



THE UNIVERSITY OF QUEENSLAND  
AUSTRALIA

**Availability and quality  
of vital registration mortality data in South Africa:  
A national and surveillance site perspective**

Janetta Debora Joubert

B.A., B.A. Hons, M.A., B.A. Hons (Demography), M.A. (Demography)

A thesis submitted for the degree of Doctor of Philosophy at  
The University of Queensland in 2014  
School of Population Health

## **Abstract**

The importance of valid and reliable mortality data as a key input for public and population health planning and research, and development programmes, is globally acknowledged. However, few low- and middle-income countries have such data. Even fewer have assessed the quality of their mortality data. Despite the particular utility of periodic, independent evaluations of national mortality statistics by countries themselves, such an evaluation has not yet been conducted for South Africa. Neither has a review been conducted of the sources of mortality data in South Africa since 1995. Regarding causes of death, a few small-scale studies assessed data quality in urban settings, but no study yet in civil registration mortality data in a rural setting. In addition, notwithstanding district health managers' critical need to know the health care needs of their district population current district mortality profiles are accompanied by strong warnings of invalid data.

These gaps in essential public- and population-health intelligence present important reasons for research on the availability and quality of mortality data in South Africa.

The primary aim of this body of research was to assess the quality of mortality data in the national CRVS system of South Africa. As alternative mortality data sources can complement or evaluate the quality of CRVS data, a secondary aim was to identify and review alternative mortality data sources, and assess the availability, strengths, and limitations of their data. A further aim was to conduct a field study to link and compare CRVS and Agincourt HDSS mortality data for the same individuals with the purpose of quantifying the level of completeness of death registration into the CRVS system, and assess the level of agreement between CRVS and verbal autopsy cause-of-death data. Seven research questions have been addressed towards achieving these aims.

Different approaches were followed, including: (a) a mortality data source review and appraisal of the availability, strengths and limitations of different sources; (b) a comprehensive evaluation of national CRVS mortality data based on demographic and epidemiological principles; (c) a detailed record-linkage study applying both deterministic and probabilistic approaches; and (d) an application of adjustment factors from the linkage study to improve district-level cause-of-death data quality.

The data source review found a rich and varied list of mortality data sources. Recent transformation in South Africa has enhanced mortality data availability from CRVS. Detailed cause-of-death data are available from civil registration and demographic surveillance. Additionally, a range of mortality data items are available in three population censuses, three demographic surveillance systems, and various national surveys, mortality audits and disease-notification programmes. Employing nine criteria to evaluate the quality of national CRVS mortality data, the data were rated satisfactory for coverage and completeness of death registration, temporal consistency, age/sex classification, timeliness and sub-national availability. Epidemiological consistency could not be assessed conclusively as the model lacks the discriminatory power to enable an assessment for South Africa. Selected studies and the extent of ill-defined/non-specific codes suggest shortcomings with detailed-cause data, and this criterion and content validity were rated unsatisfactory.

Record linkage was achieved on 2,264 records from the Agincourt HDSS accounting for 61% of deaths. The study was unable to estimate completeness of registration due to differing boundaries in the two systems. However, it revealed that records for more vulnerable people, including young children, poorer persons, and non-South Africans were less likely to be matched. An identity number increased the odds of being matched almost 14 times. Cause-of-death agreement was very low at 15% (kappa 0.1083, CI: 0.0995–0.1171) for the WHO Special Tabulation list with 103 causes of death, and 23% (kappa 0.1631, CI: 0.1511–0.1751) for the short list of 15 causes. Applying adjustment factors derived from the linkage study to the surrounding Ehlanzeni district data, caused substantial changes in the district cause-of-death profile. The reasonable agreement between the adjusted Ehlanzeni profile and the overall Mpumalanga province profile suggests that the linkage study has been useful for developing adjustment factors for cause-of-death attribution at district level.

This work reveals considerable progress from a dysfunctional mortality data system to one that offers data that can be adjusted for known biases. However, a confidence gap remains for unadjusted detailed-cause mortality data to inform health planning and resource prioritisation, and recommendations are made for improved data quality.

## **Declaration by author**

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

## **Publications during candidature**

### ***Peer-reviewed journal articles***

**Joubert JD**, Bradshaw D, Kabudula C, Rao C, Kahn K, Mee P, Tollman S, Lopez AD, Vos T. Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006–09. *Int J Epidemiol.* 2014.. Epub 21 Aug 2014; doi:10.1093/ije/dyu156.

Kabudula CW, **Joubert JD**, Tuoane-Nkhasi M, Kahn K, Rao C, Gómez-Olivé FX, Mee P, Tollman S, Lopez AD, Vos T, Bradshaw D. Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa. *Popul Health Metr.* 2014; 12(23).

**Joubert J**, Rao C, Bradshaw D, Vos T, Lopez AD. Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007. *PLoS ONE.* 2013; 8(5): e64592. doi:10.1371/journal.pone.0064592.

**Joubert J**, Rao C, Bradshaw D, Dorrington RE, Vos T, Lopez AD. Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa. *Glob Health Action.* 2012; 5(19263).

### ***Book chapter***

Hale W, **Joubert J**, Kalula S. Aging populations and chronic illness. In: Markle W, Fisher M, Smego R, editors. *Understanding Global Health*, 2nd edition. New York and others: McGraw-Hill Education; 2013. p. 363-86. ISBN: 978-0-07-179100-7

### ***Peer-reviewed technical reports***

Msemburi W, Pillay-van Wyk V, Dorrington RE, Neethling I, Nannan N, Groenewald P, Laubscher R, **Joubert J**, Matzopoulos R, Nicol E, Nojilana B, Prinsloo M, Sithole N, Somdyala N, Bradshaw D. Second national burden of disease study for South Africa:

Cause-of-death profile for South Africa, 1997–2010. Cape Town: South African Medical Research Council, 2014. ISBN: 978-1-920618-34-6. Available:

<http://www.mrc.ac.za/bod/SouthAfrica2010.pdf>

Msemburi W, Pillay-van Wyk V, Dorrington RE, Neethling I, Nannan N, Groenewald P, Laubscher R, **Joubert J**, Matzopoulos R, Nicol E, Nojilana B, Prinsloo M, Sithole N, Somdyala N, Bradshaw D. Second National Burden of Disease Study for South Africa: Cause-of-death profile for Eastern Cape, 1997–2010. Cape Town: South African Medical Research Council, 2014. ISBN: 978-1-920618-36-0. Available:

<http://www.mrc.ac.za/bod/EasternCape2010.pdf>

Msemburi W, Pillay-van Wyk V, Dorrington RE, Neethling I, Nannan N, Groenewald P, Laubscher R, **Joubert J**, Matzopoulos R, Nicol E, Nojilana B, Prinsloo M, Sithole N, Somdyala N, Bradshaw D. Second national burden of disease study for South Africa: Cause-of-death profile for Western Cape, 1997–2010. Cape Town: South African Medical Research Council, 2014. ISBN: 978-1-920618-35-3. Available:

<http://www.mrc.ac.za/bod/WesternCape2010.pdf>

Seven more provincial technical reports, available at:

<http://www.mrc.ac.za/bod/FreeState2010.pdf>

<http://www.mrc.ac.za/bod/Gauteng2010.pdf>

<http://www.mrc.ac.za/bod/KwaZuluNatal2010.pdf>

<http://www.mrc.ac.za/bod/Limpopo2010.pdf>

<http://www.mrc.ac.za/bod/Mpumalanga2010.pdf>

<http://www.mrc.ac.za/bod/NorthernCape2010.pdf>

<http://www.mrc.ac.za/bod/NorthWest2010.pdf>

Victoria Pillay-Van Wyk, Ria Laubscher, William Msemburi, Rob E. Dorrington, Pam Groenewald, Theo Vos, Richard Matzopoulos, Megan Prinsloo, Beatrice Nojilana, Nadine Nannan, Nontuthuzelo Somdyala, Nomfuneko Sithole, Ian Neethling, Edward Nicol, Anastasia Rossouw, **Jané D. Joubert**, Debbie Bradshaw. *Second South African National Burden of Disease Study: Data cleaning, validation and SA NBD List*. Cape Town: South African Medical Research Council, 2014. ISBN: 978-1-920618-17-9.

Available: <http://www.mrc.ac.za/bod/reports.htm>

***Published systematic review proposal***

Nglazi M, Muhwava M, **Joubert JD**, Pillay-van Wyk V, Bradshaw D. Major depressive disorder in South Africa (1997 - 2014): a systematic review. PROSPERO 2015: CRD42015024885. Available:

[http://www.crd.york.ac.uk/PROSPERO/display\\_record.asp?ID=CRD42015024885](http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015024885)

### ***Peer-reviewed published abstract***

Pillay-van Wyk V, Msemburi W, Laubscher R, Dorrington R E, Groenewald P, Matzopoulos R, Prinsloo M, Nojilana B, Nannan N, Gwebushe N, Vos T, Somdyala N, Sithole N, Neethling I, Nicol E, **Joubert J**, Rossouw A, Bradshaw D. Second National Burden of Disease Study South Africa: national and subnational mortality trends, 1997–2009. *The Lancet*, 381 (Supplement 2), 17–19 June 2013: S113.

### ***Conference, technical meeting, and seminar presentations***

**Joubert JD**, Bradshaw D, Vos T, Rao C, Lopez AD, Kabudula CW, Hanmer, L. *Progress in strengthening cause-of-death statistics in South Africa*. Poster and invited oral presentation at the WHO-FIC Network Annual Meeting and Conference, 17-23 Oct. 2015, Manchester, England. Poster in published Conference Proceedings book.

Nglazi M, **Joubert JD**, Pillay-van Wyk V, Bradshaw D. 2015. *What is the magnitude of major depressive disorder in South Africa for the period 1997 to 2014?* Poster presented at the 11<sup>th</sup> Public Health Association of South Africa Conference, 7–9 October 2015, Durban, South Africa. Awarded runner-up prize for best poster presentation.

