individual/community-level factors, studies have indicated maternal age, education, marital status, religion, ethnicity, parity, gravidity, income, occupation, area of residence, nutritional inadequacies, and medical co-morbidities (i.e., hypertension, HIV, TB, diabetes, malaria and other medical conditions) to be associated with maternal mortality[15–30].

Timely identification of factors associated with maternal mortality and appropriate management of these factors has the potential to reduce maternal deaths and improve maternal health in the country. Although this is not a clinical review, it is possible to explore patient characteristics and co-morbidities associated with an expanded set of maternal deaths. The aim of this analysis was to describe patient factors associated with maternal mortality in the Western Cape province, South Africa, and determine whether associations based on a more complete dataset of maternal deaths differ from those based merely on institutional reporting.

7.3 Methods

Design and setting

This was a retrospective cohort study conducted to determine factors associated with maternal mortality in the Western Cape province, South Africa. Based on the Western Cape burden of disease reports, the highest burdens of disease are from infectious diseases (HIV/AIDS and tuberculosis (TB)), injuries (interpersonal violence and accidental gunshot) and non-communicable diseases (Ischaemic heart disease, Cerebrovascular disease, diabetes etc.)[31]. In the province, 60% of TB is HIV associated and 80% of diabetics are also hypertensive[12,32].

Data sources and types

Data for this analysis came from the South African Department of Home Affairs (DHA) and Western Cape Department of Health (WCDOH). As discussed in Chapter 2, in the Western Cape province, many administrative data sources are electronically available at an individual patient-level which are managed at the Provincial Health Data Centre (PHDC). In the province, the

pregnancy data can be inferred from a variety of clinical systems, with the mortality data records being available from various sources[33,34].

The main data sources registering maternal deaths in the province include hospital death reporting, death certificates (DHA) and clinical audit tools specific to maternal mortality i.e., Confidential Enquiry into Maternal Deaths (CEMD, as described in detail in Chapter 2). The mortality data used in this analysis comes either from CEMD or the expanded dataset from the three sources.

Pregnancy data from the Provincial Health Data Centre (PHDC) database were extracted and linked with corresponding mortality information from the above sources as described in Chapter 2. Following the amendment to the South African Births and Deaths Registration Act (February 2014) which restricted access to cause-of-death data for privacy and confidentiality, we restricted our analysis to data covering the period 1st January 2010 to 31st December 2013[35].

Outcome and definitions

The primary outcome of this study was a maternal death. In this study, maternal death was defined according to the 10th revision of the International Classification of Diseases and Related Health Problems (ICD-10) as follows: "*the death of either a pregnant woman or death of a woman within 42 days of delivery, spontaneous abortion or termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes*" [36]. However, some of the deaths, in which pregnancy was incidental, or there was an ascertainment issue (not the case with

institutional reporting) might have been included as maternal deaths in the expanded dataset. Therefore, we were unable to definitively exclude incidental causes from the expanded dataset but excluded non-natural deaths.

Predictors of maternal mortality

Variables with established associations with maternal mortality from the literature (as listed in the introduction section), where available, were extracted from the PHDC database and analysed. The nature of this study (secondary data analyses of routine administrative and surveillance data) limited detailed analyses of associations with mortality. The following factors were included in our analyses: maternal age, gravidity, parity, and co-morbidities, including diabetes, hypertension, mental health, TB, and HIV and ART status.

Maternal age was categorised and together with gravidity and parity were analysed as categorical variables in all analyses. To avoid the problem of multicollinearity in the multivariable analyses, gravidity was analysed over parity due to presence of a strong association between the two variables as well as the relative importance of its information in relation to the study's outcome. HIV and ART status were merged to form a single variable with the following categories 1. "Negative or unknown" 2. "HIV positive, not on ART" 3. "HIV positive, started ART prior to pregnancy" and 4. "HIV positive, started ART during pregnancy". TB status was categorised as 1. "Never had TB" 2. "Had TB prior to pregnancy" and 3. "Had TB during pregnancy".

The medical co-morbidities included in the analysis were limited to those conditions where the PHDC had attempted to create province-wide lists of patients with these conditions. These co-

morbidities (hypertension, diabetes, epilepsy and mental health) were identified from the electronic pharmacy and Chronic Disease Dispensing unit (CDU) databases, ICD10 discharge coding, and laboratory investigations, based on similar inference approaches to those described previously for ascertaining pregnancy. A woman was assigned a code 1 “Yes” and 0 “No “, if she had ever received medication for the respective condition or not respectively. However, the available data did not distinguish gestational diabetes from pre-existing diabetes, and hypertension was based just on drugs used for chronic hypertension treatment and does not include gestational hypertension.

Initially, epilepsy was one of the conditions considered for inclusion, but after the first analysis, the very strong association (OR >80) led to a review of the data, and that the strong association between epilepsy evidence and pre-eclampsia, and inability to distinguish pre-existing epilepsy from seizure resulting from eclampsia, led to the decision not to include it in our analysis.

Statistical analysis

Data management and analysis were conducted using Stata software version 15 (Stata Corporation, College Station, Texas, USA). All analyses were restricted to cases with female gender confirmed, and those aged between 10 and 55 years old. Using the final linked dataset, univariable followed by a multivariable logistic regression analysis were used to determine factors associated with maternal mortality based on *a priori* specification of potential associations and causal pathways. Odds ratios (ORs) with their corresponding 95% confidence intervals were used to determine the magnitude and direction of associations. Two models of associations were fitted, with the first model fitting deaths only registered by the CEMD while

the other model fitted deaths registered by all sources in the province i.e., hospital death reporting, death certificates and CEMD.

7.4 Results

General characteristics of the pregnant women in the study

In total, 294,986 women were pregnant from 1st January 2010 to 31st December 2013 in the Western Cape province. Their mean (± Standard Deviation (SD)) age at pregnancy was 26 (± 6.5) years old. A total of 391 unique maternal deaths were registered by the CEMD (n=238), DHA (n=108) and PHDC (n=205) over the full 4-year period. Regarding gravidity, slightly less than a quarter of the women had electronic evidence of previous birth (22.9%). About 17% (n=50,684) of all women had HIV. Of them, 48.34% were on ART. Of 24,503 pregnant women on ART, 53.5% (n=13,116) were enrolled on ART during their current pregnancy. Regarding other medical co-morbidities, 8.8% had hypertension; 8.3% had TB with 8.6% of them having TB during pregnancy; 4.3% had TB/HIV coinfections; 3.7% had diabetes; and 3.6% had mental health condition(s). Table 37 below shows the general characteristics of women in the study.

Table 37: General characteristics of pregnant women in the study (N=294,986).

Variable	Total n	Percent %
Maternal age, years		
<15	809	0.3
15-24	115,850	39.3
25-34	137,369	46.6
35-44	39,572	13.4
45+	1,386	0.5
Mean age ± SD, years	*26 ± 6.5*	
Year of pregnancy		
2010	70,711	24.0

2011	66,236	22.5
2012	72,874	24.7
2013	85,165	28.9
Previous electronic evidence of pregnancy		
No	226,009	76.6
Yes	68,977	23.4
Previous electronic evidence of birth		
No	227,540	77.1
Yes	67,558	22.9
HIV and ART status		
Negative or Unknown	244,388	82.8
Positive, not on ART	19,827	6.7
Positive, started ART prior to pregnancy	11,387	3.9
Positive, started ART during pregnancy	19,470	6.6
TB status		
Negative or unknown	270,453	91.7
Had TB prior to pregnancy	22,407	7.6
Had TB during pregnancy	2,212	0.7
HIV/ TB co-infection		
No	282,328	95.7
Yes	12,744	4.3
Other medical co-morbidities		
Hypertension	26,013	8.8
Diabetes	10,825	3.7
Mental health condition(s)	10,726	3.6

Factors associated with maternal deaths

Table 38 below presents the results of univariable logistic regression analyses of the factors associated with maternal mortality, HIV and ART status, TB, hypertension, diabetes and mental health conditions were associated with maternal mortality in both models.

Based on the results from the expanded dataset (model 2), strong associations were observed with co-morbidities, including HIV, where patients were not on ART (OR=2.91; 95% CI: 2.22,3.82; p<0.001, compared to those who were HIV-negative or had started ART only during the

pregnancy (OR=1.81; 95% CI: 1.05,3.12; *p*<0.001); had TB during pregnancy (OR=9.56; 95% CI: 5.70,16.03; *p*<0.001); had TB/HIV co-infection (OR=2.7; 95% CI: 2.0,3.8; *p*<0.001); were hypertensive or diabetic (OR=3.61; 95% CI: 2.88,4.53; *p*<0.001 and OR=8.85; 95% CI: 7.03,11.14; *p*<0.001 respectively) and with mental health condition(s) (OR=8.46; 95% CI: 6.70,10.68; *p*<0.001). Additionally, association with maternal age was also observed among women aged 35-44 years (OR=1.58; 95% CI: 1.06,1.09; *p*<0.001) and 15-24 years (OR=0.40; 95% CI: 0.30,0.55; *p*<0.001) when compared to 25-34 years.

The effect size for the associations between medical co-morbidities, such as hypertension, diabetes and mental health conditions, were greater when looking just at CEMD deaths ,compared with all maternal deaths. Gravidity, parity and year of pregnancy were not associated with maternal mortality in either model.

Table 38: Univariable logistic regression analyses of the factors associated with maternal mortality

Factors	Total	CEMD deaths† OR [95% CI]	All deaths‡ OR [95% CI]
Maternal age, years			
<15	809	1.20 [0.28, 14.37]	1.02 [0.14, 7.27]
15-24	115,850	0.47 [0.32, 0.71]	0.40 [0.30, 0.55]
25-34	137,369	1	1
35-44	39,572	1.84 [1.28, 2.64]	1.58 [1.21, 2.07]
45+	1,386	1.17 [0.16, 8.38]	2.38 [0.88, 6.42]
Year of pregnancy			
2010	70,711	1	1
2011	66,236	0.88 [0.62, 1.50]	0.98 [0.71,1.34]
2012	72,874	0.89 [0.63, 1.49]	0.95 [0.69,1.29]
2013	85,165	0.52 [0.57, 1.33]	0.73 [0.53, 1.00]
Previous electronic evidence of pregnancy			

No	226,009	1	1
Yes	68,977	0.84 [0.58, 1.28]	1.14 [0.88, 1.48]
Previous electronic evidence of birth			
No	227,540	1	1
Yes	67,558	0.74 [0.50, 1.10]	0.77 [0.58, 1.03]
TB status			
Negative or unknown	270,453	1	1
Had TB prior to pregnancy	22,407	0.94 [0.57, 1.57]	1.09 [0.75, 1.58]
Had TB during pregnancy	2,212	10.21 [6.21, 16.77]	10.40 [7.06, 15.33]
HIV and ART status			
Negative or Unknown	244,388	1	1
Positive, not on ART	19,827	2.90 [2.05, 4.12]	2.91 [2.22, 3.82]
Positive, started ART prior to pregnancy	11,387	1.81 [1.05, 3.12]	2.05 [1.37, 3.07]
Positive, started ART during pregnancy	19,470	1.44 [0.89, 2.31]	1.33 [0.91, 1.96]
HIV/ TB co-infection			
No	282,328	1	1
Yes	12,744	2.96 [2.0,4.4]	2.7 [2.0, 3.8]
Other medical co-morbidities*			
Hypertension	26,013	4.78 [3.63, 6.27]	3.61 [2.88, 4.53]
Diabetes	10,825	13.90 [10.63, 18.17]	8.85 [7.03, 11.14]
Mental health condition[s]	10,726	12.52 [9.53, 16.46]	8.46 [6.70, 10.68]

*†Fitted deaths registered by CEMD only; ‡Fitted deaths registered by all 3 sources; *Reference categories for each condition is not having the disease; **Mean±SD; OR – Odds Ratio.*

In multivariable logistic regression (Table 39); HIV/ART status, diabetes, hypertension and TB status remained associated with maternal deaths in both models. In addition, 14% and 10% of variations in maternal mortality were explained by model 1 model 2, respectively.

The findings from the expanded dataset (model 2) indicated strong associations with medical co-morbidities, including HIV, where the patients were not on ART (AOR=3.05; 95% CI: 2.30, 4.04; p<0.001) or started ART prior to current pregnancy (AOR=1.85; 95% CI: 1.61, 2.46; p<0.001) when

compared to women with no evidence of HIV; had TB during pregnancy (AOR=8.93; 95% CI: 5.91, 13.47; p<0.001); with diabetes (AOR=6.82; 95% CI: 5.32, 8.76; p<0.001); mental health condition(s) (AOR=8.46; 95% CI: 6.70,10.68; p<0.001); and with pre-existing hypertension (AOR=0.41; 95% CI: 0.24, 0.69; p=0.001).

The association of diabetes and mental health condition(s) with maternal deaths were slightly stronger, when looking just at the institutional deaths (model 1) compared with all maternal deaths.

Table 39: Multivariable logistic regression analyses of the factors associated with maternal mortality

Factors	Total	CEMD deaths† aOR [95% CI]	All deaths‡ aOR [95% CI]
Maternal age, years			
<15	809	2.54 [0.35 18.33]	1.23 [0.17, 8.78]
15-24	115,850	0.78 [0.35, 1.08]	0.58 [0.45, 0.76]
25-34	137,369	1	1
35-44	39,572	1.08 [1.78, 1.49]	1.08 [0.83, 1.39]
45+	1,386	0.35 [0.05, 2.50]	0.99 [0.36, 2.68]
HIV and ART status			
Negative or Unknown	244,388	1	1
Positive, not on ART	19,827	3.43 [2.38, 4.92]	3.05 [2.30, 4.04]
Positive, started ART prior to pregnancy	11,387	1.61 [0.91, 2.84]	1.61 [1.05, 2.46]
Positive, started ART during pregnancy	19,470	1.40 [0.85, 2.30]	1.16 [0.78, 1.73]
TB status			
Negative or unknown	270,453	1	1
Had TB prior to pregnancy	24,520	0.77 [0.45, 1.30]	0.89 [0.61, 1.32]
Had TB during pregnancy	2,098	8.98 [5.29, 15.22]	8.93 [5.91, 13.47]
Hypertension			
No	269,062	1	1
Yes	26,010	2.74 [2.03, 5.71]	2.28 [1.78, 2.92]
Diabetes			

No	284,248	1	1
Yes	10,824	10.85 [8.07, 14.59]	6.82 [5.32, 8.76]
Mental health condition[s]			
No	284,346	1	1
Yes	10,726	12.52 [9.53, 16.46]	8.46 [6.70, 10.68]

*†Fitted deaths registered by CEMD only; ‡Fitted deaths registered by all 3 sources; *Mean±SD; aOR – Adjusted Odds Ratio.*

7.5 Discussion

Summary of findings

The findings from this analysis indicated that HIV and access to and timing of ART, TB, pre-existing hypertension, diabetes and mental health condition(s) are factors independently positively associated with maternal mortality. The results of a model fitted to maternal deaths from institutional reporting (model 1) did not materially differ from the findings of the expanded dataset (model 2), except for diabetes and mental health condition(s).

Maternal deaths and HIV/ART status

Consistent with the literature, this research found that being on ART, whether prior to, or initiated during pregnancy, resulted in lower maternal mortality compared to having HIV and not starting ART. i.e., not being on ART increased the risk of maternal deaths among HIV-infected women. Evidence from a systematic review revealed that ART is an effective intervention in reducing mortality among HIV-positive pregnant and postpartum women[37]. Additionally, this analysis has shown that being on ART prior to pregnancy increased risk of maternal deaths compared to starting ART during pregnancy. However, the dataset limited the ability to further understand the underlying mechanisms of these associations; since we did not have the CD4

counts closest to the pregnancy available and data on adherence or treatment interruptions in those who had started ART prior to pregnancy. Recent studies indicated that HIV-infected women initiating ART at an advanced disease stage or with CD counts below 350 cells/mm^3 are thought to have an increased risk of death during pregnancy and during the postpartum period[38–41]. In 2010-2013, pregnant women only initiated ART if the CD4 count was <350. The intermediate risk associated with prior ART could be due to both low CD4 counts in this group (i.e., had advanced disease once when first starting ART, therefore eligible) and probability that some might not have been on ART at time of pregnancy (adherence).

Maternal deaths and TB

An elevated risk of maternal mortality was also observed among women with TB during pregnancy. This finding broadly supports the work of other researchers, who had suggested TB as a common non-obstetric cause of maternal death, as well as a leading cause of maternal mortality in HIV high burden setting[42–47]. In this analysis, 51.8% of TB was HIV associated. Active TB disease in HIV-positive pregnant women has further been linked with a substantially high risk of maternal mortality[42]. Although maternal deaths attributed to TB decreased by 15% from 2011 to 2013 in South Africa, it is still viewed as being among the major causes of maternal deaths in the country[11]. This study therefore shows that medical comorbidities remain important associations with maternal mortality.

Maternal deaths and hypertensive disorders

Another important finding in this study was an increased risk of maternal mortality among women with hypertensive disorders. This result is consistent with existing evidence which

identified hypertension as among the key direct causes of maternal deaths[48–51]. The 7th triennial report on Confidential Enquiry into Maternal Deaths revealed that maternal deaths due to hypertension disorders in pregnancy increased by 14% from 2011 to 2016 in South Africa[11]. Although our study focused on pre-existing hypertension, it might still explain this effect.

Maternal deaths and diabetes

In this analysis, diabetes was strongly associated with maternal mortality. In spite of the strong association, the number of women with diabetes was less than the number with HIV.

Strengths and limitations

This study has a number of limitations, many resulting from the constraints inherent to the dataset which was available for analysis. Both models explained <10% of the variations in the outcome (maternal deaths). Poor model fit as indicated by low goodness-of-fit measures (pseudo R^2) can be explained by nature of data used in this analysis (secondary data analysis of routine medical records and administrative data) which limited a detailed exploration of factors potentially underlying these associations. Secondly, we were unable to definitively exclude incidental causes from the expanded dataset but only excluded non-natural deaths. The impact of this might be limited however, since >50% of the maternal deaths were registered by all three sources. The absence of data on service-factors which might be related to outcomes, and on obstetric causes of death, might have also limited the ability of the models to describe more of the variation in outcomes. The inability to distinguish chronic from pregnancy-related hypertension, and gestational from chronic diabetes, and the absence of data on other co-

morbidities not routinely ascertained at the time by the PHDC, were further important limitations.

The major strengths of this study are the large sample size with co-morbidity data for common conditions available for all pregnancies, and multiple data sources used for maternal mortality ascertainment allowing exploration of associations with the expanded dataset.

7.6 Conclusion

Consistent with the general burden of disease in the setting, this study has shown that untreated HIV and incident TB were the most strongly associated maternal factors with maternal mortality, of those for which data were available. Furthermore, the measured associations only account for a small fraction of the outcome. Although maternal mortality is under-reported in the Western Cape as indicated in Chapter 5, the associations with mortality are in most cases not that different when using the expanded dataset, which is reassuring for ongoing studies on risk factors for maternal mortality if they rely just on the institutional reporting. The results underscore the importance of interventions focusing on preventing, diagnosing, and optimising treatment for HIV, TB, diabetes, and other high-burden medical co-morbidities in pregnancy

7.7 References

1. WHO, UNICEF, UNFPA, World Bank Group & The United Nations Population Division. *Trends in Maternal Mortality: 1990 to 2015. Estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division*. (2015). doi:10

2. Chou, D. *et al.* Ending preventable maternal and newborn mortality and stillbirths. *BMJ* **351**, h4255 (2015).

3. Hunt, P. & Bueno de Mesquita, J. Reducing maternal mortality: the contribution of the right to the highest attainable standard of health.

4. Department of Health. *2014 Saving Mothers: Annual report and detailed analysis of maternal deaths due to non-pregnancy related infections*. (2015).

5. National Committee for Confidential Enquiries into Maternal Deaths. *Saving Mothers 2011-2013: Sixth report on confidential enquiries into maternal deaths in South Africa Short report*. (2014).

6. Pattinson, R. C., Fawcus, S. & Moodley, J. *Tenth interim report on Confidential Enquiries into Maternal Deaths in South Africa*. (2013).

7. National Committee for Confidential Enquiry into Maternal Deaths. *Saving Mothers 2008-2010: Fifth Comprehensive Report on Confidential Enquiries into Maternal Deaths in South Africa*. *Department of Health* **5**, (2012).

8. Moodley, J. Saving mothers: 1999 - 2001. *SAMJ South African Med. J.* **93**, (2003).

9. Moodley, J. Saving mothers in South Africa. *South African Journal of Obstetrics and Gynaecology* **6**, 13–15 (2000).

10. NCCEMD. *Saving Mothers 2005-2007 : Fourth Report on Confidential Enquiries into Maternal Deaths in South Africa*. *Health (San Francisco)* (2008).

11. NCCEMD. *Saving Mothers 2014-2016: Seventh triennial report on confidential enquiry into maternal deaths in South Africa*. (2017). doi:10.2337/dc14-S014

12. Naomi, M., Yogan, P. & Ashnie, P. *District Health Barometer 2017/2018*. (2019).

13. Weterrn Cape Provincial Government. *Annual report 2016 / 2017*. (2017).

14. Moodley, J., Fawcus, S. & Pattinson, R. Improvements in maternal mortality in South Africa. *South African Med. J.* **3**, S4–S8 (2018).

15. Tunçalp, Ö. *et al.* Education and severe maternal outcomes in developing countries: a multicountry cross-sectional survey. *BJOG An Int. J. Obstet. Gynaecol.* **121**, 57–65 (2014).

16. Macleod, J. & Rhode, R. Retrospective follow-up of maternal deaths and their associated risk factors in a rural district of Tanzania. **3**, 130–137 (1998).

17. Weyesa, J. B. *et al.* Prevalence and Risk Factors Associated with Maternal Mortality in Mizan-Aman Hospital, Bench Maji, Southwest Ethiopia. *J. Womens Heal. Care* **04**, (2015).

18. Stokoe, U. Determinants of maternal mortality in the developing world. *Aust. N. Z. J. Obstet. Gynaecol.* **31**, 8–16 (1991).

19. Bayati, M. *et al.* Determinants of maternal mortality in Eastern Mediterranean region: A panel data analysis. *Med. J. Islam. Repub. Iran* **30**, 360 (2016).

20. Abbasi, S. & Younas, M. Determinants of Maternal Mortality in Pakistan. *Mashhad Univ. Med. Sci.* **3**, 430–432 (2015).

21. UNFPA. *The Social Determinants of Maternal Death*. (2012).

22. WHO. *Social determinants approach to maternal deaths*. (World Health Organization, 2015).

23. Høj, L., da Silva, D., Hedegaard, K., Sandström, A. & Aaby, P. Factors associated with maternal mortality in rural Guinea-Bissau. A longitudinal population-based study. *BJOG* **109**, 792–9 (2002).

24. Asamoah, B. O. *et al.* Distribution of causes of maternal mortality among different socio-demographic groups in Ghana; a descriptive study. *BMC Public Health* **11**, 159 (2011).

25. Alvarez, J. L., Gil, R., Hernández, V. & Gil, A. Factors associated with maternal mortality in Sub-Saharan Africa: an ecological study. *BMC Public Health* **9**, 462 (2009).

26. Statistics South Africa. *Millennium Development Goals 5: Improve maternal health 2015 / Statistics South Africa*. (2015).

27. Mmusi-Phetoe, R. M. M. Social factors determining maternal and neonatal mortality in South Africa: A qualitative study. *Curationis* **39**, 1571 (2016).

28. KWAST, B. E. & LIFF, J. M. Factors Associated with Maternal Mortality in Addis Ababa, Ethiopia. *Int. J. Epidemiol.* **17**, 115–121 (1988).

29. Mbonye, A. K. Risk factors associated with maternal deaths in health units in Uganda. *Afr. J. Reprod. Health* **5**, 47–53 (2001).

30. Nair, M. *et al.* Factors associated with maternal death from direct pregnancy complications: a UK national case-control study. *BJOG* **122**, 653–62 (2015).

31. Groenewald, P., Bradshaw, D., Day, C. & Laubscher, R. *Western Cape Burden of Disease Reduction Project. Western Cape Goverment* (2017).

32. Folb, N. *et al.* Multimorbidity, control and treatment of noncommunicable diseases among primary healthcare attenders in the Western Cape, South Africa. *S. Afr. Med. J.* **105**, 642–7 (2015).

33. Boulle, A. *et al.* Data Centre Profile: The Provincial Health Data Centre of the Western Cape Province, South Africa. *Int. J. Popul. Data Sci.* **4**, (2019).

34. Heekes, A. *et al.* Self-enrolment antenatal health promotion data as an adjunct to maternal clinical information systems in the Western Cape Province of South Africa. *BMJ Glob. Heal.* **3**, 565 (2018).

35. Republic of South Africa. Births and Deaths Registration Act No. 51 of 1992. Regulations on the Registration of Births and Deaths. *Government Gazette* 74 (2014).

36. WHO. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. (2016).

37. Holtz, S. A. *et al.* A Systematic Review of Interventions to Reduce Maternal Mortality among HIV-Infected Pregnant and Postpartum Women. *Int. J. MCH AIDS* **4**, 11–24 (2015).

38. Marazzi, M. C. *et al.* Extended antenatal use of triple antiretroviral therapy for prevention of mother-to-child transmission of HIV-1 correlates with favorable pregnancy outcomes. *AIDS* **25**, 1611–1618 (2011).

39. Liotta, G. *et al.* Reduction of Maternal Mortality with Highly Active Antiretroviral Therapy in a Large Cohort of HIV-Infected Pregnant Women in Malawi and Mozambique. *PLoS One* **8**, e71653 (2013).

40. Matthews, L. T. *et al.* HIV-infected women on antiretroviral treatment have increased mortality during pregnant and postpartum periods. *AIDS* **27**, S105–S112 (2013).

41. Li, N. *et al.* Maternal mortality among HIV-infected pregnant women in Tanzania. *Acta Obstet. Gynecol. Scand.* **93**, 463–468 (2014).

42. Zumla, A., Bates, M. & Mwaba, P. The neglected global burden of tuberculosis in pregnancy. *Lancet. Glob. Heal.* **2**, e675-6 (2014).

43. Grange, J. *et al.* Tuberculosis in association with HIV/AIDS emerges as a major nonobstetric cause of maternal mortality in Sub-Saharan Africa. *Int. J. Gynecol. Obstet.* **108**, 181–183 (2010).

44. Adhikari, M. Tuberculosis and tuberculosis/HIV co-infection in pregnancy. *Semin. Fetal Neonatal Med.* **14**, 234–40 (2009).

45. Ahmed, Y. *et al.* A study of maternal mortality at the University Teaching Hospital, Lusaka, Zambia: The emergence of tuberculosis as a major non-obstetric cause of maternal death. *Int. J. Tuberc. Lung Dis.* **3**, 675–680 (1999).

46. Gupta, A. *et al.* Postpartum Tuberculosis Incidence and Mortality among HIV-Infected Women and Their Infants in Pune, India, 2002-2005. *Clin. Infect. Dis.* **45**, 241–249 (2007).

47. Mathad, J. S. & Gupta, A. Tuberculosis in Pregnant and Postpartum Women: Epidemiology, Management, and Research Gaps. *Clin. Infect. Dis.* **55**, 1532–1549 (2012).

48. Nair, M. *et al.* Factors associated with maternal death from direct pregnancy complications: a UK national case-control study. *BJOG An Int. J. Obstet. Gynaecol.* **122**, 653–662 (2015).

49. Khan, K. S., Wojdyla, D., Say, L., Gülmezoglu, A. M. & Van Look, P. F. WHO analysis of causes of maternal death: a systematic review. *Lancet* **367**, 1066–1074 (2006).

50. Say, L. *et al.* Global causes of maternal death: A WHO systematic analysis. *Lancet Glob. Heal.* **2**, 323–333 (2014).

51. Neal, S. *et al.* The causes of maternal mortality in adolescents in low and middle income countries: a systematic review of the literature. *BMC Pregnancy Childbirth* **16**, 352 (2016).

Chapter 8: Factors associated with neonatal mortality in the Western Cape province, South Africa

8.1 Abstract

Background: There are few individual-patient-data analyses in South Africa, or the Western Cape of association with neonatal mortality. However, there is instead a high reliance on attribution of cause of neonatal deaths, or comparisons of the characteristics of neonates, who die with aggregate data, such as comparisons of birth-weight categories for live births. This chapter has aimed to determine the factors associated with neonatal mortality in the Western Cape province, South Africa, and to determine whether associations based on a more complete dataset of neonatal deaths differ from those based only on institutional reporting.

Methods: Data of all infants delivered in public facilities between 1st January 2010 and 31st December 2013 were extracted from the Western Cape Provincial Health Data Centre (PHDC) database. Neonatal death records were ascertained from the hospital death reporting system, deaths certificates (Department of Home Affairs (DHA)) and clinical audit tool specific to perinatal mortality (Perinatal Problem Identification Programme (PPIP)). Using the provincial linkage algorithm, neonatal mortality records and the related maternal characteristics were linked to delivery data before data analysis. Multivariable logistic regression analyses were used to determine factors independently associated with neonatal mortality.

Results: In total, 162,119 livebirths were delivered from 1st January 2010 to 31st December 2013 in the Western Cape province. Overall, 8,757 (5.4%) unique neonatal deaths were registered by the three sources over the full four-year period. Neonates who died were more likely to have had their sex unrecorded (aOR=1.91; 95% CI: 1.35, 2.68; *p*<0.001) or they were identified as male (aOR=1.12; 95% CI: 1.07, 1.19; *p*<0.001). Tuberculosis and syphilis identified during pregnancy were both independently associated with higher neonatal mortality (aOR=1.47; 95% CI: 1.15, 1.90; *p*<0.001) and (aOR=1.35; 95% CI: 1.21, 1.51; *p*<0.001) respectively), but not HIV infection or treatment status. Chronic maternal conditions were also associated with modest increases in neonatal mortality. Except for diabetes, the associations with maternal co-morbidities were attenuated when additionally adjusting for birthweight, where low and very low birthweight were strongly associated with mortality. However, restriction to deaths identified through institutional reporting did not materially change the findings.