Pillay-van Wyk V, Gbabe OF, Roomaney R, Nglazi M, **Joubert JD**, Bradshaw D. *Risk of bias tool for systematic review of observational studies*. Poster presented at the 2015 Society for Epidemiologic Research 2015 Annual Meeting, June 2015, Denver, Colorado, USA. Available: [http://epiresearch.org/poster/view\\_abstract.php?id=28237](http://epiresearch.org/poster/view_abstract.php?id=28237) (Accessed: 20 Sep. 2015).

Kabudula CW, **Joubert JD**, Tuoane-Nkhasi M, Kahn K, Rao C, Gómez-Olivé FX, Mee P, Tollman S, Lopez AD, Vos T, Bradshaw D. The potential of HDSS and record linkage to estimate completeness of death registration in civil registration system in South Africa. Improving Mortality Statistics as part of Strengthening Civil Registration and Vital Statistics Systems: Technical meeting of World Health Organization Reference Group on Health Statistics; Chateau de Penthes, Geneva, Switzerland; 4-5 Nov 2014.

**Joubert JD**. Evaluating the quality of South African mortality statistics from civil registration. Invited seminar to the Department of Actuarial Sciences Seminar Series. University of Cape Town; 27 Mar 2014.

Kabudula CW, **Joubert JD**, Tuoane-Nkhasi M, Kahn K, Bradshaw D. *Linking demographic surveillance mortality data with national vital registration system data: Lessons learned from Agincourt, South Africa*. The INDEPTH Scientific Conference 2013; 28 – 31 October 2013; School of Public Health, University of Witwatersrand, Johannesburg, South Africa.

Pillay-van Wyk V, Msemburi W, Laubscher R, Dorrington RE, Groenewald P, Matzopoulos R, Prinsloo M, Nojilana B, Nannan N, Gwebushe N, Vos T, Somdyala N, Sithole N, Neethling I, Nicol E, **Joubert JD**, Rossouw A, Bradshaw D. *Second National Burden of Disease Study South Africa: national and subnational mortality trends, 1997–2009*. IHME Conference 2013; IHME, University of Washington, Seattle, USA; June 2013.

**Joubert JD**, Bradshaw D, Dorrington R. *Population ageing and older persons' health in South Africa*. XXVI International Population Conference of the International Union for the Scientific Study of Population (IUSSP), Marrakech, Morocco; 27 Sep–2 Oct 2009.



## **Publications included in this thesis**

**Incorporated in Chapter 5: Joubert JD**, Bradshaw D, Kabudula C, Rao C, Kahn K, Mee P, Tollman S, Lopez AD, Vos T. Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006–09. *International Journal of Epidemiology*. 2014; Epub 21 Aug 2014; doi:10.1093/ije/dyu156.

<b>Contributor</b>	<b>Statement of contribution</b>	
Jané Joubert (Candidate)	Conceived and conceptualized project:	20
	Conceived and conceptualized paper:	80
	Searched, sourced, and reviewed literature:	100
	Performed statistical analysis:	90
	Interpreted results:	75
	Constructed tables and figures:	100
	Wrote first draft of manuscript:	100
	Led submission process:	100%
	Revised paper according to external reviewers' comments:	90%
Debbie Bradshaw	Conceived project and conceptualized project:	20%
	Conceived and conceptualized paper:	5%
	Performed statistical analysis:	5%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	20%
	Revised paper according to external reviewers' comments:	5%
Chodziwadziwa Kabudula	Interpreted results:	5%
	Performed statistical analysis:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
	Interpreted results:	5%
	Revised paper according to external reviewers' comments:	5%
Chalapati Rao	Conceived project and conceptualized project	10%
	Conceived and conceptualized paper:	5%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
Kathy Kahn	Critically reviewed and intellectually strengthened drafts:	5%
	Revised paper according to external reviewers' comments:	5%
Paul Mee	Critically reviewed and intellectually strengthened drafts:	5%
	Revised paper according to external reviewers' comments:	5%
Stephen Tollman	Critically reviewed and intellectually strengthened drafts:	5%
	Revised paper according to external reviewers' comments:	5%
Alan Lopez	Conceived project and conceptualized project:	10%
	Conceived and conceptualized paper:	5%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
Theo Vos	Conceived project and conceptualized project:	40%
	Conceived and conceptualized paper:	5%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	20%

**Incorporated in Chapter 4:** Kabudula CW, Joubert JD, Tuoane-Nkhasi M, Kahn K, Rao C, Gómez-Olivé FX, Mee P, Tollman S, Lopez AD, Vos T, Bradshaw D. Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa. *Population Health Metrics*. 2014; 12(23).

<b>Contributor</b>	<b>Statement of contribution</b>	
Chodziwadziwa Kabudula (First author)	Conceived and conceptualized paper:	10%
	Wrote research proposal:	2%
	Searched, sourced, and reviewed literature:	90%
	Data acquisition process:	10%
	Data preparation and matching exercise:	70%
	Performed statistical analysis:	90%
	Interpreted results:	50%
	Planned and constructed tables and figures:	100%
	Wrote first draft of manuscript:	70%
	Led submission process:	100%
Revised paper according to external reviewers' comments:	65%	
Jané Joubert (Candidate)	Conceived and conceptualized paper:	10%
	Wrote research proposal:	90%
	Prepared ethics application process:	90%
	Searched, sourced, and reviewed literature:	10%
	Coordinated multi-institute collaboration:	100%
	Coordinated matching exercise:	100%
	Data preparation and matching exercise:	10%
	Interpreted results:	15%
	Wrote first draft of manuscript:	30%
Critically reviewed and intellectually strengthened drafts:	15%	
Revised paper according to external reviewers' comments:	15%	
Maletela Tuoane-Nkhasi	Data acquisition process:	50%
	Data preparation and matching exercise:	20%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	5%
Kathy Kahn	Data acquisition process:	20%
	Wrote research proposal:	2%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	5%
	Revised paper according to external reviewers' comments:	5%
Chalapati Rao	Wrote research proposal:	2%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
Xavier Gómez-Olivé	Performed statistical analysis:	5%
	Interpreted results:	3%
	Critically reviewed and intellectually strengthened drafts:	5%

	Revised paper according to external reviewers' comments:	5%
Paul Mee	Performed statistical analysis:	5%
	Interpreted results:	4%
	Critically reviewed and intellectually strengthened drafts:	5%
	Revised paper according to external reviewers' comments:	5%
Stephen Tollman	Data acquisition process:	5%
	Interpreted results:	3%
	Critically reviewed and intellectually strengthened drafts:	5%
	Revised paper according to external reviewers' comments:	5%
Alan Lopez	Data acquisition process:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
Theo Vos	Wrote research proposal:	2%
	Data acquisition process:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
Debbie Bradshaw	Conceived and conceptualized paper:	80%
	Wrote research proposal:	2%
	Data acquisition:	5%
	Interpreted results:	10%
	Critically reviewed and intellectually strengthened drafts:	15%

### Incorporated in Chapter 3:

**Joubert J**, Rao C, Bradshaw D, Vos T, Lopez AD. Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007. PLoS ONE. 2013; 8(5): e64592. doi:10.1371/journal.pone.0064592.

<b>Contributor</b>	<b>Statement of contribution</b>	
Jané Joubert (Candidate)	Conceived and conceptualized paper:	75%
	Searched, sourced, and reviewed literature:	100%
	Data acquisition:	10%
	Performed statistical analysis:	90%
	Interpreted results:	70%
	Planned and constructed tables and figures:	85%
	Wrote first draft of manuscript:	100%
	Led revision and submission process:	100%
	Revised paper according to external reviewers' comments:	80%
Chalapati Rao	Conceived and conceptualized paper:	10%
	Performed statistical analysis:	5%
	Interpreted results:	10%
	Planned and constructed tables and figures:	10%
	Critically reviewed and intellectually strengthened drafts:	50%
	Revised paper according to external reviewers' comments:	10%
	Reviewed, final comments, and approved final manuscript:	40%
Debbie Bradshaw	Data acquisition:	90%
	Conceived and conceptualized paper:	5%
	Interpreted results:	10%
	Critically reviewed and intellectually strengthened drafts:	15%
	Revised paper according to external reviewers' comments:	3%
	Reviewed, final comments, and approved final manuscript:	15%
Theo Vos	Conceived and conceptualized paper:	5%
	Performed statistical analysis:	5%
	Interpreted results:	5%
	Planned and constructed tables and figures:	5%
	Critically reviewed and intellectually strengthened drafts:	20%
	Revised paper according to external reviewers' comments:	4%
	Reviewed, final comments, and approved final manuscript:	25%
Alan Lopez	Conceived and conceptualized paper:	5%
	Interpreted results:	5%
	Critically reviewed and intellectually strengthened drafts:	15%
	Revised paper according to external reviewers' comments:	3%
	Reviewed, final comments, and approved final manuscript:	20%

## Incorporated in Chapter 2:

**Joubert J**, Rao C, Bradshaw D, Dorrington RE, Vos T, Lopez AD. Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa. *Global Health Action*. 2012; 5(19263).

<b>Contributor</b>	<b>Statement of contribution</b>
Jané Joubert (Candidate)	Conceived and conceptualized paper: 80% Searched, sourced, reviewed data sources and literature: 100% Interpreted results: 70% Planned and constructed tables and figures: 90% Wrote first draft of manuscript: 85% Led revision and submission process: 100% Revised paper according to external reviewers' comments: 85%
Chalapati Rao	Conceptualized aspects of the paper: 10% Interpreted results: 10% Planned and constructed tables and figures: 7% Wrote part of the first draft of manuscript: 15% Critically reviewed and intellectually contributed to drafts: 60% Revised paper according to external reviewers' comments: 5% Reviewed and approved final manuscript: 40%
Debbie Bradshaw	Conceptualized aspects of the paper: 4% Interpreted results: 10% Critically reviewed and intellectually contributed to drafts: 10% Revised paper according to external reviewers' comments: 4% Reviewed and approved final manuscript: 20%
Rob Dorrington	Conceptualized aspects of the paper: 2% Interpreted results: 4% Planned and constructed tables and figures: 3% Critically reviewed and intellectually contributed to drafts: 10% Revised paper according to external reviewers' comments: 2% Reviewed and approved final manuscript: 20%
Theo Vos	Conceived and conceptualized paper: 2% Interpreted results: 3% Critically reviewed and intellectually contributed to drafts: 10% Revised paper according to external reviewers' comments: 2% Reviewed and approved final manuscript: 10%
Alan Lopez	Conceptualized aspect of the paper: 2% Interpreted results: 3% Critically reviewed and intellectually contributed to drafts: 10% Revised paper according to external reviewers' comments: 2% Reviewed and approved final manuscript: 10%

### **Contributions by others to the thesis**

Significant and substantial inputs made by others to this research are acknowledged in the contributions to published papers.

Prof. Theo Vos, Prof. Alan Lopez, Assoc. Prof. Chalapati Rao and Prof. Debbie Bradshaw contributed significantly to this work by supplying detailed comments during all phases of the study; critically appraising decisions regarding the methods, findings and interpretation of the results; and providing critical comments on the structure and content.

During the initial phases of candidature, Dr Tim Adair and Dr Simon Forsyth assisted with technical analysis of the national mortality data.

### **Statement of parts of the thesis submitted to qualify for the award of another degree**

None.

## **Acknowledgements**

I thank my precious husband, Samuel, and our amazing children, Janet and Francois, for their invaluable support and priceless patience. I thank my parents, Attie and Benoni, for their encouragement and support. Mamma, your everlasting love, prayers and inspiration have been an indescribable treasure. Pappa, I miss you unremittingly.

I thank my advisors, Prof. Theo Vos, Prof. Alan Lopez, Assoc. Prof. Chalapati Rao, and Prof. Debbie Bradshaw, for their distinguished intellectual inputs, remarkable knowledge and academic guidance, and ongoing support. Your critique and expectations have been tough at times. It was, however, enormously rewarding to learn from giants.

Special thanks are due to Ms Mary Roset, Postgraduate Administration Officer, and Prof. Deirdre McLaughlin and Dr Ruth Crowther, Postgraduate Coordinators, for their tremendous support and encouragement during my candidature.

I gratefully acknowledge the Endeavour International Postgraduate Research Scholarship, the University of Queensland Research Scholarship, and the UQ International Scholarship awarded to me during my enrolment at the University of Queensland, Australia. I also thank the South African Medical Research Council for its financial contributions during my studies in Australia.

I gratefully acknowledge Statistics South Africa and the Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt) for making available mortality data for the linkage study, and the important contribution of the Agincourt HDSS field workers and other staff members responsible for the collection and custodianship of the demographic and mortality data used in this work.

I thank Ms Elize de Kock for her precious and outstanding administrative and wellness support, and Mr Alexander Samuelowicz-Foster for editorial assistance.

Thank you Elize, Nadine, Rosana, Linda, Betsie, Petra, Mam, Tan, Twinz, Lyn, Vicky, Victoria, Monique, Claudette, Sulaiman, Edward, Anita and Johan, Mary and Frantz, Ina, Hilda and André, Joan, Sarie, Elmarie, Mweete, Toyin, Rifqah, Tracy, and other colleagues and friends, for your encouraging words and gestures.

## **Keywords**

mortality, cause of death, data quality, civil registration, vital statistics, record linkage, death certification, verbal autopsy, public health, epidemiology

## **Australian and New Zealand Standard Research Classifications (ANZSRC)**

ANZSRC code: 160304, Mortality, 50%

ANZSRC code: 111706, Epidemiology, 50%

## **Fields of Research (FoR) Classification**

FoR code: 1117, Public Health, 50%

FoR code: 1603, Demography, 50%



**Availability and quality  
of vital registration mortality data in South Africa:  
A national and surveillance site perspective**

---

## Table of Contents

<b>1</b>	<b>Chapter 1: Introduction.....</b>	<b>1</b>
1.1	Aim and scope of the thesis.....	1
1.2	Graphical overview and activity pathway of the thesis.....	1
1.3	Overview of Chapter 1 .....	1
1.4	Global importance and utility of mortality data for public health.....	1
1.5	Mortality data: sources, techniques and data-collection methods.....	4
1.6	National civil registration and vital statistics (CRVS) systems.....	6
1.6.1	Development of the recording of vital events: an international overview .....	6
1.6.2	A recommended and preferred source of mortality data .....	8
1.6.3	Advantages of a civil registration and vital statistics system.....	9
1.7	Study setting 1: South Africa, a brief introduction .....	10
1.7.1	Geographic and population features .....	10
1.7.2	Socio-economic features.....	13
1.7.3	Health and health system features .....	15
1.7.3.1	Pre-democracy fragmentation, separation, and inequalities.....	15
1.7.3.2	Transition and a new trajectory of redress, redistribution and reduced inequality .....	16
1.7.3.3	The paradox remains .....	17
1.8	Study setting 2: Agincourt Health and Demographic Surveillance System.....	18
1.8.1	Geographic and population features .....	18
1.8.2	Socio-economic features.....	19
1.8.3	Health and health service features .....	19
1.8.3.1	Mortality indicators of health, and causes of death.....	19
1.8.3.2	Local health services .....	20
1.9	Availability, utility and quality of mortality data in South Africa .....	23
1.9.1	Collection of death and cause-of-death data in South Africa .....	23
1.9.1.1	Registration of deaths into the civil registration system .....	24
1.9.1.2	Cause-of-death information from death notification forms .....	25
1.9.2	Collection of death and cause-of-death data in Agincourt HDSS .....	26

1.9.2.1	Recording of deaths in the Agincourt HDSS .....	26
1.9.2.2	Cause-of-death information from Agincourt HDSS verbal autopsy questionnaires.....	27
1.10	Rationale for the research.....	28
1.11	Aim and research questions.....	29
1.12	Outline of thesis.....	30
<b>2</b>	<b>Chapter 2: Identification, review and documentation of post-apartheid mortality data sources for South Africa .....</b>	<b>33</b>
2.1	Introduction .....	33
2.2	Under-developed CRVS systems in most African countries, but transformation in South Africa.....	34
2.3	A next review of mortality data sources in South Africa.....	35
2.4	Paper 1: “Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa” .....	35
2.5	Concluding remarks.....	55
<b>3</b>	<b>Chapter 3: Civil registration and vital statistics in South Africa: history, current status, and the quality of national mortality statistics .....</b>	<b>59</b>
3.1	Introduction .....	59
3.2	CRVS in South Africa: a concise overview .....	59
3.2.1	Historical development.....	60
3.2.1.1	Challenges facing CRVS systems over the 1900s .....	60
3.2.1.2	Transformation affecting CRVS since the 1990s.....	63
3.2.2	Current status of CRVS in South Africa .....	63
3.2.3	The CRVS system in South Africa: concluding remarks .....	64
3.3	Evaluating CRVS mortality data: evaluation approaches, frameworks, and criteria .....	65
3.3.1	Organizations involved in evaluating health information and CRVS systems .....	65
3.3.2	Comparative evaluation of the quality of national mortality data and the criteria used in assessing national mortality data in institutional databases.....	66
3.3.3	Single-country evaluation of the quality of national mortality data and the criteria used in assessing national mortality data in countries’ own databases .....	68

3.4	Suitable criteria for evaluating South Africa’s national mortality data from CRVS .....	69
3.5	Paper 2: “Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997 – 2007” .....	72
3.6	Concluding remarks.....	86
3.6.1	Evaluation studies, tools, frameworks, and criteria.....	86
3.6.2	Quality of national mortality statistics from civil registration in South Africa.....	87
3.6.3	<sup>8,9</sup> Suggested comprehensive framework for South Africa and countries with comparable statistical development.....	87
<b>4</b>	<b>Chapter 4: Assessing the feasibility, utility, and quality of linking mortality data from the civil registration system and a health and demographic surveillance site .....</b>	<b>91</b>
4.1	Introduction .....	91
4.2	Record linkage.....	93
4.2.1	Origin and definition of record linkage .....	93
4.2.2	Record linkage approaches.....	94
4.2.3	Identifiers used in record linkage .....	94
4.2.4	Wide and increased use of data linkage for research in developed countries .....	95
4.2.5	Increased linkage: reasons and enabling factors.....	97
4.2.6	Data linkage in Africa.....	98
4.3	Studies linking CRVS data with mortality data from other sources .....	98
4.3.1	Assessing the feasibility, quality and optimal identifiers of linking CRVS and other mortality data.....	98
4.3.2	Assessing numbers of deaths and completeness of death registration, and identifying sources of bias via record linkage .....	99
4.3.3	Assessing the reliability of cause-of-death data in mortality data sources through record linkage.....	101
4.4	Proposing a linkage study with South African mortality data.....	102
4.5	Paper 3: “Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa” .....	103
4.6	Concluding remarks.....	114

<b>5</b>	<b>Chapter 5: Record-linkage as a tool for comparing cause-of-death data from civil registration and other data sources .....</b>	<b>117</b>
5.1	Introduction .....	117
5.2	Methods for evaluating and validating cause-of-death information.....	118
5.2.1	Electronic tools .....	118
5.2.2	Statistical and mathematical methods.....	118
5.2.3	Validation studies against a reference standard.....	118
5.2.3.1	Pathological autopsy review.....	119
5.2.3.2	Medical record review.....	119
5.2.3.3	Verbal autopsy review.....	121
5.2.3.4	Comparison with cause-of-death information from long-term follow-up studies, administrative and health databases, and disease registers .....	122
5.2.3.5	Record linkage.....	123
5.3	Potential sources for evaluating causes of death in South Africa .....	124
5.4	Drawing on the Agincourt HDSS/Statistics South Africa linkage study .....	125
5.5	Paper 4: “Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006-09” .....	125
5.6	Cause-of-death agreement according to selected characteristics of the decedent, venue of death, and type of ascertainment of the cause of death.....	140
5.7	Concluding remarks regarding published Paper 4 in the context of Chapter 5.....	142
5.7.1	Low cause-of-death agreement.....	142
5.7.2	Misclassification patterns .....	143
5.7.3	Cause-specific mortality fractions .....	144
5.8	Application of cause-specific mortality fractions as adjustment factors.....	144
5.8.1	Exploration with adjustment factors.....	145
5.9	Conclusion.....	149
<b>6</b>	<b>Chapter 6: Synopsis of findings, conclusions, and recommendations .....</b>	<b>153</b>
6.1	Synopsis of key findings, conclusions and recommendations .....	153
6.1.1	Review of South Africa’s mortality data sources.....	153
6.1.1.1	Recommendations .....	154