Conclusion: Mothers at the extremes of maternal age, and probable first-time mothers, were at higher risk of neonatal death. There were modest associations between a range of maternal co-morbidities and neonatal death, which were mostly attenuated by birthweight, which was strongly associated with neonatal mortality, and is likely also on the causal pathway between maternal health conditions and neonatal death. These results suggest the importance of appropriate follow-up and close monitoring of mothers with high-risk pregnancies and with medical co-morbidities in reducing risk of neonatal deaths.

Keywords: Neonatal mortality, neonatal deaths, factors, Western Cape, South Africa.

8.2 Introduction

Neonatal mortality is a significant public health problem worldwide. Globally, neonatal mortality dropped to 19 (90% Uncertainty Interval (UI): 18, 20) deaths per 1,000 live births in 2016 from 37 (90% UI: 36, 38) deaths per 1,000 live births in 1990, translating to a 49% decline in neonatal deaths worldwide[1]. Despite the decreasing trend in neonatal deaths from 1990 to 2016 globally, it contributes to a growing proportion of under-five mortality each year i.e., neonatal deaths accounted for 46% in 2016 from 41% in 2000 of all under-five deaths[1].

In South Africa, the neonatal mortality rate dropped to 12 deaths per 1,000 live births in 2016 from 20 deaths per 1,000 live births in 1990, a slower decline than that of children aged 1 – 59 months in the country, from 57 deaths per 1,000 live births in 1990 to 43 deaths per 1,000 live births in 2016[1]. Similar to the global pattern, neonatal deaths in South Africa accounted for about half of all under-five deaths in the country[2]. Nevertheless, the Western Cape province remained among the three provinces (Western Cape, Gauteng and KwaZulu-Natal) with an increasing number of early neonatal deaths over time[3].

Most neonatal deaths are considered preventable/avoidable with existing proven cost-effective interventions[4–8]. Evidence suggests that interventions to reduce maternal deaths also prevent neonatal mortality[9–14]. In South Africa, the causes of neonatal deaths have been consistently and adequately reported in the 'Saving Babies' reports based on data from Perinatal Problem Identification Programme (PPIP). Recent reports show that most neonatal deaths in the country are caused by preventable or treatable conditions such as intrapartum-related injuries and infections which contribute about 28% and 8% of these deaths respectively[3]. Prematurity

remains the leading cause of neonatal deaths, accounting for 49.2% of all neonatal deaths in the country[3]. However, much of the current work is based on the audits of explicitly reported deaths, with limited electronic data at a population level available for all neonates.

Previous research findings have shown various factors associated with neonatal mortality, including but not limited to: birthweight; infant sex; birth order; number of living children; birth interval; malnutrition; infants' medical co-morbidities: residence; family socio-economic status; maternal age; maternal education; partner's education; maternal illnesses; and complication during delivery[15–23]. Collectively, these studies outlined a critical role for proper management of high-risk pregnancies, complications of delivery and high-quality postnatal care in neonatal survival.

Despite the slow reduction in neonatal mortality in the country, as compared to the reduction in infant and under-five mortality rates; South Africa has made good progress in achieving the Sustainable Development Goal (SDG) target of reducing neonatal mortality rates to 12 deaths per 1,000 live births or below by 2030[1,24,25]. Nevertheless, the number of neonatal deaths in the country are unacceptably high given the level of government investments in healthcare over time[3].

Identification of factors associated with neonatal mortality may contribute substantially in further reducing the neonatal mortality and improving child health, perhaps through proper planning and improvements of quality of prenatal, childbirth and new-born care in the country. Studies have indicated that a reduction in infant deaths often does not require new interventions,

but rather depends on the availability and effective use of existing scientific evidence and technological innovations[4,5,16,26–28].

Although extensive research has been carried out on causes of neonatal deaths in South Africa and in the Western Cape province, there have been no previous attempts to determine factors associated with neonatal mortality in the Western Cape province, using neonatal deaths ascertained from multiple population-representative data sources. This study aimed to explore factors associated with neonatal mortality in the Western Cape province, South Africa, and determine whether associations based on a more complete dataset of neonatal deaths differ from those based just on institutional reporting.

8.3 Methods

Design and setting

This was a retrospective cohort study conducted to determine factors independently associated with neonatal mortality in the Western Cape, South Africa. The Western Cape province is estimated to account for 11.5% of the national population[29]. The majority of people in the province are estimated to be using public-sector facilities and the vast majority of women deliver in health facilities (institutional deliveries)[2,30]. In 2015/16, the immunisation coverage was estimated at 89% in the province[2,31].

Data source and type

This analysis used data from the South African Department of Home Affairs (DHA) and Western Cape Department of Health (WC-DOH). The Western Cape province is unique in that important

health data are available electronically at an individual patient-level and can be linked across sources using a unique health identifier assigned to each patient when they register at any public facilities for the first time (as presented in Chapter 2). In the province, neonatal mortality data are available and were extracted from the death certificates with known cause of deaths from the DHA, hospital death registrations as collated by the Provincial Health Data Centre (PHDC) and a clinical audit tool specific to perinatal mortality i.e., Perinatal Problem Identification Program (PPIP).

Infant delivery data were obtained from the PHDC SQL database and linked with the neonatal mortality and respective maternal data to form two datasets such that, one includes only deaths registered by PPIP and the second one using deaths registered by all three sources. Historically, most analyses of avoidable causes of neonatal mortality in the province have been based on the PPIP data which comprehensively includes neonates who die in facilities and stillbirths but does not include individuated data on other neonates.

For this analysis, data covering a period between 1st January 2010 and 31st December 2013 were used. The choice of this study period was motivated by the fact that, the study duration concluded long ago, thereby assuring the availability of mortality data across all the sources. Moreover, the amendment to the South African Births and Deaths Registration Act (February 2014) restricted access to the cause-of-death data from 2014 for privacy and confidentiality, which would have limited access to the most recent DHA data[33,34].

Case definition and analysis period

The main outcome in this study was a neonatal death which was defined according to the 10th revision of the International Classification of Diseases and Related Health Problems (ICD-10) as "*death of a baby within the first 28 days of life"* [32]*.*

Independent variables

Due to the nature of this study (secondary data analyses of routine administrative and surveillance data), we were limited to routinely collected data. We included both infant and maternal factors in the analysis in order to have a comprehensive understanding of potential associations. Maternal variables included were as follows: maternal age, gravidity, parity, diabetes, hypertension, rhesus factor, mental health, TB, and HIV and ART status. Infant characteristics included in analyses were sex, birth weight, HIV and ART status.

Ascertainment of other co-morbidities (such as maternal diabetes, hypertension and mental health) were from the electronic pharmacy and Chronic Disease Dispensing unit (CDU) databases, laboratory data and diagnostic coding. We were unable to distinguish gestational diabetes from pre-existing diabetes. Maternal hypertension reflects pre-existing hypertension based on drugs used for chronic hypertension treatment, and not gestational hypertension which could not be enumerated in this analysis.

Measures and coding

Maternal and infant age was categorised and together with gravidity and parity were analysed as categorical variables in all analyses. Due to collinearity between parity and gravidity, we

included gravidity in our multivariable models. Maternal TB status was categorised as 0. “Never had TB” 1. “Had TB prior to pregnancy” and 2. “Had TB during pregnancy”. HIV and ART status were collapsed to form one variable with the following categories: 0. “Negative or unknown”, 1. “HIV positive, no ART”, 2. “HIV positive, enrolled to ART prior to pregnancy”, and 3. “HIV positive, enrolled to ART during pregnancy”. Other medical co-morbidities were categorised as binary variables based on whether or not the woman/infant ever had the condition.

Statistical analysis

Data were managed and analysed using Stata software version 15.1 (Stata Corporation, College Station, Texas, USA). Odds ratios (ORs) with their corresponding 95% confidence intervals were used to determine the magnitude and direction of associations. We fitted two multivariable models of associations with neonatal mortality, the first model with neonatal deaths only registered by PPIP while the other model fitted deaths registered by all three-sources in the province.

Sensitivity analyses

We conducted four sensitivity analyses to understand the effects of birth weight in our models. The first sensitivity analysis fitted deaths ascertained from PPIP with all available potential factors known to be associated with neonatal mortality but including birth weight. Similar to the first sensitivity analysis, we fitted another model including deaths ascertained from all sources as mentioned above and including birth weight. The third and fourth sensitivity analyses fitted deaths ascertained from PPIP and from all sources among neonates who had very low birth

weights against other factors. We hypothesized that; the effects of birth weight might be masking other potential associations since they are in the causal pathway.

8.4 Results

General characteristics of the study subjects

Infant's characteristics

In total, data were available for 161,121 infants delivered from 1st January 2010 to 31st December 2013 in the Western Cape province. Overall, 8,757 (5.4%) unique neonatal deaths were registered by the three sources (1,763 by PIPP only) over the full four-year period. Slightly more than half were males (82,089; 50.6%). Nearly one quarter of the infants (35,810; 22.7%) had low birth weight. A total of 218 infants were documented to have HIV (0.1%). Of them, 81.2% were on ART at the time the data were accessed. Table 40 below shows the infant characteristics.

Table 40: Demographic and clinical characteristics of the infants (N=161,121).

Characteristics	Total N (%)	Neonatal death* n (%)	
		Yes	No
Infant sex			
Female	77,609 (47.9)	3,108 (35.5)	74,501 (48.6)
Male	82,089 (50.6)	3,698 (42.2)	78,391 (51.1)
Not documented	2,421 (1.5)	1,951 (22.3)	470 (0.3)
Year of birth (N=160,216)			
2010	34,293 (21.4)	1,816 (26.5)	32,477 (21.2)
2011	35,157 (21.9)	1,577 (23.0)	33,580 (21.9)
2012	42,567 (26.6)	1,744 (25.4)	40,823 (26.6)
2013	48,199 (30.1)	1,717 (25.1)	46,482 (30.3)
Birth weight (N=157,538)			
Very low (< 1.5 kg)	9,611 (6.1)	3,607 (52.7)	6,004 (4.0)
Low (1.5-2.5 kg)	26,199 (16.6)	1,742 (25.5)	24,457 (16.2)
Normal (≥ 2.5 kg)	121,728 (77.3)	1,490 (21.8)	120,238 (79.8)
Infant HIV status			

Negative or unknown	161,901 (99.9)	8,757 (100)	153,144 (99.9)
Positive	218 (0.1)	0 (0.0)	218 (0.1)
Infant ART status (N=218)			
No	41 (18.8)	-	41 (18.8)
Yes	177 (81.2)	-	177 (81.2)
Total	**162,119**	**8,757 (5.4)**	**153,362 (94.6)**

***** *Ascertained from all 3 sources*

Maternal characteristics

Maternal characteristics are shown in Table 41 below. The mean maternal age (± Standard Deviation [SD]) at delivery was 26.8 (± 6.5) years old. Less than one quarter of the mothers had previous electronic evidence of pregnancy (19.3%), and 19.4% had previous electronic evidence of births. A total of 25,865 (16.0%) mothers had HIV. Of these, 16,051 (62.1%) were on ART. Of 16,051 mothers on ART, 63.1% (n=10,122) were enrolled on ART during their current pregnancy.

Regarding medical co-morbidities, 12.9% of the mothers had hypertension; 8.4% TB; 6.5% diabetes; 5.6% mental health conditions; and 5.1% syphilis.

Table 41: Demographic and clinical characteristics of the mothers (N=161,121)

Characteristics	Total N (%)	Neonatal death* n (%)	
		Yes	No
Maternal age (N=159,462)			
<15	430 (0.3)	23 (0.3)	407 (0.3)
15-24	63,234 (39.7)	2,615 (38.9)	60,619 (39.7)
25-34	74,068 (46.4)	3,122 (46.5)	70,946 (46.4)
35-44	21,297 (13.4)	923 (13.7)	20,374 (13.3)
45+	433 (0.3)	34 (0.5)	399 (0.3)
HIV and ART status			
Negative or Unknown	136,254 (84)	7,404 (84.5)	128,850 (84)
Positive, no ART	9,814 (6.1)	546 (6.2)	9,268 (6.0)
Positive, ART prior to pregnancy	5,929 (3.7)	267 (3.0)	5,662 (3.7)

Positive, ART during pregnancy	10,122 (6.2)	540 (6.2)	9,582 (6.2)
Previous electronic evidence of pregnancy (N=159,462)			
No	128,658 (80.7)	5,313 (79.1)	123,345 (80.8)
Yes	30,804 (19.3)	1,404 (20.9)	29,400 (19.2)
Previous electronic evidence of birth (N=159,462)			
No	128,548 (80.6)	5,171 (77.0)	123,377 (80.8)
Yes	30,914 (19.4)	1,546 (23.0)	29,368 (19.2)
Mother rhesus factor (N=84,217)			
Negative	3,599 (4.3)	169 (4.5)	3,430 (4.3)
Positive	80,672 (95.7)	3,587 (95.5)	77,085 (95.7)
Syphilis (N=81,482)			
Negative	77,357 (94.9)	5,348 (93.5)	72,009 (95.0)
Positive	4,125 (5.1)	370 (6.5)	3,755 (5.0)
Diabetes			
No	151,619 (93.5)	6,230 (71.1)	145,389 (94.8)
Yes	10,500 (6.5)	2,527 (28.9)	7,973 (5.2)
Hypertension			
No	141,234 (87.1)	5,502 (62.8)	135,732 (88.5)
Yes	20,885 (12.9)	3,255 (37.2)	17,630 (11.5)
Mental health			
No	153,049 (94.4)	6,395 (73.0)	146,654 (95.6)
Yes	9,070 (5.6)	2,362 (27.0)	6,708 (4.4)
Mother TB status			
Never had TB or Unknown	148,429 (91.6)	8,046 (91.9)	140,383 (91.5)
Had TB before pregnancy	12,492 (7.7)	630 (7.2)	11,862 (7.7)
Had TB during pregnancy	1,198 (0.7)	81 (0.9)	1,117 (0.7)
Total	**162,119**	**8,757 (5.4)**	**153,362 (94.6)**

***** *Ascertained from all 3 sources*

Factors associated with neonatal mortality

Univariable analysis

Table 42 below shows the results of univariable logistic regression analyses of factors associated with neonatal mortality. Only maternal age and rhesus factor status were not associated with neonatal mortality in both models. Additionally, the model fitting only PPIP deaths (model 1)

indicated a strong association between infants born to HIV mothers who were not on ART or those who started ART during pregnancy. These results differ from those of the expanded dataset whereby only infants born to HIV mothers who started ART prior to pregnancy, were associated with neonatal deaths.

Based on the results of the expanded dataset (model 2), strong associations were also observed with maternal co-morbidities, including syphilis (OR=1.33; 95% CI: 1.18, 1.48; *p*<0.001); diabetes (OR=7.40; 95% CI: 7.03, 7.79; *p*<0.001); hypertension (OR= 4.55; 95% CI: 4.35, 4.77; *p*<0.001); mental health conditions (OR=8.07; 95% CI: 7.66, 8.52; *p*<0.001); and TB during pregnancy (OR=1.27; 95% CI: 1.01, 1.59; *p*=0.011).