6.1.2	Evaluation of South Africa’s mortality statistics from CRVS .....	155
6.1.2.1	Recommendations .....	157
6.1.3	Data linkage for data quality assessment.....	158
6.1.3.1	Recommendations .....	159
6.1.4	Exploring an application of adjustment factors .....	161
6.1.4.1	Recommendations .....	162
<b>7</b>	<b>Chapter 7: Consolidation and contextualization of the thesis findings .....</b>	<b>165</b>
7.1	CRVS systems: Global progress slow, but moving forward with renewed interest .....	165
7.2	CRVS systems: Globally a preferred source of mortality data, but... ..	165
7.3	CRVS systems in Africa: A prolonged bleak picture.....	166
7.4	CRVS in Africa: Regional initiatives and high-level commitments .....	166
7.5	CRVS country initiatives: mixed stages and results, but largely moving forward .....	167
7.6	CRVS in South Africa .....	167
7.6.1	Progress and improved ratings over time .....	167
7.6.2	Key persisting challenges .....	168
7.6.2.1	Accuracy of cause-of-death data .....	168
7.6.2.2	Completeness of death registration .....	169
7.7	Concluding call: Reforming a “data-rich, but information-poor” situation .....	169
7.8	Contribution to existing knowledge and relevance of the research.....	171
	<b>References .....</b>	<b>175</b>
	<b>Notes .....</b>	<b>203</b>

## List of Tables in Thesis text

Table 2.1:	Contributions of different approaches to measuring key mortality and population health indicators.....	34
Table 2.2:	Contributions, broadly in terms of availability and suitability for mortality measurement, of different approaches to measuring key mortality and population health indicators in democratic South Africa .....	57
Table 3.1:	Criteria for evaluating the quality of country-level mortality data from civil registration systems .....	70
Table 3.2:	Comprehensive evaluation framework for assessing the quality of national mortality data from civil registration.....	89
Table 5.1:	Cause-of-death agreement according to selected demographic characteristics of the decedent, venue of death, and type of ascertainment of the cause of death.....	141

## Tables in published PDF papers

### Paper 1:

Table 1:	Post-democracy data sources for mortality analysis in South Africa by enumeration years.	Table 2: . National surveys measuring mortality, by year of survey, number of households and persons enumerated, and different methods of mortality measurement.
Table 2:	National surveys measuring mortality, by year of survey, number of households and persons enumerated, and different methods of mortality measurements.	
Table 3:	Selected facility-based data sources that may complement vital registration mortality data.	

**Paper 2:**

Table 1: Percentage of total deaths assigned selected ill-defined and non-specific codes by province of death occurrence, South Africa, 1997-2007.

**Paper 3:**

Table 1: Deterministic matches.

Table 2: Weights for the probabilistic linkage approach with blocking on sex and year of death.

Table 3: Weights for the probabilistic linkage approach with blocking on sex and year of birth.

Table 4: Factors predictive of successful matching of death records between Agincourt HDSS and South African CR system.

Table 5: Background characteristics of all 2009 Agincourt HDSS deaths compared to those matched with CR records.

**Paper 4:**

Table 1: WHO Mortality Tabulation List 1 and ICD-10 codes for the short list of causes of death.

Table 2: Matching rate and informant-reported level of death registration in the study population.

Table 3: Characteristics of the study populations, Agincourt HDSS and VR data, 2006–09.

Table 4: Ten leading causes of deaths from VA and CR, according to the WHO List: Agincourt HDSS, 2006–2009 (N=2137).

Table 5: Misclassification patterns for selected causes/cause groups in the Agincourt HDSS study site, 2006–2009.

Table 6: Agreement characteristics of civil registration and verbal autopsy diagnoses for the Short list causes/cause groups: Agincourt HDSS, 2006–2009.



Table 7: External cause from VA for under-specified CR injury deaths, Agincourt HDSS 2006–2009.

**List of Figures in Thesis text**

Figure 1.1: Map of Africa, with South Africa on the southern tip of the continent..... 11

Figure 1.2: Gross Domestic Product (GDP) per capita, at current US dollars, in selected countries, 2000 – 2012 ..... 14

Figure 1.3: Provincial delineations of South Africa, with geographic indications and insets of three INDEPTH surveillance sites. ....21

Figure 1.4: Map indicating the location of the villages in the Agincourt HDSS site (with red boundary) in north-east South Africa. ....22

Figure 1.5: The death notification, certification, and registration process to be followed after a death or stillbirth.....26

Figure S1A. Chapter R codes (R00–R99) in each age group as a percentage of total deaths in that age group, South Africa, 1997– 2007. B. Trends in the number of registered deaths by province of death, South Africa, 1997–2007. C. Trends in the number of deaths assigned an R code (R00–R99) by province of death, South Africa, 1997–2007. Source: StatsSA vital registration data. ....83

Figure S2A. Unadjusted age-specific death rates by year of death for three major cause groups, South Africa, 1997–2007. Source: Mortality data from StatsSA vital registration; population data from ASSA2008. ....84

Figure S2B. Unadjusted age- and sex-specific rates for cerebrovascular deaths, by province and nationally, 1997–2007: Log scale. Source: Mortality data from StatsSA vital registration; population data from ASSA2008. ....85

Figures 5.1 & 5.2: Top 20 causes of death for Ehlanzeni district, based on raw (Figure 5.1) and adjusted CRVS data (Figure 5.2)..... 146

Figure 5.3:	Raw and adjusted cause-specific mortality fractions for Ehlanzeni district.	147
Figures 5.4 & 5.5:	Top 20 causes of death for Ehlanzeni district based on adjusted CRVS data (Figure 5.4), and top 20 causes of death for Mpumalanga province (Figure 5.5).....	148
Figures 5.6 & 5.7:	Top 20 causes of death for Ehlanzeni district based on adjusted CRVS data (Figure 5.6), and top 20 causes of death for Ehlanzeni district based on the profile presented by the <i>District Health Barometer</i> report (Figure 5.7). ....	149
Figure 5.8:	Provincial and district administrative delineations of South Africa with insets of Ehlanzeni district and its five municipalities, and Agincourt HDSS with its villages.....	151

**Figures in published PDF papers**

**Paper 1:**

Figure 1:	The probability of dying between ages 15 and 50 (35q15) from different data sources.
-----------	--

**Paper 2:**

Figure 1:	Estimated completeness of death reporting in South Africa: 1994–2007.
Figure 2A:	Number of standard deviations by age and sex for South Africa, 2007: HIV/AIDS deaths included.
Figure 2B:	Number of standard deviations by age and sex for South Africa, 2007: HIV/AIDS deaths excluded.
Figure 3:	Proportion of total deaths due to leading categories and causes of death, 1997–2007.

**Paper 3:**

Figure 1: Location of Agincourt HDSS.

Figure 2: Agincourt HDSS records matched with CR records.

**Paper 4:**

Figure 1: Agincourt HDSS villages within the study site borders (light orange), stretching across three tribal areas, Amashangana, Jongilanga, and Hoxana, within the Bushbuckridge Municipality.

Figure 2: Establishing the study population: data sources, number of deaths, and number of comparable records, Agincourt HDSS, 2006 – 2009.

## Abbreviations and Acronyms

AIDS	-	Acquired Immunodeficiency Syndrome
BI-1663	-	Form for the notification of a death or stillbirth
CRVS	-	Civil registration and vital statistics
DHA	-	Department of Home Affairs
GBD	-	Global Burden of Disease
HIS Hub	-	Health Information Systems Knowledge Hub
HIV	-	Human Immunodeficiency Virus
ICD-10	-	International Statistical Classification of Diseases and Related Health Problems, 10th Revision
MDG	-	Millennium Development Goal
NCD	-	Non-communicable disease
SDG	-	Sustainable Development Goals
Stats SA	-	Statistics South Africa
VA	-	Verbal autopsy
WHO	-	World Health Organisation
UN	-	United Nations
US	-	United States of America
UNPD	-	United Nations Population Division