Infant factors strongly associated with neonatal mortality included the male sex (OR=1.13; 95% CI: 1.08, 1.19; *p*<0.001) or undocumented infant's sex (OR=99.5; 95% CI: 89.41, 110.73; *p*<0.001) compared to female sex; born to mothers with electronic evidence of previous pregnancy (OR=1.11; 95% CI: 1.04, 1.18; *p*=0.001); and born to HIV mothers who started ART prior to pregnancy (OR=0.82; 95% CI: 0.72, 0.93; *p*=0.002) were maternal factors associated with neonatal mortality.

Associations of diabetes, hypertension and mental health conditions with maternal deaths were generally stronger when looking just at all neonatal deaths (model 2) compared to the PPIP deaths (model 1).

42: Univariable logistic regression analyses of factors associated with neonatal mortality (N=161,121)

	PPIP deaths†	**All deaths**‡
Factors	OR [95% CI]	OR [95% CI]

Infant sex		
Female	1	1
Male	1.22 [1.11, 1.34]	1.13 [1.08, 1.19]
Not documented	0.67 [0.41, 1.10]	99.50 [89.41, 110.73]
Year of birth (N=160,216)		
2010	1	1
2011	0.87 [0.76, 0.99]	0.84 [0.78, 0.90]
2012	0.88 [0.78, 1.00]	0.76 [0.71, 0.81]
2013	0.63 [0.55, 0.72]	0.66 [0.62, 0.71]
Maternal age, years		
<15	1.64 [0.78, 3.48]	1.28 [0.84, 1.96]
15-24	1.03 [0.93,1.14]	0.98 [0.93, 1.03]
25-34	1	1
35-44	1.26 [1.10. 1.45]	1.02 [0.96, 1.11]
45+	1.63 [0.77, 3.46]	1.94 [1.36, 2.75]
Previous electronic evidence of pregnancy (N=159,462)		
No	1	1
Yes	1.05 [0.93, 1.18]	1.11 [1.04, 1.18]
Previous electronic evidence of birth (N=159,462)		
None	1	1
One birth	1.08 [0.95, 1.23]	1.16 [1.09, 1.23]
More than one births	1.47 [1.14, 1.91]	1.91 [1.69, 2.16]
Maternal HIV and ART status		
Negative or Unknown	1	1
Positive, no ART	1.47 [1.24, 1.4]	1.03 [0.94, 1.12]
Positive, ART prior to pregnancy	1.09 [0.86, 1.404]	0.82 [0.72, 0.93]
Positive, ART during pregnancy	1.34 [1.13, 1.60]	0.98 [0.90, 1.07]
Mother TB status		
Never had TB or Unknown	1	1
Had TB before pregnancy	1.30 [1.11, 1.53]	0.93 [0.85, 1.01]
Had TB during pregnancy	1.75 [1.15, 2.68]	1.27 [1.01, 1.59]
Other maternal co-morbidities*		
Syphilis (N=81,482)	1.06 [0.85,1.33]	1.33 [1.18, 1.48]
Diabetes	1.82 [1.57, 2.12]	7.40 [7.03, 7.79]
Hypertension	1.89 [1.69, 2.12]	4.55 [4.35, 4.77]
Mental health condition(s)	1.89 [1.61, 2.21]	8.07 [7.66, 8.52]

†Fitted deaths registered by PPIP only; ‡Fitted deaths registered by all 3 sources; OR – Odds Ratio;
**Reference categories for each condition is not having the disease/condition or missing records.*

Multivariable analysis

Table 43 shows the multivariable logistic regression analyses of factors associated with neonatal mortality. Infant's sex, year of birth and mother's hypertension status, TB and ART prior to pregnancy were the factors independently associated with neonatal mortality in both models.

Maternal conditions such as syphilis, diabetes and mental health condition(s) were the only factors not associated with neonatal deaths on model 1 (PPIP deaths); but they were associated with neonatal deaths in the expanded dataset (model 2).

Focusing on results of the expanded dataset (model 2); infant sex, year of birth, previous electronic evidence of pregnancy, mother's enrolment on ART prior to pregnancy, mother's diabetes, TB and hypertension status were factors independently associated with neonatal deaths. Positive associations were observed among male infants (aOR=1.12; 95% CI: 1.07, 1.19; p<0.001) or infants with undocumented sex infants (aOR=1.91; 95% CI: 1.35, 2.68; p<0.001); and those born to mothers with diabetes (aOR=1.61; 95% CI: 1.42, 1.82; p<0.001) or syphilis (aOR=1.35; 95% CI: 1.21, 1.51; p<0.001). Similar effect sizes but more certainty in the expanded dataset for maternal age, were observed, but with tighter confidence intervals. TB association showed a gradient based on prior (aOR=1.11; 95% CI: 1.00, 1.22; p<0.001) or active TB (aOR=1.47; 95% CI: 1.15, 1.90; p<0.001).

Negative associations were observed among infants born to mothers with electronic evidence of pregnancy (aOR=0.86; 95% CI: 0.80, 0.93; p<0.001); previous electronic evidence of pregnancy (aOR=0.88; 95% CI: 0.83, 0.94; p<0.001); born to mothers who started ART prior to pregnancy

(aOR=0.85; 95% CI: 0.74, 0.97; *p*=0.001); and those born to mothers with history of hypertension (aOR=0.90; 95% CI: 0.83, 0.98; *p*=0.017).

Table 43: Multivariable logistic regression analyses of the factors associated with neonatal mortality (N=161,121)

Factors	PPIP deaths † aOR [95% CI]	All deaths‡ aOR [95% CI]
Infant sex		
Female	1	1
Male	1.22 [1.10, 1.36]	1.12 [1.07, 1.19]
Not documented	1.67 (0.86, 3.26)	1.91 (1.35, 2.68)
Year of birth		
2010	1	1
2011	0.79 [0.68, 0.91]	0.82 [0.76, 0.88]
2012	0.87 [0.76, 0.99]	0.74 [0.69, 0.80]
2013	0.89 [0.77, 1.03]	0.89 [0.83, 0.96]
Maternal age, years		
<15	1.86 [0.68, 0.91]	1.67 [1.04, 2.67]
15-24	1.10 [0.98, 1.23]	1.06 [1.00, 1.12]
25-34	1	1
35-44	1.25 [1.07, 1.45]	0.98 [0.90, 1.06]
45+	1.70 [0.80, 3.64]	1.99 [1.37, 2.92]
Previous electronic evidence of pregnancy		
No	1	1
Yes	0.85 [0.74, 0.96]	0.88 [0.90, 1.10]
Mother HIV and ART status		
Negative or Unknown	1	1
Positive, no ART	1.04 [0.86, 1.24]	0.99 [0.90, 1.10]
Positive, ART prior to pregnancy	0.75 [0.57, 0.98]	0.85 [0.74, 0.97]
Positive, ART during pregnancy	0.97 [0.80, 1.18]	0.99 [0.90, 1.10]
TB history		
Never had TB or Unknown	1	1
Had TB before pregnancy	1.22 [1.02, 1.46]	1.11 [1.00, 1.22]
Had TB during pregnancy	1.68 [1.08, 2.62]	1.47 [1.15, 1.90]
Other maternal co-morbidities*		
Syphilis	1.07 [0.86, 1.35]	1.35 [1.21, 1.51]
Diabetes	0.96 [0.78, 1.20]	1.23 [1.11, 1.37]

Hypertension	1.34 [1.17, 1.54]	1.52 [1.41, 1.63]
Mental health condition(s)	1.14 [0.90, 1.45]	1.14 [1.00, 1.29]

†Fitted deaths registered by PPIP only; ‡Fitted deaths registered by all 3 sources;
**Reference categories for each condition is not having the disease/condition or missing records; aOR – Adjusted Odds Ratio.*

Sensitivity analyses

Table 44 below presents the sensitivity analyses, when including birth weight in the models determining the factors independently associated with neonatal mortality.

Including birth weight in the multivariable model that fitted PPIP deaths only (model 1) did not change the presence of associations, but only the effect sizes, with the exception of TB which was no longer associated with neonatal mortality. Focusing on the expanded dataset (model 2); including birth weight in the model, resulted in a reversed association with hypertension and the attenuated effect of syphilis.

Table 44: Multivariable logistic regression analyses of factors associated with neonatal mortality when including the infant's birth weight

	PPIP deaths†	**All deaths‡**
Factors	aOR [95% CI]	aOR [95% CI]
Infant sex		
Female	1	
Male	1.32 [1.18, 1.47]	1.26 [1.19, 1.34]
Not documented	1.40 (0.69, 2.86)	2.08 (1.36, 3.17)
Year of birth		
2010	1	1
2011	0.86 [0.74, 1.01]	0.88 [0.81, 0.96]
2012	1.10 [0.95, 1.27]	0.88 [0.81, 0.96]
2013	1.11 [0.95, 1.30]	1.12 [0.03, 1.23]
Birth weight		
Very low (< 1.5 kg)	41.99 [36.52, 48.28]	48.21 [44.61, 52.09]
Low (1.5-2.5 kg)	4.50 [3.80, 5.34]	6.11 [5.64, 6.61]

Normal (≥ 2.5 kg)	1	1
Maternal age, years		
<15	1.82 [0.75, 4.44]	1.87 [1.08, 3.25]
15-24	1.06 [0.94, 1.19]	1.01 [0.95, 1.09]
25-34	1	1
35-44	1.28 [1.09, 1.49]	0.97 [0.88, 1.06]
45+	1.67 [0.74, 3.75]	2.35 [1.49, 3.69]
Previous electronic evidence of pregnancy		
No	1	1
Yes	0.85 [0.75, 0.97]	0.86 [0.80, 0.93]
Mother HIV and ART status		
Negative or Unknown	1	1
Positive, no ART	1.05 [0.87, 1.27]	0.99 [0.89, 1.11]
Positive, ART prior to pregnancy	0.70 [0.53, 0.93]	0.76 [0.65, 0.89]
Positive, ART during pregnancy	0.92 [0.75, 1.12]	0.94 [0.84, 1.05]
TB history		
Never had TB or Unknown		
Had TB before pregnancy	1.01 [0.84, 1.22]	0.90 [0.80, 1.00]
Had TB during pregnancy	1.08 [0.68, 1.73]	0.96 [0.72, 1.28]
Other maternal co-morbidities*		
Positive rhesus factor		
Syphilis	0.87 [0.69, 1.11]	1.14 [1.00, 1.29]
Diabetes	1.11 [0.88, 1.40]	1.61 [1.42, 1.82]
Hypertension	0.79 [0.69, 0.91]	0.90 [0.83, 0.98]
Mental health condition(s)	1.08 [0.84, 1.39]	1.08 [0.93, 1.24]

†Fitted deaths registered by PPIP only; ‡Fitted deaths registered by all 3 sources
**Reference categories for each condition is not having the disease/condition or missing records; aOR – Adjusted Odds Ratio.*

8.5 Discussion

Summary of findings

Mortality in the first 28 days of life plays a crucial role in the childhood mortality through its high rates and a slow decline over time[1,2]. It provides an indicator for quality of maternal and child health care services in the country. Excluding birth weight, fitting a multivariable model with neonatal deaths ascertained from the three sources in the Western Cape province between 2010

to 2013, this study has identified later year of birth; mother's enrolment on ART prior to pregnancy as being negatively associated with neonatal deaths; and infant male sex; maternal age (youngest and oldest age categories); and Maternal morbidities including TB, syphilis, diabetes, hypertension and mental health condition(s) being positively associated with neonatal mortality. In the dataset based on the smaller number of outcomes, many of the confidence intervals were wider and the associations would have been considered less certain in the absence of the expanded dataset, except for syphilis and diabetes in which the effect was not evident in the model based on PPIP events. Including birth weight in the multivariable model, attenuated some of the maternal morbidity associations with neonatal mortality (similar in both datasets) including TB, hypertension, and syphilis.

Infant characteristics associated with neonatal characteristics

It has been thoroughly documented that male infants have an increased risk of perinatal morbidity and mortality compared to their female counterparts.[35–39]. In this study. although the predominance of adverse neonatal outcomes in males has been proposed in many singleton and twin studies, the underlying mechanism contributing to this difference is not well elucidated[35–39]. The observed gender disparity in this study is speculated to be partially explained by the fact that female macerated deaths are less likely to have sex recorded, which might be consistent with a 22.3% (1,951/8,757) of neonatal deaths without sex documented.

In this analysis, we have observed that neonates born in later years had slightly lower mortality than in 2010 but there was no consistent downward trend. Decreased risk of neonatal mortality among infants born in later years might be explained by the improved birth registrations which

resulted in the improved mortality reporting systems. A previous report has indicated that, by 2011, less than 10% of the registered neonatal deaths were captured on the national population register[8]. The possible reason for this discrepancy could be failure to register births or to register deaths, as both registrations are required for a death to appear on the population register.

As expected, based on previous studies, there was a very strong association between birth weight and neonatal mortality. This result corroborates the findings of a great deal of the previous work from both retrospective and prospective studies which showed low birth weight is a risk factor for neonatal mortality and low birth weight can be a symptom of maternal, placental or intra-uterine fetal disease[19,40–48].

Maternal associations with neonatal deaths

Findings from this study corroborate extensively described associations with adverse pregnancy outcomes in the very young and much older women as well as infants born to mothers with electronic evidence of pregnancy. This accords with previous research indicating that mothers, who had a history of births, had less risk of losing their babies to neonatal mortality, although this is not the case with mothers with more than 5 pregnancies[49,50]. It seems possible that this result is due to the level and quality of antenatal, delivery and new-born healthcare services offered to women with previous history of pregnancy/birth (especially the multigravidity/multiparous) due to increased risks of adverse neonatal outcomes associated with gravidity/parity such as prematurity, intrauterine growth restrictions and mortality[51–53]. There are, however, other possible explanations.

Another important finding was that babies born to mothers who were enrolled on ART prior to the pregnancy had lesser odds of neonatal mortality compared to those born from mothers with unknown or negative HIV status. A possible explanation for this result is that mothers with history of medical comorbidities might be receiving close follow-up during pregnancy and childbirth compared to the general population in order to improve their maternal and new-born outcomes. Pre-pregnancy interventions such as prevention, treatment and management of STIs including HIV have significant impacts on reducing maternal and neonatal morbidity and mortality[11].

All maternal co-morbidities on which there were data available, showed modest associations with neonatal mortality. These are mostly expected based on prior studies. In particular, this study supports evidence from previous research that indicated infants born from mothers who had diabetes had high risk of neonatal mortality[54,55]. Maternal diabetes is associated with an increased risks of maternal and new-born complications, morbidity and mortality[55,56]. This relationship can be explained through the high morbidity risks and large size among infants born to diabetic women they possess. The major morbidities associated with these infants includes macrosomia, hypoglycaemia, hypomagnesaemia, hypocalcaemia, congenital malformations, respiratory distress, and hyperbillirubinaemia[56]. The findings support recommendations for close follow-up among mothers with co-morbidities such as diabetes and provision of quality care during pregnancy, labour and delivery for better maternal and new-born outcomes.

Differences in association based on the single versus expanded mortality datasets

When compared to the expanded dataset (model 2), the model of association with deaths from PPIP (model 1) showed similar effect sizes, with the exception of maternal co-morbidities. Particularly, larger effect sizes for the co-morbidity association were observed in the expanded dataset. The possible reason for this discrepancy might be the level of completeness of neonatal mortality registration PPIP had relative to the full model in our analysis i.e., 20% mortality registration level.

However, these findings highlight the importance of including mortality data ascertained from multiple data sources, in order to make inferences about the population parameters. Analyses of associations based on only PPIP deaths should be interpreted with care especially in relation to maternal co-morbidities.

Findings from the sensitivity analyses

The results of the sensitivity analyses indicated that including birth weight in the multivariable models did not change the associations largely but only the effect sizes especially for the co-morbidities associations. The possible explanation for this might be the influence of birth weight in describing neonatal mortality since it is in the causal pathway of this association i.e., risk factor for neonatal death.

Strengths and limitations

The key strengths for this study are the large data set (sample size) covering a 4-year period, and ascertainment of mortality data from three independent sources which ensured representativeness and completeness of registrations. Some variables where data were missing, the missingness was considered informative, and preferable to include rather than using multiple imputation. Negative ascertainment was not always possible for some health conditions where absence was assumed. In addition, we failed to clearly differentiate diabetes and gestational diabetes, and hypertension and gestational hypertension since enumeration of these variables were done based on drugs used for chronic diabetes and hypertension treatment, respectively.

8.6 Conclusion

Findings from this analysis indicated a clear association of medical co-morbidities in the mother with increased neonatal deaths, more apparent in the expanded dataset except for the hypertension effect in the model with birth weight. Although speculative, our findings indicated being in care for a well-managed co-morbidity (ART) can result in improved outcomes, suggesting value to antenatal interventions. Additionally, low birth weight remains independently associated with neonatal mortality, with a clear increasing risk based on birth weight stratification. Taken together, these results suggest the importance of appropriate follow-up and close monitoring of mothers with high-risk pregnancies and with medical co-morbidities, in reducing risk of neonatal deaths.

8.7 References

1. UNICEF, WHO, World Bank Group & United Nations. *Levels and Trends in Child Mortality: Report 2017*. (2017).

2. National Department of Health (NDoH), Statistics South Africa Council (Stats SA), South African Medical Research (SAMRC) & ICF. *South Africa Demographic and Health Survey 2016: Key Indicators.* (2017).

3. Rhoda, N., Velaphi, S., Gebhardt, G. S., Kauchali, S. & Barron, P. Reducing neonatal deaths in South Africa: Progress and challenges. *South African Med. J.* **108**, 9 (2018).

4. Gülmezoglu, A. M. *et al. Interventions to Reduce Maternal and Newborn Morbidity and Mortality*. *Reproductive, Maternal, Newborn, and Child Health: Disease Control Priorities, Third Edition (Volume 2)* (The International Bank for Reconstruction and Development / The World Bank, 2016). doi:10.1596/978-1-4648-0348-2_CH7

5. Wardlaw, T., You, D., Hug, L., Amouzou, A. & Newby, H. UNICEF Report : enormous progress in child survival but greater focus on newborns urgently needed. *Reprod. Health* **11**, 1–4 (2014).

6. Dorrington, R., Bradshaw, D. & Laubscher, R. *Rapid Mortality Surveillance Report 2012*. (2014). doi:ISBN: 978-1-920618-00-1

7. Pattinson, R. & Rhoda, N. *Saving Babies 2012-2013: Ninth report on perinatal care in South Africa*. *RC Pattinson* (2014).

8. Dorrington, R., Bradshaw, D., Laubscher, R. & Nannan, N. *Rapid Mortality Surveillance Report 2015*. (2016). doi:ISBN: 978-1-920618-00-1

9. Lassi, Z. S., Salam, R. A., Das, J. K. & Bhutta, Z. A. Essential interventions for maternal, newborn and child health: background and methodology. *Reprod. Health* **11 Suppl 1**, S1 (2014).

10. Lassi, Z. S. *et al.* Essential interventions: implementation strategies and proposed packages of care. *Reprod. Health* **11 Suppl 1**, S5 (2014).

11. Lassi, Z. S., Mansoor, T., Salam, R. A., Das, J. K. & Bhutta, Z. A. Essential pre-pregnancy and pregnancy interventions for improved maternal, newborn and child health. *Reprod. Health* **11 Suppl 1**, S2 (2014).

12. Gates, M. & Binagwaho, A. Newborn health: A revolution in waiting. *The Lancet* **384**, 2013–2014 (2014).

13. United Nations. *The Millennium Development Goals Report*. (2014).

14. Darmstadt, G. L. *et al.* Evidence-based, cost-effective interventions: How many newborn

babies can we save? *Lancet* **365**, 977–988 (2005).

15. Kinney, M. V *et al.* Sub-Saharan Africa's mothers, newborns, and children: Where and why do they die? *PLoS Med.* **7**, 1–9 (2010).

16. Kassar, S. B., Melo, A. M. C., Coutinho, S. B., Lima, M. C. & Lira, P. I. C. Determinants of neonatal death with emphasis on health care during pregnancy, childbirth and reproductive history. *J. Pediatr. (Rio. J).* **89**, 269–277 (2013).

17. Ezeh, O. K., Agho, K. E., Dibley, M. J., Hall, J. & Page, A. N. Determinants of neonatal mortality in Nigeria: evidence from the 2008 demographic and health survey. *BMC Public Health* **14**, 521 (2014).

18. Adetola, A. O., Tongo, O. O., Orimadegun, A. E. & Osinusi, K. Neonatal Mortality in an Urban Population in Ibadan, Nigeria. *Pediatr. Neonatol.* **52**, 243–250 (2011).

19. Lukonga, E. & Michelo, C. Factors associated with neonatal mortality in the general population: evidence from the 2007 Zambia Demographic and Health Survey (ZDHS); a cross sectional study. *Pan Afr. Med. J.* **20**, 64 (2015).

20. Mturi, A. J. & Curtis, S. L. The determinants of infant and child mortality in Tanzania. *Health Policy Plan.* **10**, 384–94 (1995).

21. Khadka, K. B., Lieberman, L. S., Giedraitis, V., Bhatta, L. & Pandey, G. The socio-economic determinants of infant mortality in Nepal: analysis of Nepal Demographic Health Survey, 2011. *BMC Pediatr.* **15**, 152 (2015).

22. Selemani, M. *et al.* The effect of mother's age and other related factors on neonatal survival associated with first and second birth in rural, Tanzania: evidence from Ifakara health and demographic surveillance system in rural Tanzania. *BMC Pregnancy Childbirth* **14**, 240 (2014).

23. Machado, C. J. & Hill, K. Determinants of neonatal and post-neonatal mortality in the City of São Paulo. *Rev. Bras. Epidemiol.* **6**, 345–358 (2003).

24. United Nations. *Transforming our world: the 2030 Agenda for Sustainable Development: Sustainable Development Knowledge Platform*. (2015).

25. United Nations Sustainable Development. Sustainable Development Goals: 17 Goals to Transform Our World. *2016* Available at: https://www.un.org/sustainabledevelopment/. (Accessed: 23rd June 2018)

26. De Carvalho, M. & Gomes, M. Mortality of very low birth weight preterm infants in Brazil: reality and challenges. *J. Pediatr. (Rio. J).* **81**, S111–S118 (2005).

27. Institute of Medicine (US) Committee on Improving Birth Outcomes., Bale, J. R., Stoll, B. J. & Lucas, A. O. Reducing Neonatal Mortality and Morbidity. in *Improving Birth Outcomes:*

Meeting the Challenge in the Developing World. (National Academies Press (US), 2003).

28. Lassi, Z. S., Middleton, P. F., Crowther, C. & Bhutta, Z. A. Interventions to Improve Neonatal Health and Later Survival: An Overview of Systematic Reviews. *EBioMedicine* **2**, 985–1000 (2015).

29. Statistics South Africa. *Statistical release P0302: Mid - year population estimates 2017*. (2017).

30. Western Cape Provincial Centre Health Data Centre. *Strategic approach to patient - level health data harmonisation and integration Western Cape Department of Health International Classification of Diseases*. (2013).

31. Western Cape Government. *City of Cape Town Socio-Economic profile*. (2016).

32. Republic of South Africa. Births and Deaths Registration Act No. 51 of 1992. Regulations on the Registration of Births and Deaths. *Government Gazette* 74 (2014).

33. Groenewald, P. *et al.* The importance of identified cause-of-death information being available for public health surveillance, actions and research. *South African Med. J.* **105**, 528–530 (2015).

34. WHO. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. (2016).

35. Zhao, D., Zou, L., Lei, X. & Zhang, Y. Gender Differences in Infant Mortality and Neonatal Morbidity in Mixed-Gender Twins. *Sci. Rep.* **7**, 8736 (2017).

36. Melamed, N., Yogev, Y. & Glezerman, M. Effect of Fetal Sex on Pregnancy Outcome in Twin Pregnancies. *Obstet. Gynecol.* **114**, 1085–1092 (2009).

37. Steen, E. E., Källén, K., Maršál, K., Norman, M. & Hellström-Westas, L. Impact of sex on perinatal mortality and morbidity in twins. *J. Perinat. Med.* **42**, 225–231 (2014).

38. Stevenson, D. K. *et al.* Sex differences in outcomes of very low birthweight infants: the newborn male disadvantage. *Arch. Dis. Child. Fetal Neonatal Ed.* **83**, F182-5 (2000).

39. Hou, L. *et al.* Cross sectional study in China: fetal gender has adverse perinatal outcomes in mainland China. *BMC Pregnancy Childbirth* **14**, 372 (2014).

40. McCormick, M. C. The Contribution of Low Birth Weight to Infant Mortality and Childhood Morbidity. *N. Engl. J. Med.* **312**, 82–90 (1985).

41. Ezeaka, V. C., Ekure, E. N., Iroha, E. O. & Egri-Okwaji, M. T. C. Outcome of low birth weight neonates in a tertiary health care centre in Lagos, Nigeria. *Afr. J. Med. Med. Sci.* **33**, 299–303 (2004).

42. Schoeps, D. *et al.* [Risk factors for early neonatal mortality]. *Rev. Saude Publica* **41**, 1013–22 (2007).

43. Nascimento, R. M. do, Leite, Á. J. M., Almeida, N. M. G. S. de, Almeida, P. C. de & Silva, C. F. da. Determinants of neonatal mortality: a case-control study in Fortaleza, Ceará State, Brazil. *Cad. Saude Publica* **28**, 559–572 (2012).

44. Barros, F. C. *et al.* The challenge of reducing neonatal mortality in middle-income countries: findings from three Brazilian birth cohorts in 1982, 1993, and 2004. *Lancet* **365**, 847–854 (2005).

45. Yasmin, S., Osrin, D., Paul, E. & Costello, A. Neonatal mortality of low-birth-weight infants in Bangladesh. *Bull. World Health Organ.* **79**, 608–14 (2001).

46. McIntire, D. D., Bloom, S. L., Casey, B. M. & Leveno, K. J. Birth Weight in Relation to Morbidity and Mortality among Newborn Infants. *N. Engl. J. Med.* **340**, 1234–1238 (1999).

47. Hinderaker, S. G. *et al.* Perinatal Mortality in Northern Rural Tanzania. *Journal of Health, Population and Nutrition* **21**, 8–17 (2003).

48. Lansky, S. *et al.* Birth in Brazil survey: neonatal mortality, pregnancy and childbirth quality of care. *Cad. Saude Publica* **30 Suppl 1**, S1-15 (2014).

49. Yirgu, R., Molla, M. & Sibley, L. Determinants of neonatal mortality in rural Northern Ethiopia: A population based nested case control study. *PLoS One* **12**, e0172875 (2017).

50. Kozuki, N. *et al.* The associations of parity and maternal age with small-for-gestational-age, preterm, and neonatal and infant mortality: a meta-analysis. *BMC Public Health* **13 Suppl 3**, S2 (2013).

51. Kiely, J. L., Paneth, N. & Susser, M. An assessment of the effects of maternal age and parity in different components of perinatal mortality. *Am. J. Epidemiol.* **123**, 444–54 (1986).

52. Shah, P. S. & Knowledge Synthesis Group on Determinants of LBW/PT births. Parity and low birth weight and preterm birth: a systematic review and meta-analyses. *Acta Obstet. Gynecol. Scand.* **89**, 862–875 (2010).

53. Scholl, T. O. *et al.* Young maternal age and parity. Influences on pregnancy outcome. *Ann. Epidemiol.* **2**, 565–75 (1992).

54. Kheir, A. E. M., Berair, R., Gulfan, I. G. I., Karrar, M. Z. & Mohammed, Z. A. O. Morbidity and mortality amongst infants of diabetic mothers admitted into Soba university hospital, Khartoum, Sudan. *Sudan. J. Paediatr.* **12**, 49–55 (2012).

55. Opara, P. I., Jaja, T. & Onubogu, U. C. Morbidity and mortality amongst infants of diabetic

mothers admitted into a special care baby unit in Port Harcourt, Nigeria. *Ital. J. Pediatr.* **36**, 77 (2010).

56. Ogata, E. S. Problems of the Infant of the Diabetic Mother. *Neoreviews* **11**, e627–e631 (2010).

CHAPTER 9: Synthesis and Conclusions

9.1 Summary of findings

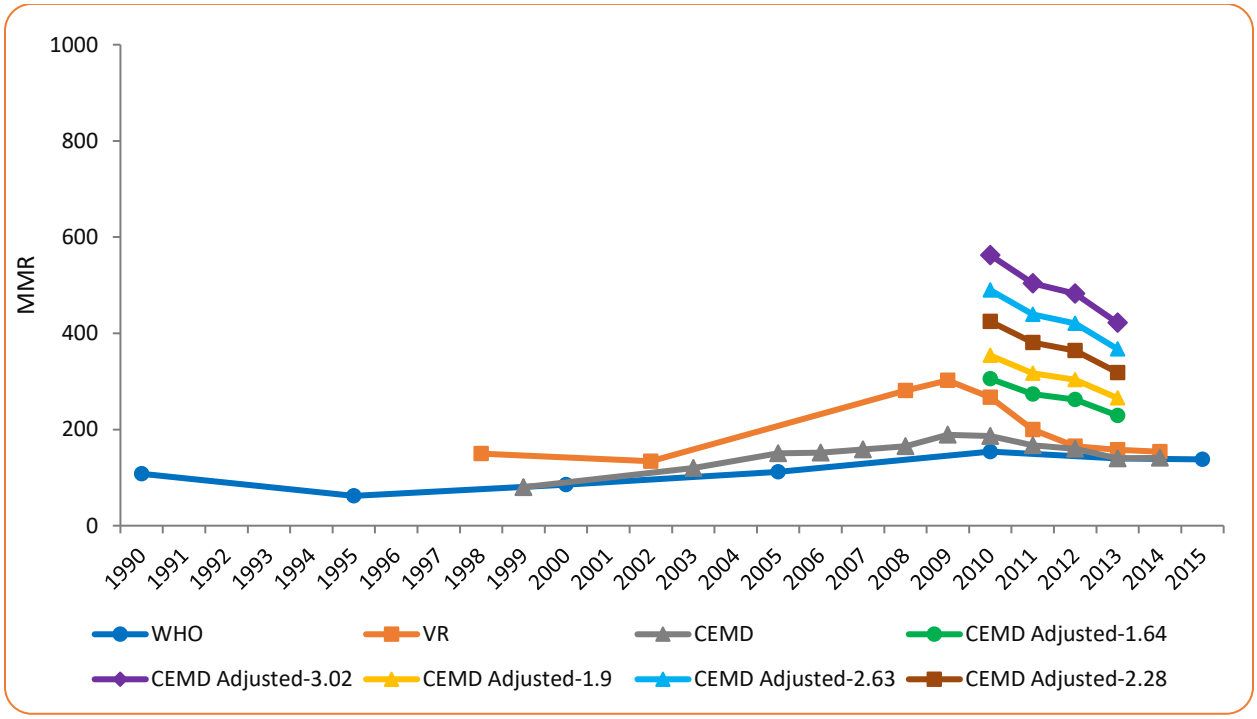
9.1.1 Level of mortality *vs.* what has been reported or estimated

Between 2010 and 2013, maternal deaths were substantially under-reported in the Western Cape province. Neonatal deaths were also under-reported in the same period, by any of the three sources evaluated, though to a lesser extent. Even if data from all sources had been combined, it is estimated that there would still have been under-reporting in maternal and neonatal mortality. Assuming the data from all sources were combined, the estimation of maternal and neonatal mortality in the Western Cape province, South Africa for the year 2010-2013 would probably require adjustment by factors of up to 1.84 and 1.13, respectively, in order to account for the under-rereporting. However, for maternal mortality, the adjustment would be higher, if one were to rely on the most used sources alone. These data suggest that the ascertainment of these outcomes using a single data source would provide biased point estimates.