# 1 Chapter 1: Introduction

---

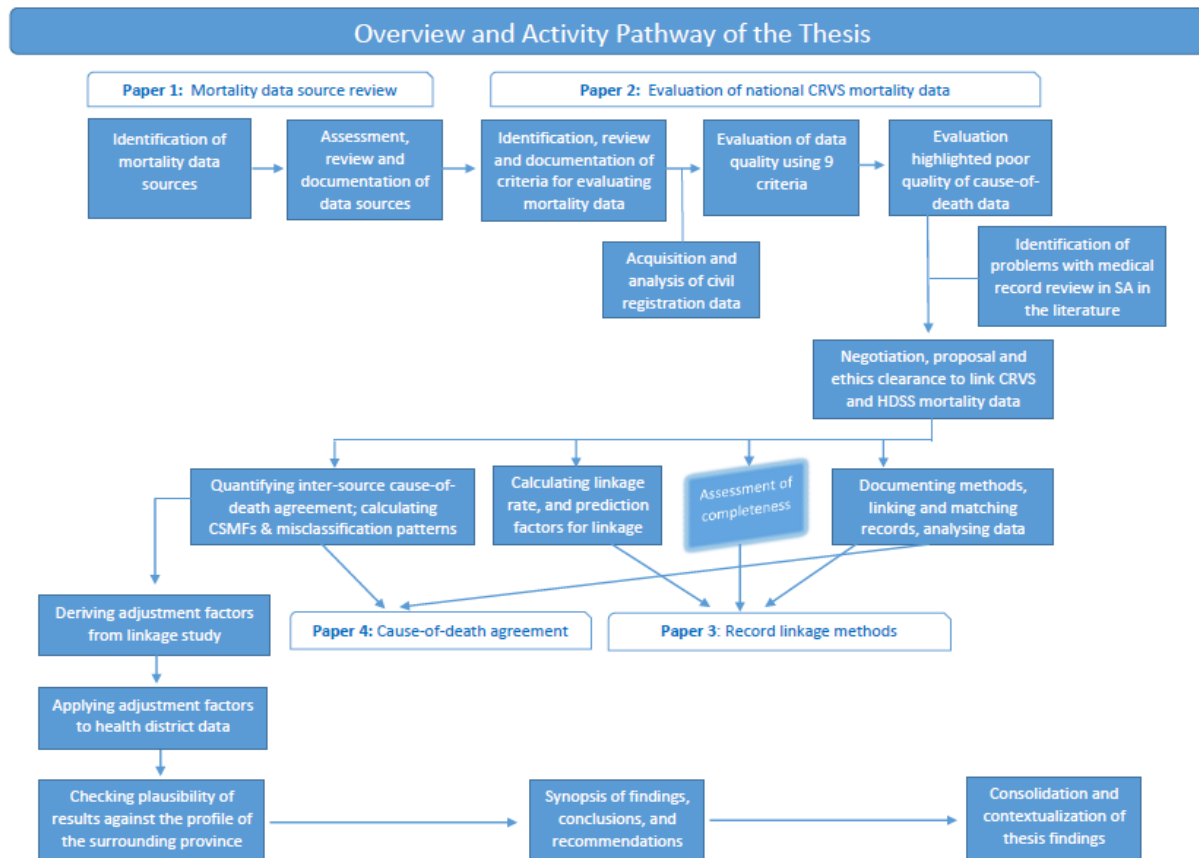
## 1.1 Aim and scope of the thesis

The importance of valid and reliable mortality data as a key input for public and population health planning and research, and development programmes, is globally acknowledged.<sup>1-7</sup> However, few low- and middle-income countries have such data. Even fewer have assessed the quality of their mortality data.<sup>8,9</sup> Despite the particular utility of periodic, independent evaluations of national mortality statistics by countries themselves,<sup>9-11</sup> such an evaluation has not yet been conducted for South Africa. Neither has a review been conducted of the sources of mortality data in South Africa since 1995.<sup>12</sup> Regarding causes of death, a few small-scale studies assessed data quality in urban settings, but no study yet in civil registration mortality data in a rural setting. In addition, notwithstanding district health managers' critical need to know the health care needs of their district population, current district mortality profiles are accompanied by strong warnings of invalid data.<sup>13</sup>

These gaps in essential public- and population-health knowledge present important reasons for research on the availability and quality of mortality data in South Africa. The primary aim of the thesis therefore was to assess the quality of mortality data from the national civil registration and vital statistics (CRVS) system of South Africa. Given the 20-year lapse since the previous review of CRVS mortality data,<sup>12</sup> and acknowledging the complementary value of alternative sources of mortality data,<sup>2,4,14</sup> a secondary aim was to identify and review alternative mortality data sources, and data availability from these sources.

Different approaches were followed to achieve these aims, including a review and critical appraisal of different mortality data sources (Chapter 2); a comprehensive evaluation of CRVS mortality data based on demographic and epidemiological principles (Chapter 3); a record-linkage study applying deterministic and probabilistic methods (Chapter 4); an application of adjustment factors from the linkage study aiming to improve the quality of district-level cause-of-death data (Chapter 5); and a summary of key findings, conclusions and recommendations from these chapters as they relate to the seven research questions investigated in the study (Chapter 6). These content chapters have been preceded by an introduction to the thesis (Chapter 1). In the final chapter (Chapter 7) the thesis findings have been consolidated in the context of the latest directions in the literature.

## 1.2 Graphical overview and activity pathway of the thesis



**Note:** The assessment of completeness of death registration was an objective of the linkage study, but could not be calculated due to differences in geographic borders of the village and tribal areas reported onto the death notification form.

### **1.3 Overview of Chapter 1**

*Chapter 1* describes the importance and utility of mortality data for decision-making in public health and indicator measurement in population health, and highlights different sources of mortality data. Given the strong focus on mortality data from CRVS systems throughout the thesis, their historical development, attributes, and advantages for mortality measurement are described. As further background to the content chapters, selected features of the study settings referred to in the thesis, i.e. South Africa and the Agincourt HDSS, are provided. This is followed by the rationale, aims and research questions of the thesis.

### **1.4 Global importance and utility of mortality data for public health**

A death constitutes a significant vital event, not only in the life of an individual, their kin and community, but also as an important incident for which basic information about the deceased and causes of the death can be collected to inform research agendas, public-health policy, and funding decisions. Ideally, the event and basic information are recorded into a well-functioning national civil registration and vital statistics (CRVS) system, which is widely valued as the optimal source of mortality data.<sup>3,9,15-18</sup> Upon such registration, personal documents of legal value can be prepared for individual use, and mortality statistics compiled for national use and international comparison.

Information on deaths, by age and sex, is indispensable for a number of purposes, including annual population estimates and projections at different levels of geographic aggregation. These, in turn, are important for the planning and evaluation of social and economic programmes, including those dealing with health care, maternal and child health, public housing, education, and social security. Death and population data by age and sex are basic, but essential, for the measurement of important demographic, health, and developmental indicators, such as crude, age-specific, and age-standardized death rates; neonatal, infant and under-five mortality; life expectancy at birth, and other life-table calculations, such as adult mortality. Mortality estimates are also needed for measuring the risk of dying at specific ages, for social security and insurance purposes.<sup>2,19,20</sup>

When complete, deaths statistics derived from civil registration are the only nationally-representative source of data on mortality by cause of death, as well as on sub-national

differences and inequalities.<sup>2</sup> Reliable and valid cause-of-death information is essential for assessing the leading causes of death, premature death, and mortality differentials in a population, and, accordingly, planning and prioritizing appropriate health interventions, and for evaluation of the health outcomes of such interventions. While mortality data are widely acknowledged as key input at the country level, for guiding health, development and research agendas,<sup>8,21</sup> the importance ascribed to data on deaths and causes of death is also reflected in a number of global and multi-national initiatives aimed at measuring and monitoring health, as well as informing health policies.<sup>2,19</sup>

A number of targets from international protocols and instruments are expressed in terms of mortality. Two of the targets and five progress indicators of the Millennium Development Goals (MDGs) are measures of mortality;<sup>22</sup> one of three components of the Human Development Index<sup>a</sup> – life expectancy at birth – is a mortality measure;<sup>23-25</sup> two of 11 indicators specified by the Commission on Information and Accountability for Women’s and Children’s Health are mortality indicators;<sup>26</sup> and the WHO Global Non-communicable Disease (NCD) Action Plan 2013–2020 includes an overarching target of reducing premature mortality from four major NCDs<sup>b</sup> by 25% relative to their 2010 levels by 2025 (referred to as the 25×25 target).<sup>27</sup>

To facilitate international and regional comparisons, and help detect priority areas for local research, interventions and donor assistance, a number of country-level mortality tabulations are available from electronic and hard-copy international data compendiums. Tabulations of country-level mortality data have been published, for example, in the United Nations’ Demographic Yearbook since 1948, as well as the WHO’s World Health Statistics Annual since 1948, followed by the WHO’s Annual Epidemiological and Vital Statistics, and currently the WHO World Health Statistics Report, which contain extensive statistics on deaths and causes of death from member countries. The Demographic Yearbook has included special topics on mortality statistics in a number of years,<sup>28</sup> and currently includes various tabulations of foetal mortality, infant and maternal mortality, and general mortality with information on deaths, crude death rates, age-specific mortality rates, and the probability of dying.<sup>29</sup> Since 2006, the United Nations Statistics Division has produced the Population and Vital Statistics Report. This is a short report with three tables, the third of which includes data for deaths and infant deaths, by country, for the latest available year.<sup>30</sup>



While these initiatives emphasize the importance of mortality data as integrated into broad health, demographic or developmental themes, over the past two decades a number of other initiatives have established a strong focus on strengthening and improving the collection, production and use of mortality data per se. These include the establishment of the Mortality Section in the United Nations Population Division (UNPD);<sup>31</sup> publication of the UN's first World Mortality Report;<sup>32</sup> the establishment of the WHO High-Level Advisory Panel on Health Statistics;<sup>33,34</sup> various research and partner engagements in different countries by the Health Information Systems Knowledge Hub (HIS Hub) at The University of Queensland;<sup>18,19,35,36</sup> and the launch of the Health Metrics Network, with its drive to record every birth, death, and cause of death under the *MOVE-IT for the MDGs* programme (Monitoring of Vital Events for the MDGs through innovation).<sup>37-39</sup>

The numbers of deaths and the patterns of causes of death are endorsed by the Global Burden of Disease (GBD) initiative as a “crucial starting point” and “fundamental building block” for determining a population's burden of disease.<sup>21,40</sup> Mortality was a principal data component in quantifying and comparing the health of the world's populations in the first GBD Study in 1990,<sup>41</sup> and remains an indispensable part of measuring the past and projected health of populations globally in ongoing GBD work.<sup>21,42,43</sup> Reliable rates of age-specific and premature mortality and valid cause-of-death data, are an important impetus for evidence-based priority setting and health policy action, as well as a tool to help measure accountability in achieving results and the effective use of resources.<sup>21,44</sup> The common metric used in burden of disease studies, the Disability-Adjusted Life Year or DALY,<sup>c</sup> inextricably links mortality and health.<sup>45</sup> The GBD work generated widespread collaboration and widely-published findings, and spawned various national burden of disease studies and preparations towards future burden of disease studies, thereby further imparting the value of generating quality mortality data from local sources.