The continuing use of crude maternal and neonatal mortality estimates from a single source (where MDG estimates are derived from) is subsequently misleading, and potentially of concern, given the importance of this metric in assessing the progress in global goals (i.e. SDG 3.1 & 3.2), the evaluation of health interventions focusing on reducing maternal and neonatal mortality, health planning and evidence-based decision-making.

Assuming the level of maternal mortality under-registration is the same for the whole country, as in the Western Cape, fitting the CEMD data for the years 2010-2013, using a range of inflation

factors over the four years derived from this study, resulted in higher maternal mortality estimates when compared to other sources, as presented in Figure 20 below.

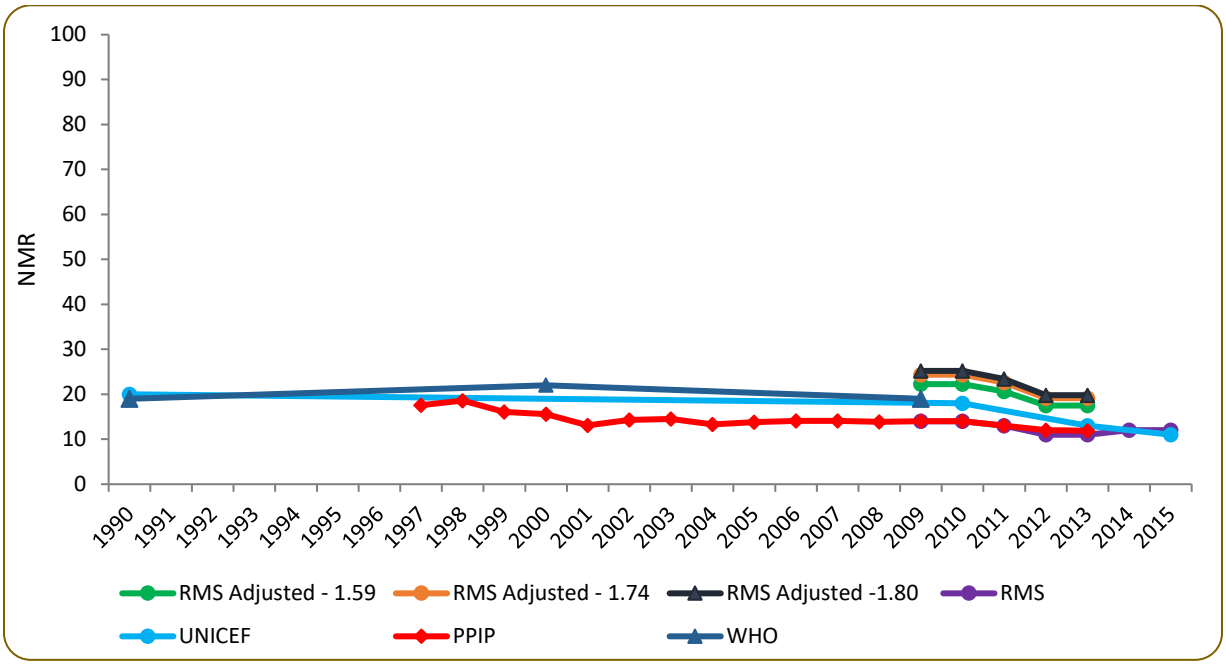


***WHO** - World Health Organisation; **VR** - Vital registration; **CEMD** - Confidential Enquiry to Maternal Deaths; **CEMD adjusted -1.64**: Assuming no missing ascertainment in CEMD (adjusted by 1.64); **CEMD adjusted -2.28**: Assuming independence (adjusted by 2.28); **CEMD adjusted -1.90**: Assuming independence and including 50% of PHDC non-overlaps (adjusted by 1.90); **CEMD adjusted -3.02**: Assuming dependence (adjusted by 3.02); **CEMD adjusted -2.63**: Assuming dependence and including 50% of PHDC non-overlaps (adjusted by 2.63).*

Figure 20: Estimates of maternal mortality from 1990-2015 in South Africa

Figure 21 presents estimates of neonatal mortality in South Africa, as contrasted with the estimates obtained from this study. Assuming the level of neonatal mortality under-registration is constant across the provinces and in the country in general, as in the Western Cape (as presented in Chapter 6, fitting the data from the Rapid mortality surveillance (RMS) report for the years 2010-2013, by using a range of adjustment factors over the four years derived from this

study, resulted in slightly higher neonatal mortality estimates than are commonly reported for the period[1–3].



WHO - World Health Organisation; PPIP - Perinatal Problem Identification Programme; RMS-Rapid Mortality Surveillance; RMS Adjusted - 1.59: Assuming no missing ascertainment; RMS Adjusted - 1.74: Assuming independence; RMS Adjusted - 1.80: Assuming dependence.

Figure 21: Estimates of neonatal mortality from 1990-2015 in South Africa

9.1.2 Trends in maternal and neonatal mortality

Estimates from global agencies and in-country institutional reporting, although widely divergent, indicate that South Africa has not achieved the MDG targets for maternal and neonatal mortality; but it has made significant progress in reducing these outcomes during the last decade. Discrepancies in data sources and quality from which these estimates were obtained, and highly variable estimates, highlight the existence of uncertainties about the true estimates of maternal and child mortality in the Western Cape. Divergent estimates of MMR and NMR by different

sources compromise the interpretation of trends over time. Based on the available data, it is difficult to estimate the trends in the CRC analysis. As confirmed by the CRC method, under-reporting in maternal and neonatal mortality in the WC leaves room for changes to be biased up or down, if these trends are based on a single source. However, the extent of decline in the context of improving reporting is likely to be biased towards the null.

9.1.3 Maternal mortality *vs.* mortality during the pregnancy period

This study defined maternal deaths, according to the 10th revision of the International Classification of Diseases and Related Health Problems (ICD-10), as "*the death of either a pregnant woman or the death of a woman within 42 days of delivery, spontaneous abortion, or the termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to, or aggravated by the pregnancy, or by its management, but not from accidental or incidental causes."* [4] It is possible that, some of the deaths in which pregnancy was incidental, or there was an ascertainment issue, might have been included as maternal deaths in the expanded dataset. However, this was not the case with institutional reporting, but the deaths enumerated through the PHDC, or from death certificates, and then linked to pregnancy episodes.

Findings from the sensitivity analyses, when assuming only a proportion of the non-overlapping deaths were true maternal mortality, resulted in decreasing maternal mortality estimates from both CRC models, i.e., independent models and a model controlling for source dependencies. Based on these results, there is potential value for additional clinical data to be used to validate all, or a sample of those patients identified as possible missed maternal deaths.

9.1.4 Factors associated with maternal and neonatal mortality

Although maternal and neonatal mortality are under-reported in the Western Cape, as indicated in Chapters 5 and 6, the associations obtained were largely consistent with previously described risk factors for both maternal and neonatal mortality rates. Additionally, the expanded dataset resulted in similar estimates of these associations, which is reassuring for ongoing studies on risk factors for maternal and neonatal mortality, if they rely just on the institutional reporting.

Based on the expanded dataset, ART status, TB and all measured chronic co-morbidities (i.e. pre-existing hypertension, diabetes and mental health condition(s)) are factors positively associated with maternal mortality. Similarly, more recent births and maternal enrolment on ART prior to pregnancy, were negatively associated with neonatal deaths; whereas infant male sex; maternal age; and mother's syphilis and diabetes were positively associated with neonatal mortality. Although birth weight was strongly correlated with neonatal mortality, it is on the causal pathway and was excluded from the models exploring the associations with antenatal risk factors.

Taken altogether, these results suggest the importance to reducing neonatal deaths, of appropriate follow-up and close monitoring of mothers and infants, where pregnancies are high risk, or the mothers have medical co-morbidities.

Even though reliance on secondary data limits the level of clinical detail available for exploration of associations with mortality; there is potential value of working with secondary data, thereby allowing for cohort analyses of associations with outcomes, rather than the analyses of case-series limited just to those pregnancies resulting in maternal or neonatal death.[5,6,15–20,7–14]

9.2 Emerging common themes

Levels of maternal and neonatal mortality

The levels of maternal and neonatal mortality in the Western Cape province are high. Even without accounting for under-ascertainment, they are still higher than those reported[7,10–12,21–26]. The recent NHI review indicated that South African officially reported maternal and neonatal mortality is high relative to other LMICs, although lower than that obtained in this study.[2,3,27] Even though the Western Cape had lower mortality estimates than other provinces, the review indicated that it is high, compared to other LMICs[3,6,7,10,11,17,27]. When compared to other LMICs, the high level of mortality might be partially ascribed to the HIV burden in the country. This is substantiated in the analysis of factors associated with maternal mortality (Chapter 6). Lower MMR in the Western Cape province compared to other provinces has been linked with the structural differences across provinces in managerial capabilities[3].

Effect of missing data on the resulting estimates

The findings indicate that there is greater variability with respect to the potential for missing data points in the estimates of ascertainment of maternal mortality, compared to neonatal mortality. The possible reasons include the nature of ascertainment of maternal mortality, requiring both death and pregnancy to be ascertained, unlike for neonatal mortality, in which only death is ascertained. Additionally, direct obstetric causes are less likely to be missed; but the indirect causes from medical conditions are more likely to be missed; and are at risk of being over attributed, in the absence of clinical review.

Trends in maternal and neonatal mortality

Based on the MDG country report, South Africa did not achieve MDG 4a and 5a targets; but it has made substantial progress in reducing maternal and neonatal mortality in the country[28]. Reduction in maternal and neonatal mortality was driven in the main by the improved access to promotive and preventive health services, along with a decline in the leading preventable causes of maternal and child mortality[28,29]. Recent estimates of maternal and neonatal mortality from both global estimates of South African maternal and neonatal mortality, and from in-country institutional reporting, have appeared to converge; and these have been decreasing over time.

A four-year three-source completeness assessment in the WC estimated lower ascertainment for maternal mortality (Overall - 54%, ranging from 35% - 79% between 2010 and 2013) compared to neonatal mortality (overall - 89%, ranging from 86%-92% between 2010 and 2013). The estimated completeness of neonatal death registration in the Western Cape remains higher than the current level of most LMICs[30]. It is possible that the increased utilisation of mortality information for research and decision-making, and the provincial mortality surveillance system, which included special training initiative in medical certification of the cause of death at a provincial level, might have contributed on the quality of medical certification. However, until the quality of medical certification has improved, and the completeness of death registration is consistently high across all sources, the provincial/national-level mortality profiles need to be interpreted with caution.

The feasibility of capture-recapture methods

As stated in Chapters 2, 5 and 6; over the last few decades, CRC has increasingly been gaining a prominent role in epidemiology and public health, particularly in the estimation of maternal and child mortality[31,32,41,33–40]. Currently, the application of CRC to estimate maternal and child mortality has expanded somewhat. In particular, a Medline search on 15th December 2021, using the following key terms: "Capture-recapture", "maternal", "neonatal", and "mortality" produced 13 articles[42,43,52–54,44–51]. Eight of these articles were after 2015[42–46,53–55]. Of these, most were from Asia[42,43,47,56], followed by Europe[44,54], North America[45,53] and South America[46]. However, there have been limited attempts to apply the CRC method in Africa in estimating maternal and neonatal mortality under-ascertainment since the inception of this study.

The application of CRC has provided valuable estimates of potential under-ascertainment of maternal and neonatal mortality in the Western Cape province for the period 2010 to 2013. However, it was not possible to fit ecological models accounting adequately for heterogeneous capture probabilities in our data; since these are only recommended when the number of data sources are sufficiently large (at least five)[37,57,58]. Even when the CRC estimates are quite widely divergent, they remain valuable in generating statistically supported estimates of under-ascertainment to caution against false certainty.

Value and challenges to the use of linked data for exploring maternal and neonatal mortality

As presented in Chapter 4, independent record linkage was performed to test the performance of the existing PHDC linkage engine, as the basis for the subsequent thesis work. The probabilistic approach used for this independent linkage proved feasible; and it might add further value to the

PHDC. Having linked clinical data; there is potential value in undertaking cohort analyses of the risk factors and not only exploring characteristics in a case-series of mothers or neonates who have died. The PHDC could also support designs where more clinical data are collected on a subset of pregnancies, which do not result in the outcome of interest. There are few comparable cohort analyses based on secondary data, with more cohort analyses from the region being from demographic surveillance sites[17–20].

9.3 Conclusions

This study has found that estimates of maternal and neonatal mortality are widely divergent across data sources and estimation methods over the MDG period. Between 2010 and 2013, these outcomes were substantially under-reported in the Western Cape, highlighting the need for the application of respective adjustment factors, in order to correct for under-reporting, or at least to appreciate the potential extent of under-ascertainment, and catalyse efforts to clinically validate additional potential cases which have been ascertained from administrative data. The findings lacked the precision to refine the estimation of trends in maternal and neonatal mortality. Analyses of antenatal risk factors for these outcomes were broadly consistent with clinical expectations; and they were not materially impacted by the inclusion or exclusion of additional data sources with additional outcome events.

The provincial linkage algorithm demonstrated similar performance to PRL in linking mortality records. This finding was reassuring with respect to the use of linked data in the WC province.

Capture-recapture methods were demonstrated to be feasible; and can be recommended as a promising approach to resolve the problems of under-ascertainment in the estimation of these outcomes in the setting.

The associations with mortality are in most cases not that different, when using the expanded dataset, which is reassuring for ongoing studies on risk factors for maternal and neonatal mortality, if they rely only on cohorts derived from institutional reporting.

9.4 Limitations

Case definitions

Maternal and neonatal mortality were defined, according to the ICD-10 classifications. However, there was uncertainty regarding the pregnancy-association of deaths, particularly those derived from the PHDC routine systems and from death certificates, as described in Chapter 2. Including pregnancy-related deaths from accidental or incidental causes is expected to have over-estimated the maternal mortality in this study, as demonstrated in the sensitivity analysis (Chapter 5, in which maternal mortality estimates decreased as the included proportion of non-overlapping PHDC-derived outcomes was reduced).Uncertainty in case definitions with administrative data further affects the assessment of potential trends.

Data fidelity

Where pregnancy might have been incidental, or there was a potential ascertainment issue, we were unable to definitively exclude incidental pregnancy in deaths from the expanded dataset, after excluding recorded non-natural deaths. Although it was reassuring that 50% of the maternal

deaths were registered by all three sources combined, sensitivity analyses were used to explore the potential impact of this, especially as the higher estimates or MMR exceeded all previous estimates.

Limited duration of the three sources

Restricted access to cause-of-death data for confidentiality purposes following the amendment to the South African Births and Deaths Registration Act in February 2014, limited DHA data linkage across the systems. As one of the implications, we limited our analysis to data from 2010 to 2013 in the capture-recapture analysis, thereby limiting the comparison period and the assessment of trends. This highlights the importance of re-enabling access to death certificates by the health department.

Limited number of variables in the analysis of risk factors for mortality

Regarding the choice of medical co-morbidities to be included in the analysis, we were limited to conditions where the PHDC had attempted to create province-wide lists of patients with these conditions. For instance, co-morbidities, such as hypertension, diabetes, epilepsy and mental health, were frequently identified from the electronic pharmacy and Chronic Disease Dispensing unit (CDU) databases, based on the history of a woman receiving medication for respective conditions. However, the available data did not distinguish gestational diabetes from pre-existing diabetes. In addition, we failed to distinguish pre-existing epilepsy from seizures resulting from eclampsia.

Consequently, we decided not to include it in our analysis. In the analysis, hypertension is considered to be pre-existing hypertension and is based on drugs used for chronic hypertension treatment. As the data quality improves, with respect to the outcomes, it is hoped that the data on risk factors will also improve.

9.4 Recommendations

- The institutional reporting focusing on rigorous verification of maternal mortality cases by clinicians remains essential and should continue to be strengthened.
- This analysis could not be completed for more recent years, since one of the data sources was no longer available. Given that there is good vital registration, advocacy for access to cause-of-death data by the department of health is therefore strongly recommended.
- Active clinical validation of administratively identified cases, which were not included in formal institutional reporting, especially for maternal deaths, would assist in estimating under-reporting more accurately in the future.
- There is a need for strengthening routine provincial surveillance systems in the province, and in the country at large. DHA and DoH should ensure comprehensive registration and wide coverage of maternal mortality data across sources.
- Capture-recapture methods should continue to be explored as epidemiological tools in surveillance contexts such as the WC province, with the caveats that an appropriate number of data sources are available, and that there is consideration of dependence between them.
- Formal consideration of under-ascertainment should be factored into country reporting.

- ❖ It would be ideal if global agencies worked closely with local researchers to agree on optimal calibration of South African estimates in multi-country models.
- ❖ Given the associations between concurrent health conditions and negative outcomes, programmes to assess and manage associated antenatal risk need to be encouraged.

9.5 References

1. Rhoda N, Velaphi S, Gebhardt GS, Kauchali S, Barron P. Reducing neonatal deaths in South Africa: Progress and challenges. *South African Med J*. 2018;108(3a):9. doi:10.7196/SAMJ.2017.v108i3b.12804

2. Damian DJ, Njau B, Lisasi E, Msuya SE, Boulle A. Trends in maternal and neonatal mortality in South Africa: a systematic review. *Syst Rev*. 2019;8(1):76. doi:10.1186/s13643-019-0991-y

3. Heever A van den. *NATIONAL HEALTH INSURANCE BILL REVIEW: Expert Review of the National Health Insurance Bill Submitted by the Minister of Health to Parliament in 2019 for Submission to Parliament as a Response to the Request for Public Comment.*; 2019.

4. WHO. *International Statistical Classification of Diseases and Related Health Problems 10th Revision*.; 2016. Accessed May 14, 2018. http://apps.who.int/classifications/icd10/browse/2016/en

5. Dorrington R, Bradshaw D, Laubscher R. *Rapid Mortality Surveillance Report 2012*.; 2014. doi:ISBN: 978-1-920618-00-1

6. Pattinson R, Rhoda N. *Saving Babies 2012-2013: Ninth Report on Perinatal Care in South Africa*.; 2014. http://www.ppip.co.za/wp-content/uploads/Saving-Babies-2012-2013.pdf

7. Dorrington R, Bradshaw D, Laubscher R, Nannan N. *Rapid Mortality Surveillance Report 2015*.; 2016. doi:ISBN: 978-1-920618-00-1

8. National perinatal morbidity and mortality committee. *National Perinatal Morbidity and Mortality Committee Report 2014*.; 2014. http://www.ppip.co.za/saving-babies/

9. NAPEMMCO. *National Perinatal Mortality and Morbidity Committee (NaPeMMCo) Triennial Report (2008-2010)*.; 2011. http://www.doh.gov.za/docs/reports/2011/perireport.pdf

10. Department of Health. *2014 Saving Mothers: Annual Report and Detailed Analysis of Maternal Deaths Due to Non-Pregnancy Related Infections*.; 2015.

11. National Committee for Confidential Enquiries into Maternal Deaths. *Saving Mothers 2011-2013: Sixth Report on Confidential Enquiries into Maternal Deaths in South Africa Short Report*.; 2014. http://www.doh.gov.za/docs/reports/2007/savingmothers.pdf

12. Pattinson RC, Fawcus S, Moodley J. *Tenth Interim Report on Confidential Enquiries into Maternal Deaths in South Africa*.; 2013.

13. National Committee for Confidential Enquiry into Maternal Deaths. *Saving Mothers 2008-2010: Fifth Comprehensive Report on Confidential Enquiries into Maternal Deaths in South Africa*. Vol 5.; 2012. doi:10.2337/dc14-S014

14. Moodley J. Saving mothers: 1999 - 2001. *SAMJ South African Med J*. 2003;93(5).

15. Moodley J. Saving mothers in South Africa. *S Afr J Obstet Gynaecol*. 2000;6(2):13-15. http://search.ebscohost.com/login.aspx?direct=true&db=awn&AN=000130-2&site=ehost-live

16. NCCEMD. *Saving Mothers 2005-2007 : Fourth Report on Confidential Enquiries into Maternal Deaths in South Africa*.; 2008. http://www.doh.gov.za/docs/reports/2007/savingmothers.pdf

17. Tlou B. Underlying determinants of maternal mortality in a rural South African population with high HIV prevalence (2000–2014): A population-based cohort analysis. Gebhardt S, ed. *PLoS One*. 2018;13(9):e0203830. doi:10.1371/journal.pone.0203830

18. Garenne M, McCaa R, Nacro K. Maternal mortality in South Africa: An update from the 2007 Community Survey. *J Popul Res*. 2011;28(1):89-101. doi:10.1007/s12546-010-9037-y

19. Bomela NJ. A cross-sectional analysis of the geographic distribution and causes of maternal mortality in South Africa: 2002-2006. *BMC Public Health*. 2015;15:273. doi:10.1186/s12889-015-1597-5

20. Bomela NJ. Maternal mortality by socio-demographic characteristics and cause of death in South Africa: 2007-2015. *BMC Public Health*. 2020;20(1):157. doi:10.1186/s12889-020-8179-x

21. WHO. *Trends in Maternal Mortality : 1990 to 2015 Estimates by WHO,UNICEF,UNFPA,The World Bank and the United Nations Population Division*.; 2015. doi:ISBN 978 92 4 156514 1

22. Statistics South Africa. *Millennium Development Goals 5: Improve Maternal Health 2015 / Statistics South Africa*.; 2015.

23. Dorrington R, Bradshaw D, Laubscher R, Nannan N. *Rapid Mortality Surveillance Report 2014*.; 2015. http://www.mrc.ac.za/bod/RapidMortalitySurveillanceReport2014.pdf

24. World Health Organization. Trends in Maternal Mortality : 1990 to 2013 Executive Summary. Published online 2014. http://apps.who.int/iris/bitstream/10665/112697/1/WHO_RHR_14.13_eng.pdf?ua=1

25. Kassebaum NJ, Bertozzi-Villa A, Coggeshall MS, et al. Global, regional, and national levels and causes of maternal mortality during 1990-2013: A systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014;384(9947). doi:10.1016/S0140-6736(14)60696-6

26. Lozano R, Wang H, Foreman KJ, et al. Progress towards Millennium Development Goals 4 and 5 on maternal and child mortality: An updated systematic analysis. *Lancet*.

2011;378(9797):1139-1165. doi:10.1016/S0140-6736(11)61337-8

27. National Department of Health (NDoH), Statistics South Africa Council (Stats SA), South African Medical Research (SAMRC), ICF. *South Africa Demographic and Health Survey 2016: Key Indicators.*; 2017.

28. Statistics South Africa. *Millennium Development Goals: Country Report 2015 / Statistics South Africa.*; 2015.

29. Chola L, Pillay Y, Barron P, Tugendhaft A, Kerber K, Hofman K. Cost and impact of scaling up interventions to save lives of mothers and children: taking South Africa closer to MDGs 4 and 5. *Glob Health Action*. 2015;8(1):27265. doi:10.3402/gha.v8.27265

30. Kasasa S, Natukwatsa D, Galiwango E, et al. Birth, stillbirth and death registration data completeness, quality and utility in population-based surveys: EN-INDEPTH study. *Popul Health Metr*. 2021;19(Suppl 1). doi:10.1186/s12963-020-00231-2

31. Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev*. 1995;17(2):243-264. Accessed May 24, 2016. http://www.ncbi.nlm.nih.gov/pubmed/8654510

32. International Working Group for Disease Monitoring and Forecasting. Capture-Recapture and Multiple-Record Systems Estimation II: Applications in Human Diseases. *Am J Epidemiol*. 1995;142(10):1059-1068. Accessed May 17, 2015. http://aje.oxfordjournals.org.ezproxy.uct.ac.za/content/142/10/1059.short

33. Héraud-Bousquet V, Lot F, Esvan M, et al. A three-source capture-recapture estimate of the number of new HIV diagnoses in children in France from 2003-2006 with multiple imputation of a variable of heterogeneous catchability. *BMC Infect Dis*. 2012;12:251. doi:10.1186/1471-2334-12-251

34. Hay G. The selection from multiple data sources in epidemiological capture-recapture studies. *J R Stat Soc Ser D (The Stat*. 1997;46(4):515-520. doi:10.1111/1467-9884.00105

35. McCarty DJ, Tull ES, Moy CS, Kwoh CK, LaPorte RE. Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol*. 1993;22(3):559-565. Accessed May 17, 2015. http://www.ncbi.nlm.nih.gov/pubmed/8359975

36. Aghaei A, Ahmadi-Jouibari T, Baiki O, Mosavi-Jarrahi A. Estimation of the gastric cancer incidence in Tehran by two- source capture-recapture. *Asian Pac J Cancer Prev*. 2013;14(2):673-677.

37. Chao A, Tsay PK, Lin SH, Shau WY, Chao DY. The applications of capture-recapture models to epidemiological data. *Stat Med*. 2001;20(20):3123-3157. Accessed August 14, 2016. http://www.ncbi.nlm.nih.gov/pubmed/11590637

38. Abubakar I. Assessing the coverage of TB notification using capture recapture methods.

39. Coeli CM, Veras RP, da Silva Freire Coutinho E. [Capture-recapture methodology: an option for surveillance of non-communicable diseases in the elderly]. *Cad Saude Publica*. 16(4):1071-1082. Accessed May 17, 2015. http://www.ncbi.nlm.nih.gov/pubmed/11175530

40. Domingo-Salvany a, Hartnoll RL, Maguire a, et al. Analytical considerations in the use of capture-recapture to estimate prevalence: case studies of the estimation of opiate use in the metropolitan area of Barcelona, Spain. *Am J Epidemiol*. 1998;148(8):732-740.

41. Qomariyah SN, Braunholtz D, Achadi EL, et al. An option for measuring maternal mortality in developing countries: a survey using community informants. *BMC Pregnancy Childbirth*. 2010;10(1):74. doi:10.1186/1471-2393-10-74

42. Mir AM, Shaikh MS, Qomariyah SN, Rashida G, Khan M, Masood I. Using community informants to estimate maternal mortality in a rural district in Pakistan: A feasibility study. *J Pregnancy*. 2015;2015. doi:10.1155/2015/267923

43. Qomariyah SN, Sethi R, Izati YN, et al. No one data source captures all: A nested case-control study of the completeness of maternal death reporting in Banten Province, Indonesia. *PLoS One*. 2020;15(5). doi:10.1371/journal.pone.0232080

44. Shai S, Perez-Becker R, Andres O, et al. Incidence of SCID in Germany from 2014 to 2015 an ESPED* Survey on Behalf of the API*** Erhebungseinheit für Seltene Pädiatrische Erkrankungen in Deutschland (German Paediatric Surveillance Unit) ** Arbeitsgemeinschaft Pädiatrische Immunologie. *J Clin Immunol*. 2020;40(5):708-717. doi:10.1007/s10875-020-00782-x

45. Ávila J, Tavera M, Carrasco M. Epidemiological characteristics of neonatal mortality in Peru, 2011-2012. *Rev Peru Med Exp Salud Publica*. 2015;32(3):423-430. doi:10.17843/rpmesp.2015.323.1670

46. Boyd AT, Hulland EN, Grand'Pierre R, et al. Use of Rapid Ascertainment Process for Institutional Deaths (RAPID) to identify pregnancy-related deaths in tertiary-care obstetric hospitals in three departments in Haiti. *BMC Pregnancy Childbirth*. 2017;17(1). doi:10.1186/s12884-017-1329-1

47. Anwar J, Torvaldsen S, Sheikh M, Taylor R. Completeness of a Maternal and Perinatal Mortality Enhanced Surveillance System in Pakistan: Evidence from Capture–Recapture Methods. *Matern Child Health J*. 2018;22(12):1743-1750. doi:10.1007/s10995-018-2571-5

48. Eisele TP, Lindblade KA, Rosen DH, Odhiambo F, Vulule JM, Slutsker L. Evaluating the completeness of demographic surveillance of children less than five years old in western Kenya: a capture-recapture approach. *Am J Trop Med Hyg*. 2003;69(1):92-97. doi:10.4269/ajtmh.2003.69.92

49. Checchi F, Nyasulu P, Chandramohan D, Roberts B. Rates and causes of death in Chiradzulu District, Malawi, 2008: A key informant study. *Trop Med Int Heal*. 2011;16(3):375-378. doi:10.1111/j.1365-3156.2010.02702.x

50. Aslan D, Özcebe H, Bertan M, Karaaoaoolu E. Capture-recapture methods for estimation of fertility and mortality in a rural district of Turkey. *East Mediterr Heal J*. 2004;10(1-2):56-63. doi:10.26719/2004.10.1-2.56

51. Ronsmans C, Scott S, Qomariyah SN, et al. Professional assistance during birth and maternal mortality in two Indonesian districts. *Bull World Health Organ*. 2009;87(6):416-423. doi:10.2471/BLT.08.051581