The importance of good-quality mortality data for assessing and improving population health has often been accentuated by scholarly comments and recommendations in the literature. In two international reviews of mortality data collected from CRVS in countries that report mortality data to the WHO, deaths and their causes are acknowledged as the most widely available and most commonly used data in health policy and research,<sup>8</sup> and as essential for understanding countries' public-health status.<sup>9</sup> In 2000 and 2006, the *Bulletin of the WHO* published theme issues on the topics of *Child Mortality* and *Estimating Mortality*, respectively.

In the latter issue, information regarding causes of death is highlighted as key for understanding the nature and magnitude of health problems in populations. Over the past decade, *The Lancet* has dedicated a number of series to mortality, including *Child Survival* (2003), *Neonatal Survival* (2005), *Maternal Survival* (2006), *Who Counts?* (2007), *Stillbirths* (2011), and *Every Newborn* (2014).

The advantages of continuous and representative death and cause-of-death data from a CRVS system have been emphasized strongly in the *Lancet Who Counts?* series which describes vital statistics and cause-of-death data as essential public goods of crucial importance to the health sector and beyond.<sup>5,9,14,46,44</sup> Recognizing that many countries have deficient CRVS systems and lack medically-certified cause-of-death information, but that verbal autopsy methods could help bridge the cause-of-death data gap, the Population Health Metrics Research Consortium and the *Population Health Metrics* journal launched in 2011 the thematic series “Verbal autopsy: Innovations, applications, opportunities—Improving cause-of-death measurement”.<sup>47,48</sup> The results are intended to help decision-makers choose the best VA techniques to determine causes of death in their respective study populations.<sup>47</sup>

These initiatives, recognizing the crucial role of mortality data, contribute to a shared vision, among global, country and independent role-players to improve population health globally. To optimize population health and reduce health inequalities, governments need to make important decisions about health expenditure and the distribution of resources. Such decisions should be influenced by a comprehensive understanding of the health needs of the population and the health outcomes of health interventions, which, in turn, requires a careful assessment of the leading causes of disease, injury and death.<sup>49</sup> Assessing the leading causes of death should preferably be informed by an effective vital statistics system rooted in a well-functioning civil registration system.<sup>2,8,9,19</sup>

### **1.5 Mortality data: sources, techniques and data-collection methods**

While there is wide agreement that records of vital events collected in a fully-functional civil registration system are the optimal source of mortality data for the production of a country’s mortality statistics,<sup>2,5,8,9,19</sup> other sources of data are also used to collect information on deaths and other vital events.<sup>2,6,14,15</sup> These data sources are often used as alternative or substitute sources in countries with deficient or insufficiently reliable systems of civil registration, but also as complementary data sources in countries with well-functioning CRVS systems for in-

depth analysis of demographic and health topics.<sup>2,14</sup> Other sources include records from health facilities; vital records from sample registration; information from specific questions on births and deaths in household surveys and population censuses; and information from the application of indirect techniques of demographic estimation.<sup>2</sup>

A number of techniques and data-collection methods are being used to detect the occurrence of individual deaths, and assemble information about these deaths, contributing towards the goal of providing reliable statistics about the level and causes of mortality in a population. The techniques include (a) continuous and compulsory recording of births and deaths, shortly after death, as typically done in a civil registration system;<sup>50-53</sup> (b) recording of births and deaths by a part-time registrar in sampled areas which are also visited by an independent survey team that ask all sampled households about births and deaths during the previous six months, as done in a sample registration system.<sup>54</sup> These two sets of information concerning vital events are then matched, from which a final number of events are calculated. Other techniques include (c) retrospective questions about deaths occurring in a specified period of time (e.g. over the 12 months prior to the interview), as may be asked in a census or survey questionnaire;<sup>55,56</sup> (d) retrospective questions about the cumulative number of deaths that have occurred, as may be asked in a household survey, for example, to gather information about the number of surviving parents of the interviewee, or the number of children ever born and surviving of each eligible women in a sampled household;<sup>56-58</sup> (e) the listing of people in a household at a point in time and follow-up of these persons over time by repeated visits and the recording of deaths that occurred between visits, as done in demographic surveillance sites;<sup>59-63</sup> and (f) comparing the number of persons by age and sex at two points in time towards deriving the number of deaths in between time points.<sup>6,14 64,65</sup>

Data collection methods other than those used in CRVS systems include population censuses, household sample surveys, sample vital registration, record keeping at health facilities, and health and demographic surveillance.<sup>2,14</sup> The strengths and limitations of the different techniques and methods are discussed in the literature; for example, in United Nations publications,<sup>2,6,15,66</sup> demography textbooks,<sup>67,68</sup> and research papers.<sup>14,46,54</sup> Such strengths and limitations are included in the *Introduction* and *Discussion* sections of the published review article in Chapter 2. These methods and techniques can be, and often are, used together as alternative data sources to gather information on epidemiological or demographic processes

where CRVS systems are lacking. These methods are also useful as complementary data sources to enrich, evaluate, and, where feasible, validate civil registration data.<sup>2</sup>

Given that the primary focus of this thesis is on mortality data from CRVS systems, civil registration and vital statistics will be defined, and their historical development, attributes, and advantages for mortality measurement highlighted in Section 1.3.

## **1.6 National civil registration and vital statistics (CRVS) systems**

A recent United Nations publication defines civil registration as “the continuous, permanent, compulsory and universal recording of the occurrence and characteristics of vital events pertaining to the population” in a country.<sup>2</sup> Civil registration is carried out as a governmentally operated administrative system, mandated by law to create a record of each vital event occurring within the boundaries of a country, and in accordance with the legal requirements of that country.

Vital statistics are defined as “a collection of statistics on vital events in a lifetime of a person, as well as relevant characteristics of the events themselves and of the person and persons concerned”.<sup>2</sup> Vital events are events that concern the life and death of individuals, and the formation, termination or invalidation of legal relationships, which in turn affect the civil status of those individuals. For statistical purposes, the United Nations currently recommends data to be collected on the following vital events: live births, deaths and foetal deaths (vital events proper); marriages, registered partnerships, separations, divorces, legal dissolutions of registered partnerships, annulments of marriage (dual events); adoptions, legitimations, and recognitions (vertical family events).<sup>2</sup>

### **1.6.1 Development of the recording of vital events: an international overview**

Evidence exists of fragmentary vital registration by civil authorities in the era before year 0; for example, in Egypt around 1250 B.C. during the reign of King Rameses II, and in Rome in the years 578 – 534 B.C., where citizens were required to report births, and officials were appointed to record facts related to births, deaths and adolescents.<sup>15</sup>

The first systematic, compulsory registration of births, deaths and marriages is reported to have been functioning in some parts of Japan as early as A.D. 720. While it is not known whether

this Japanese system was functioning as a religious or secular system,<sup>15</sup> the recording of vital events during the Middle Ages was commonly the concern of ecclesiastical authorities. A chronology of important events in the development of CRVS is given in the United Nations *Handbook of Vital Statistics Methods*, covering 3000 years, and reflecting the gradual development of ecclesiastical record-keeping of baptisms, burials and weddings of church members into secular record-keeping with compulsory registration of vital events for all.<sup>15</sup> Important milestones include: the 1532 ordinance in England that required weekly records of burials; the “Bills of Mortality”, to be issued by parish priests to inform about the number of plague and total deaths in the parishes of London; the passing of a law by the Catholic Church to keep registers of baptisms and marriages, as ordered by the Council of Trent in 1563, with registers subsequently instituted in many European countries and their colonies elsewhere; and the spreading of vital-event registration to Protestant Scandinavian countries, with the first systematic parish register established in Sweden in 1608, followed by a system of compulsory civil registration introduced in Finland in 1628, Denmark in 1646, Norway in 1685, and Sweden in 1686.<sup>15</sup>

Gradually, from the latter part of the 1600s, vital-event registration evolved from an ecclesiastical function into a secular system with compulsory registration, as shown in the chronology produced by the United Nations<sup>15</sup>, and country review of national CRVS practices.<sup>16</sup> Compulsory civil registration spread widely in European and British countries during the 1800s, but also to countries and regions in other parts of the world, including the United States of America, South America, Asia and the Caribbean.<sup>15</sup>

During the 1900s, further progress had been made in all continents; not only in terms of additional countries enacting civil registration laws and establishing civil registration systems, but also in terms of countries extending their systems beyond the registration of births and deaths to also include foetal deaths, marriages and divorces.<sup>16</sup> While most North American, European, then-USSR, and South American countries that participated in the United Nations country review, reported established civil registries during the 1800s or earlier, the majority of participating African and Asian countries had set up their systems – established colonially or independently – in the 1900s. In Oceania, the island countries of Fiji and French Polynesia established their civil registries in 1874; civil registration laws for the eight states of Australia were enacted from 1838 to 1856; those in New Zealand around 1850; and civil registries for Samoa, Vanuatu and Tonga were established after 1950.<sup>16</sup>

The historical development of CRVS in South Africa is presented in Section 3.2.1.