52. Yazdizadeh B, Mohammad K, Nedjat S, et al. Application of capture-recapture for fine-tuning uncertainties about national maternal mortality estimates. *Int J Prev Med*. 2014;5(5):624-631. Accessed February 11, 2015. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4050684&tool=pmcentrez&rendertype=abstract

53. Matthias J, du Bernard S, Schillinger JA, Hong J, Pearson V, Peterman TA. Estimating Neonatal Herpes Simplex Virus Incidence and Mortality Using Capture-recapture, Florida. *Clin Infect Dis*. 2021;73(3):506-512. doi:10.1093/cid/ciaa727

54. Phillips A, Beard F, Macartney K, et al. Vaccine-preventable child deaths in New South Wales from 2005 to 2014: How much is preventable? *J Paediatr Child Health*. 2018;54(4):356-364. doi:10.1111/jpc.13835

55. Anwar J, Torvaldsen S, Sheikh M, Taylor R. Completeness of a Maternal and Perinatal Mortality Enhanced Surveillance System in Pakistan: Evidence from Capture–Recapture Methods. *Matern Child Health J*. 2018;22(12):1743-1750. doi:10.1007/s10995-018-2571-5

56. Mony PK, Varghese B, Thomas T. Estimation of perinatal mortality rate for institutional births in Rajasthan state, India, using capture-recapture technique. *BMJ Open*. 2015;5(3):e005966. doi:10.1136/bmjopen-2014-005966

57. Amstrup SC, McDonald TL, Manly BFJ. *Handbook of Capture-Recapture Analysis*. Princeton University Press; 2005.

58. Chao A, Tsay PK, Shau WY, Chao D. Population size estimation for capture-recapture models with applications to epidemiological data. *Proc Biometrics Sect Am Stat Assoc*. Published online 1996:108-117. doi:10.1080/02664760120011572

APPENDICES

Appendix 1: Records of literature review search strategy

Database	Search terms	Retrieved
PubMed	(("mothers"[MeSH Terms] OR "mothers"[All Fields] OR "maternal"[All Fields]) OR ("infant, newborn"[MeSH Terms] OR ("infant"[All Fields] AND "newborn"[All Fields]) OR "newborn infant"[All Fields] OR "neonatal"[All Fields])) AND (("mortality"[Subheading] OR "mortality"[All Fields] OR "mortality"[MeSH Terms]) OR ("death"[MeSH Terms] OR "death"[All Fields])) AND (estimation[All Fields] OR estimates[All Fields]) AND ("South Africa"[Mesh] OR ("south africa"[MeSH Terms] OR ("south"[All Fields] AND "africa"[All Fields]) OR "south africa"[All Fields])) AND (("1990/01/01"[PDAT] : "2015/12/31"[PDAT]) AND "humans"[MeSH Terms] AND English[lang])	140
	Limits: English language Year: 2005-2015 Study on humans	
Africa-wide Information	(Maternal OR neonatal) AND (mortality OR death) AND estimat* AND ('developing countries' OR Africa)	355
	Limits: English language Year: 2005-2015	
Web of Science	(Maternal OR neonatal) AND (mortality OR death) AND estimat* AND ('developing countries' OR Africa)	264
	Limits: English language Year: 2005-2015	
Scopus	(Maternal OR neonatal) AND (mortality OR death) AND estimat* AND ('developing countries' OR 'Sub Saharan Africa' OR Africa)	189
	Limits:	

	English language Year: 2005-2015	
Extra	**Reference collected from bibliography of key articles**	**81**
Total	**Total unique reference used**	**323**

Appendix 2. PHDC linking criteria

Criteria ID	Match criteria name	Match criteria description	Match Score
1	Exact (folder, fullname, dob)	folder_number = folder_number AND fullname = firstnames + surname AND date_of_birth = date_of_birth	9.9
2	Exact (folder, surname, dob)	folder_number=folder_number AND surname=surname AND date_of_birth=date_of_birth	10.4
3	Exact (internalpatientnumber)	internal_patient_number = internal_patient_number	10
4	Highly Possible (dob, firstnames, ~surname (incl sa_id check))	date_of_birth=date_of_birth AND JW (surname)>=@surnamescore AND firstnames =firstnames AND (sa_id=sa_id if not null)	7.8
5	Highly Possible (dob, surname, firstnames (excl sa_id check))	date_of_birth = date_of_birth AND surname = surname AND firstnames = firstnames (no SA ID check)	7.4
6	Highly Possible (firstnames, ~surname, ~dob (incl sa_id check))	Firstname = firstname AND JW (surname)>=@surnamescore AND YOB=YOB AND (sa_id=sa_id if not null) AND (JW(DOB)>=@dateofbirthscore and ((MOB=MOB) or (MOB=DOB and DOB=MOB))	7.5
7	Highly Possible (surname, ~firstnames, ~dob (incl sa_id check))	surname = surname AND JW (firstnames)>=@firstnamescore AND YOB=YOB AND (sa_id=sa_id if not null) AND (JW(DOB)>=@dateofbirthscore and ((MOB=MOB) or (MOB=DOB and DOB=MOB))	7.5
8	Highly Possible (surname, dob, ~firstnames (incl sa_id check))	date_of_birth = date_of_birth AND surname = surname AND JW (firstnames) >= @firstnamescore AND (sa_id=sa_id if not null)	7.8
9	Highly Possible (surname, firstnames ~dob (incl sa_id check))	surname = surname AND firstnames =firstnames AND YOB=YOB AND (sa_id=sa_id if not null) AND [JW (date_of_birth)>=@DOBscore OR ((MOB=MOB) or (MOB=DOB and DOB=MOB))]	7.7
10	Highly Probable (folder, ~dob, ~fullname)	folder_number = folder_number AND JW (fullname,firstnames+surname)>=@firstnamesscore AND JW(date_of_birth,date_of_birth)>=@dateofbirthscore	9.4
11	Highly Probable (folder, ~surname,~dob)	folder_number = folder_number AND JW (surname) >= @surnamescore AND [(YOB = YOB AND DayOB = DayOB) OR (MOB = MOB AND DayOB = DayOB) OR (YOB = YOB AND MOB = MOB)]	10.2
12	Highly Probable (folder, dob, ~fullname)	folder_number = folder_number AND JW (fullname,firstnames+surname)>=@firstnamesscore AND date_of_birth = date_of_birth	9.7

13	Highly Probable (folder, firstnames, surname)	folder_number = folder_number AND firstnames = firstnames and surname = surname	9.8
14	Highly Probable (folder, fullname)	folder_number = folder_number AND fullname = firstnames + surname	9.8
15	Highly Probable (folder, surname, ~firstnames)	folder_number = folder_number AND JW (firstnames)>=@firstnamesscore and surname = surname	9.7
16	Highly Probable (sa_id, ~dob)	sa_id=sa_id AND JW(DOB)>@dateofbirthscore	9.8
17	Highly Probable (sa_id, ~firstnames)	sa_id=sa_id AND JW (firstname)>@firstnamescore	9.8
18	Highly Probable (sa_id, ~surname)	sa_id=sa_id AND JW (surname)>@surnamescore	9.8
19	Highly Probable (sa_id, ~surname,~dob)	sa_id <> '' AND (both sa_id AND sa_id is valid AND sa_id = sa_id) AND JW (surname) >= @surnamescore AND JW (date_of_birth) >= @dateofbirthscore	10.3
20	Highly Probable (sa_id, ~switch (surname, firstnames))	sa_id = sa_id AND JW (surname,firstname) >=@surnamescore and JW(firstname, surname)>=@surnamescore	8.9
21	Highly Probable (sa_id, surname, firstname)	sa_id = sa_id and surname=surname and firstname=firstname	10.6
22	Highly Probable (sa_id, switch(surname,firstnames))	sa_id=sa_id AND surname = firstname and firstname = surname	10.5
23	Low Possible (~ (surname, firstnames))	JW (surname) >= @surnamescore AND JW (firstnames) >= @firstnamesscore	2
24	Low Possible (~fullname)	fullname = firstnames + surname	4.2
25	Low Possible (surname, ~(firstnames))	surname = surname AND JW (firstnames) >= @firstnamesscore	3
26	Low Possible (surname, firstnames)	surname = surname AND firstnames = firstnames	4.2
27	Possible (~address, ~surname, ~firstname, ~dob)	AddressLine1 <> '' AND JW(addressline1, AddressLine2) >= @addresslinescore AND JW(surname1, Surname2) >= @surnamescore AND JW(forenames1 , Forenames2) >= @forenamesscore AND DATEDIFF(YY,@dateofbirth,DateOfBirth) BETWEEN -15 AND 15) OR JW(surname1, Surname2) >= @surnamescore AND JW (DOB1, DOB2) >= @dateofbirthscore OR JW(forenames1 , Forenames2) >= @forenamesscore AND JW (DOB1, DOB2) >= @dateofbirthscore	5.5

28	Possible (dob, firstnames, ~surname (excl sa_id check))	date_of_birth=date_of_birth AND JW (surname)>=@surnamescore AND firstnames =firstnames (no SA ID check)	7.2
29	Possible (dob, surname, ~firstnames (excl sa_id check))	date_of_birth = date_of_birth AND surname = surname AND JW (firstnames) >= @firstnamescore (no SA ID check)	7.3
30	Possible (firstnames, ~surname, ~dob, (incl sa_id check))	Firstname = firstname AND JW (surname)>=@surnamescore AND YOB=YOB AND (sa_id=sa_id if not null) AND (JW(DOB)>=@dateofbirthscore and ((MOB=MOB) or (MOB=DOB and DOB=MOB))	5.7
31	Possible (sex,~surname, ~firstnames, ~dob)	JW (surname) >= @surnamescore AND JW (firstnames) >= @firstnamescore AND JW(DOB) >= @dateofbirthscore) OR (DOB1 = DOB2 AND DIFFERENCE (surname1, Surname2) > 3) AND DIFFERENCE (Forenames1, Forenames2) > 2 AND (Gender1= Gender2 OR Postalcode1 = PostalCode2)	5.4
32	Possible (soundex(surname,firstnames), dob)	Soundex (Surname) + Soundex (Firstname) + date_of_birth	4.1
33	Possible (surname, ~firstnames, ~dob (excl sa_id check))	surname = surname AND JW (firstnames)>=@firstnamescore AND (JW(DOB)>=@dateofbirthscore and (YOB=YOB AND [(MOB=MOB) or (MOB=DOB and DOB=MOB)])	5.8
34	Possible (surname, firstnames, ~dob, (excl sa_id check))	surname = surname AND firstnames =firstnames AND [JW (date_of_birth)>=@DOBscore OR (YOB=YOB AND [(MOB=MOB) or (MOB=DOB and DOB=MOB)])]	5.9
35	Probable (dob, firstnames, surname, ~address, (incl sa_id check))	date_of_birth=date_of_birth AND firstnames=firstnames AND JW (surname)>=@surnamescore AND (sa_id=sa_id if not null) AND JW (address)>=@addressscore	7.5
36	Probable (firstnames, surname, dob (incl sa_id check))	date_of_birth = date_of_birth AND surname = surname AND firstnames = firstnames AND (sa_id=sa_id if not null)	7.9
37	Probable (folder, ~firstnames, ~surname)	folder_number = folder_number AND JW (firstnames)>=@firstnamesscore and JW (surname)>=@surnamescore	9.5
38	Probable (folder, ~fullname)	folder_number = folder_number AND JW (fullname,firstnames + surname)>=@firstnames_score	9.7
39	Probable (folder, ~surname)	folder_number = folder_number AND fullname like '%'+ surname +'%'	9.2
40	Probable (folder, dob)	folder_number = folder_number AND dateofbirth=dateofbirth	8.8

41	Probable (folder, firstnames)	folder_number = folder_number AND firstname = firstname	8.5
42	Probable (folder, surname)	folder_number = folder_number AND surname = surname	8.7
43	Probable (fullname, dob)	Fullname = firstnames + surname AND date_of_birth=date_of_birth	7.4
44	Probable (multiline match, firstname, surname, dob)	firstname, surname and dob match on different lines with the same folder number	7.5
45	Probable (surname, dob, ~firstnames,~address (incl sa_id check))	date_of_birth=date_of_birth AND surname=surnames AND JW (firstnames) >=@firstnamescore AND (sa_id=sa_id if not null) AND JW (address)>=@addressscore	7.5
46	Probable (surname, firstnames, ~address, ~dob (incl sa_id check))	surname=surname AND firstnames=firstnames AND (sa_id=sa_id if not null) AND JW(address)>=@addressscore AND [JW(date_of_birth)>=@DOBscore OR (YOB=YOB AND [(MOB=MOB) or (MOB=DOB and DOB=MOB)])]	8.2
47	Probable (switch (surname, firstnames), dob)	Firstname = surname and surname = firstname and date_of_birth = date_of_birth	7.9

Appendix 3: Closed population capture-recapture log-linear models for maternal deaths fitted with STATA (2010 - 2013)

Model	DoF	G^2	*p*-value	AIC	BIC	$\widehat{X}$	$\widehat{N}$	95% CI
M_{000}	3	30.0	<0.001	24.0	24.2	152	543	501, 597
M_{100}	4	24.3	<0.001	20.3	20.4	129	520	480, 572
M_{020}	4	4.6	0.101	0.6	0.7	328	719	607, 885
M_{003}	4	27.6	<0.001	23.6	23.8	134	525	483, 581

Key:

DoF: Number of degrees of freedom

G^2: Deviance statistic

$\widehat{X}$: Estimate of the number of deaths not reported to any source

$\widehat{N}$: Estimate of the number of deaths

95% CI: 95% confidence interval for $\widehat{N}$;

Sources: **S_1** – PHDC; **S_2** – DHA; **S_3** – CEMD

M_{000}: Independent model i.e., model with no interaction

M_{100}: Model with one-way interaction i.e., $S_2 * S_3$

M_{020}: Model with one-way interaction i.e., $S_1 * S_3$

M_{003}: Model with one-way interaction i.e., $S_1 * S_2$

Appendix 4: Closed population capture-recapture log-linear models for neonatal deaths fitted with STATA (2010 - 2013)

Model	DoF	G^2	*p*-value	AIC	BIC	$\widehat{X}$	$\widehat{N}$	95% CI
M_{000}	3	561.0	<0.001	555.0	555.2	838	9,819	9743, 9901
M_{100}	4	69.0	<0.001	65.0	65.1	1,165	10,146	10045, 10254
M_{020}	4	485.3	<0.001	481.3	481.4	407	9,388	9304, 9485
M_{003}	4	557.0	<0.001	553.0	553.2	857	9,838	9759, 9924

Key:

DoF: Number of degrees of freedom

G^2: Deviance statistic

$\widehat{X}$: Estimate of the number of deaths not reported to any source

$\widehat{N}$: Estimate of the number of deaths

95% CI: 95% confidence interval for $\widehat{N}$;

Sources: **S_1** – PHDC; **S_2** – PPIP; **S_3** – DHA.

M_{000}: Independent model i.e., model with no interaction

M_{100}: Model with one-way interaction i.e., $S_2 * S_3$

M_{020}: Model with one-way interaction i.e., $S_1 * S_3$

M_{003}: Model with one-way interaction i.e., $S_1 * S_2$