### 1.6.2 A recommended and preferred source of mortality data

Although a number of initiatives are at present being used for collecting mortality data, including household sample surveys, population censuses, sample registration, health and demographic surveillance, and health records, a well-functioning civil registration system is a widely-recommended and preferred source of vital-event records.<sup>2,8,9,19</sup> Information collected from a civil registration system has legal authority, and the primary purpose of civil registration is to establish personal legal documents of direct interest to individuals, required by citizens as proof of certain facts, such as parentage, marital status, and death. Civil registration records of births, for example, provide legal proof of an individual's identity and civil status, including name, nationality and age. Records of death provide evidence and date of death, as well as legal proof to process claims to inheritance or insurance benefits.<sup>2,19,66</sup>

Beyond recording vital events, civil registration functions include the reporting of selected information on vital events for statistical purposes, and providing timely and reliable information and data to other government agencies, such as the official statistical office and ministry of health; a population register, where operative; and research institutions.<sup>2</sup> Indeed, vital statistics, as a collection of statistics on vital events that occur in a specified population, form a key component of a country's national statistical system.<sup>69</sup>

Recording of vital events and converting them into vital statistics have significant value at both the individual and country level. While deaths, in particular, can be recorded in complementary or alternative data sources, recording mortality data into a well-functioning civil registration system with national coverage, remains the preferred source.<sup>2,5,8,9,19,46</sup> Such preference stems from the fundamental principles and attributes of a civil registration system,<sup>2,7</sup> i.e.:

- the *compulsory* nature of vital-event registration as mandated by legislation;
- *universal* coverage, by which the registration requirement applies to the entire population, regardless of geographical location or population subdivision;
- *permanence* of the registration system, contingent upon the authority given to those administering the system by the enactment of relevant legislation;
- *continuity* of the registration, for which permanence of the system is a requirement;
- protection of the *confidentiality* of the information supplied to the registration system;

- *regular dissemination* on a time schedule that will ensure the effective use of the analysis of the data to plan, operate and evaluate public health, economic and social programmes.

### **1.6.3 Advantages of a civil registration and vital statistics system**

These principles ensure a number of advantages when using mortality data from a civil registration system compared to data derived from other methods and sources. These include individual legal and protective advantages, like the possibility of establishing and issuing personal legal documents that are required by citizens as proof of certain facts, such as a death certificate providing evidence, date and, where possible, cause of death, as well as the legal proof necessary for inheritance claims, property transfer, or insurance benefits, and the legal right of a widowed spouse to remarry. Universal death registration holds certain administrative advantages, such as the maintenance of a population register, updating of pension fund and medical aid registers, and updating of electoral lists. A well-kept civil registration system has a cost advantage: once established, it is relatively inexpensive to obtain vital statistics because such statistics are a by-product of an existing administrative process. Statistical advantages of a well-functioning system include the generation of records which are not subject to sampling errors and relatively free from certain types of response errors. Where a well-functioning civil registration system is in place, large numbers of records are routinely informed by a country's standardised death notification forms, the international form of the medical certificate of the cause of death, and coding of the cause of death according to a standard diagnostic tool such as the International Statistical Classification of Diseases and Related Health Problems (ICD).<sup>2,19,66</sup>

A well-administrated civil registration system with universal, continuous coverage can therefore provide representative death and cause-of-death data to inform research, health policy or interventions, and resource allocation at different geographic or administrative levels, offering epidemiological patterns, cause-of-death profiles, and changing trends over time for different aggregations of the population. It can also provide data which can be evaluated against records from other data sources, providing a powerful resource for studying completeness of death registration and validity or plausibility of cause-of-death attribution.<sup>2,9,15,19,43</sup>

## **1.7 Study setting 1: South Africa, a brief introduction**

As further background to the research conducted for the thesis, Sections 1.4 and 1.5 provide a brief introduction to South Africa and the Agincourt Health and Demographic Surveillance System (HDSS). The Agincourt HDSS is the area in which a field study has been conducted to extend the body of research presented in Chapters 2 and 3.

### **1.7.1 Geographic and population features**

South Africa is the southernmost country of Africa, stretching latitudinally from 22° to 35° South and longitudinally from 17° to 33° East. Its surface covers an area of 1 219 090 km<sup>2</sup>, surrounded on the eastern, southern and western border by a coastline of about 3 000 kilometres, and having common boundaries with Namibia in the northwest, Zimbabwe and Botswana in the north, and Mozambique and Swaziland in the northeast. The Kingdom of Lesotho and Swaziland lie within the southeast part of the country, bordered by South African territory.<sup>70</sup> (Figure 1.1.)

Paleontological research into the origins of humanity indicates that modern humans have lived in South Africa for over 100,000 years, and their ancestors for some 3,3 million years.<sup>70</sup> Among the first peoples known to have lived in the country, are the San and KhoiKhoi people, estimated to have lived in southern Africa for between 10,000 and 20,000 years.<sup>71</sup> About 2000 years ago, Bantu-speaking agro-pastoralists from the north began arriving in southern Africa, settling in the eastern coastal region of southern Africa and introducing domesticated crops and an Iron-Age culture into the region.<sup>70</sup>



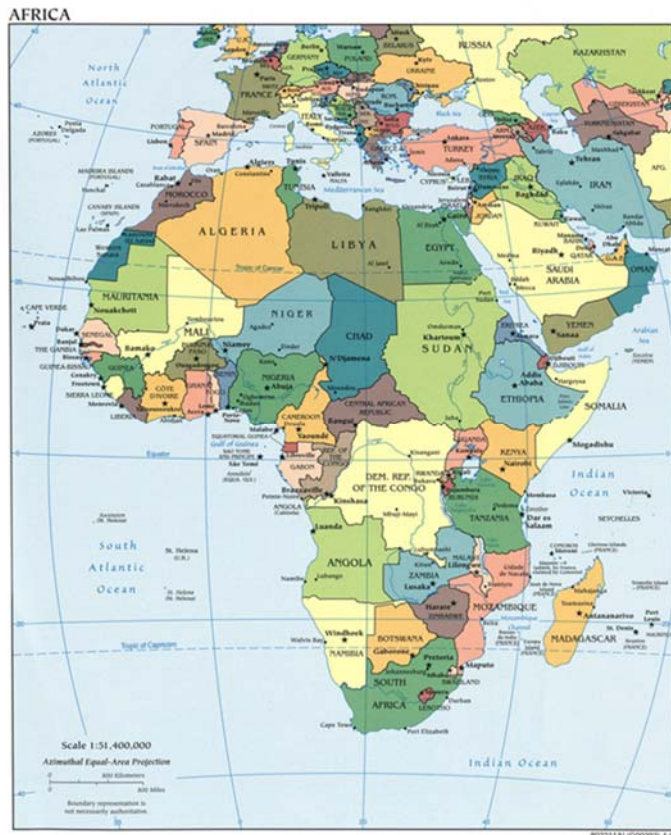


Figure 1.1: Map of Africa, with South Africa on the southern tip of the continent.

Source: [http://www.lib.utexas.edu/maps/africa/africa\\_pol](http://www.lib.utexas.edu/maps/africa/africa_pol)

Since the 1400s, Portuguese seafarers on route to the east visited the South African coast from time to time. In 1652, the Dutch established the first permanent European settlement in what was then the Cape of Good Hope (Cape Town). The first population enumerations were conducted shortly thereafter.<sup>72</sup> From 1654, slaves were imported to the country from Indonesia, Madagascar, Mozambique, and western Africa to work on the settler farms, and from 1859, indentured labourers from India were brought to what was then Natal, on the north-eastern coast.

South Africa's people and geography were subjected to imperialistic and ideological segregation and demarcation for an extended period. In 1910, the Union of South Africa was founded, with four provinces.<sup>71</sup> Since 1910, and before democratization in 1994, in addition to the four provinces, four independent states (Bophuthatswana, Ciskei, Transkei, and Venda, or the TBVC countries), and six self-governing territories (Gazankulu, KaNgwane, KwaNdebele, KwaZulu, Lebowa, and Qwaqwa) were established. These areas were ideological outcomes of separate-development policies and legislation, which will be referred to in Chapter 2. After the

first fully democratic election in 1994, the 10 geo-political territories were re-incorporated into the remainder of the country, and the country was divided into nine provinces, as shown in Figure 1.3, each with its own legislature, premier and executive council.<sup>70</sup>

The mid-year population for 2013 was estimated at 52,982,000 people, with a 48.7/51.3% male/female split. The overall annual population growth rate increased from an estimated 1,30% in the period 2002–2003 to 1,34% for 2012–2013, despite a declining rate of natural increase. This increase in the growth rate has been attributed to the high number of international immigrants assumed in the projection.<sup>73</sup> The African<sup>i</sup> population group constitutes an estimated 42,284 million people, or 79.8% of the total population; Whites, 4.602 million (8.7%); Coloureds, 4.766 million (9.0%); and Indians/Asians, 1.329 million (2.5%). Three in ten, 29.2% (11,640 million), of the population, is estimated to be younger than 15 years; 7.8% (4.133 million) 60 years or older;<sup>73</sup> and the median age was 25 years at Census 2011.<sup>74</sup> Unlike most countries in sub-Saharan Africa, population figures indicate fast demographic ageing in South Africa over the past few decades, with population ageing in 2013 most pronounced in Whites (19.7%) and Indians (10.7%), and least in Coloureds (7.4%) and Africans (6.6%), but older Africans making up the majority, 66%, of the total older (60+ years) population.<sup>75</sup>

Although population ageing in South Africa is less pronounced than in developed nations, it has the highest proportion of older persons in mainland sub-Saharan Africa. Second to Nigeria, it accommodates the largest number of older persons in the subcontinent.<sup>76</sup> Population estimates show steep increases in the South African population above the age of 60, and growth in this population is estimated to be considerably more extensive than in younger cohorts. Moreover, the majority of older South Africans have lived through several decades of discriminating policies during colonization and apartheid, expected to have impacted detrimentally on the prospects they have for healthy ageing.<sup>77</sup>

---

<sup>i</sup> The term ‘population group’ is used for a collective term that may elsewhere be called ‘race’ or ‘ethnic group’. The use of the terms “African”, “Coloured”, “Indian/Asian” and “White” is not intended to support or legitimise a racial or ethnic classification system, neither to denote biological difference. Under previous governments, South Africans were classified according to skin colour, and this went hand in hand with disparities in many spheres of life, including health. To acknowledge this impact, and to contribute to tracking progress in redressing past inequalities on the basis of this classification, these terms are used in the thesis. Additionally, the thesis takes a historical perspective at times, necessitating references to the classification terms.

Partly due to natural population growth and high levels of rural-to-urban migration, the rate of urbanisation has increased substantially over the past few years.<sup>78</sup> The population census of 2001 indicated that 56% of South Africans lived in urban areas,<sup>79</sup> a proportion which increased to about 60% by 2011.<sup>78</sup> While urbanization can be associated with increased living standards, in South Africa many rural-to-urban migrants have no choice but to settle down in informal settlements on the periphery of cities. These settlements are often characterized by rapid growth, placing a strain on local and national authorities to provide basic services to residents,<sup>80</sup> and by pervasive levels of poverty, inadequate housing, poor sanitation, and high levels of unemployment.<sup>81</sup> These conditions increase the risk of disease, injury and death for people who experience them.

### **1.7.2 Socio-economic features**

South Africa has the largest economy on the continent, contributing to 40% of the gross domestic product (GDP) of Africa. A high level of economic stability has been achieved since the transition to democracy in 1994.<sup>82</sup> The World Bank classifies South Africa as an upper middle-income country. GDP per capita,<sup>d</sup> at current US dollars, increased from US\$ 5,186 in 2005 to US\$ 7,352 in 2012, compared, for example, to Australian GDP per capita that almost doubled over the same period from US\$ 34,012 to US\$ 67,442. Since about 2006, GDP per capita has been performing weaker than in Brazil and the Russian Federation, but better than in China and India<sup>83</sup> (cf Figure 1.2).

Economic growth and human development investments over the past two decades have led to a decline in income poverty. However, the percentage of the population living on less than \$2.00 a day (PPP at 2005 international prices) was 43% (18.9 million) in 2000, and remained high at 31% (15.7 million) in 2009. The percentage living below the national poverty line (based on population-weighted subgroup estimates from household surveys) was 38% in 2000, and declined towards 2008, but remained high at 22%.<sup>84</sup> The World Bank estimated South Africa's Gini Index, or the extent to which the distribution of income among individuals or households within the country's economy deviates from a perfectly equal distribution, for the year 2000 at 0.58, and for 2009, at 0.63.<sup>85</sup> In 2009, the Gini Index was 0.42 for China, and 0.40 for the Russian Federation. The South African levels indicate enormous inequality, among the highest in the world, and the gap between rich and poor is widening<sup>82</sup>.

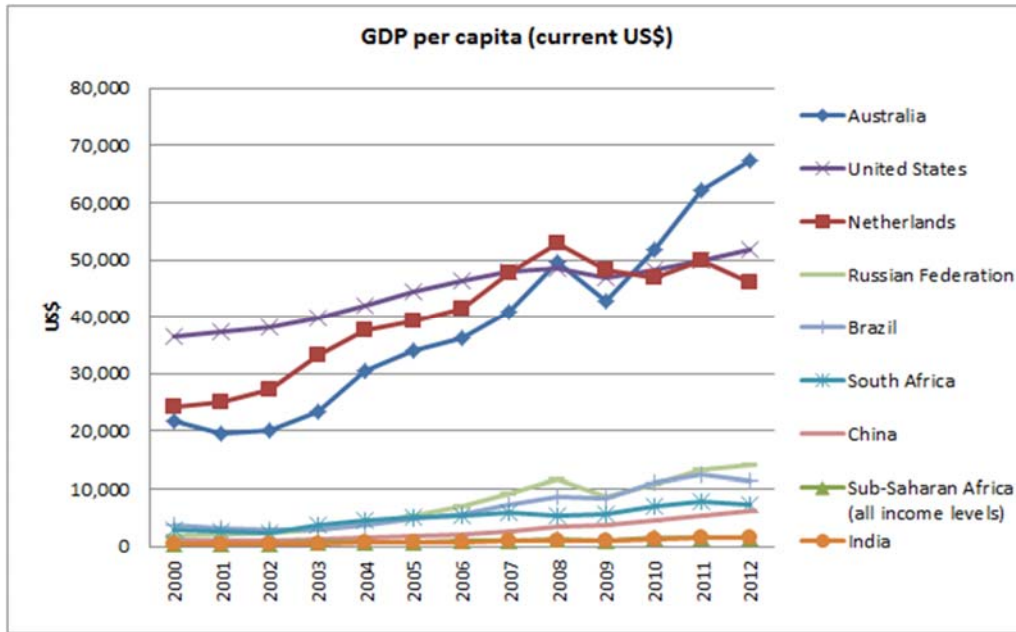


Figure 1.2: Gross Domestic Product (GDP) per capita, at current US dollars, in selected countries, 2000 – 2012

Source: Graph created using data from The World Bank, 2014<sup>83</sup>

South Africa has been suffering a very high unemployment rate for years. The unemployment rate, under the official definition (persons who did not work, but who looked for work and were available to work), was 29.8% in 2011. Using the expanded definition (persons who did not work, but were available to work in the reference period), the rate is higher, at 40%. This is in stark contrast to 2012 levels in other BRICS countries: 3.4% in India, 4.5% in China, and 6.9% in Brazil (according to The World Bank definition: the share of the labour force that is without work, but available for and seeking employment).<sup>86</sup> While the average annual growth rate of the labour force was 0.6% during the period 2006 to 2012, the average annual growth rate of the population was 1.3%. Comparative figures for upper middle-income countries are 0.9% growth in the labour force and 0.8% in the population.<sup>87</sup>

Despite being an upper middle-income country and having a rising GDP per capita, poverty, unemployment, and income inequality continue to present major challenges to health and human development. A number of programmes have been proposed that emphasize employment, agricultural growth, and land reform. Furthermore, the government invests heavily in cash transfers and free services to alleviate poverty and increase access to social services, with initiatives that include free primary health care for all at clinics; subsidized housing and transport; and the subsidized supply of electricity, water, sanitation, and rubbish

removal. Social grants were shown to have had a considerable impact on both health outcomes and poverty indicators; for example, the older persons' grant was found to reduce poverty-related vulnerability,<sup>88-90</sup> including a significant reduction in household reports of food insecurity.<sup>91</sup> Child support grants, which reached 11,228 million children in 2012,<sup>92</sup> help reduce poverty and have resulted in better child nutritional status for those receiving them, compared to children in households not receiving the grants.<sup>82</sup> Despite the considerable increase in social and welfare benefits and the expansion of basic services, including considerable investment in free primary health care services, social inequalities and disparities in access to health services and extensive health problems persist.<sup>93-95</sup>

### **1.7.3 Health and health system features**

#### **1.7.3.1 Pre-democracy fragmentation, separation, and inequalities**

The public sector health system was characterized by fragmentation and geographical and racial inequalities.<sup>96</sup> As a result of 'homeland' and separate development policies, 14 separate health departments existed towards the end of the 1980s.<sup>71</sup> Three separate 'own affairs' Departments of Health Services and Welfare for Coloureds, Indians and Whites were established under the Tri-cameral Parliament, formed in 1983.<sup>97</sup> Moreover, health services were primarily hospital-based; primary care services were underdeveloped and not universally available—especially poorly or undeveloped in rural and 'homeland' areas; many facilities were racially segregated; and for some time, preventive and curative services were separated.<sup>71</sup>

With the transformation from an agricultural to an industrial and mining-based economy, profound social and economic changes ensued. Massive foreign investments followed the discovery of diamonds and gold, and the demand for labour in these mining areas escalated. A combination of restrictions on access to land, coercive legislation, and taxes promoted the migration of male labourers to mining areas, where a lack of urban housing resulted in overcrowded, unsanitary accommodation for migrant workers, often in hostels and urban slums. Coovadia et al (2009) have highlighted the adverse impact that this, together with the policies of apartheid, had on health and health service delivery.<sup>71</sup>

### 1.7.3.2 Transition and a new trajectory of redress, redistribution and reduced inequality

During the political transition period from 1990 until 1994, a health commission was established to formulate a health plan to transform the health sector into a single system, with an equitable distribution of resources and expanded service delivery.<sup>96</sup> Post-1994, the 14 health administrations were consolidated into one national and nine provincial health departments. A strong emphasis was placed on reorienting health services towards primary health care delivered through a district health system, offered without cost in primary health care clinics as the first point of contact, usually staffed by nurses and community health workers. When medical needs exceeded clinic capacities, patients were referred to district hospitals, staffed by local medical doctors and nurses. The next levels in the referral chain are regional hospitals serviced by specialists; tertiary hospitals offering advanced surgical services; and national central hospitals with highly specialized services and technology.<sup>71,82</sup>

Post-1994 transformation included progress in redistributing resources among different levels of care and between under and well-resourced geographic areas, with significantly increased expenditures in poor districts; expansion of the number and range of health programmes, particularly so in primary health care settings; the building of more than 1300 new clinics; free primary health care for all; public health legislation to strengthen post-rape care, and reduce tobacco consumption and firearm violence; and free government health services for pregnant and breastfeeding women, and children under 6 years.<sup>82,95</sup> By 2012, an outcome of these changes was that 60% of children had access to a clinic or health centre within 2 kilometres of their homes, and 53% of children lived within 10 kilometres of the nearest hospital.<sup>92</sup>

A concerted government response to the devastating, interlinked HIV/tuberculosis epidemic included increased funding from R4.5 billion in 2009/10 to R8.4 billion budgeted in 2010/11 for expanding antiretroviral therapy (ART), promotion of the integration of HIV and tuberculosis treatment, scaling up the prevention of mother-to-child transmission of HIV, increased investments and efforts in HIV prevention, and a forceful national campaign to encourage 15 million sexually-active persons to be tested for HIV, resulting in 13 million tests for HIV being administered.<sup>94</sup> With an estimated 1.8 million people taking antiretrovirals by mid-2011, and ART services available in over 2500 approved facilities countrywide, South Africa remains the country with the largest ART programme in the world.<sup>94</sup>

A national strategic plan for the prevention and control of NCDs, with reference to cost-effective interventions and targets for reducing NCDs, has been launched.<sup>98</sup> The impact of reducing salt intake in the population has been studied, and the Ministry of Health has taken action towards formulating legislation to reduce the mean population intake.<sup>94</sup> Three national committees, respectively relating to maternal, perinatal, and child mortality, have been established and report directly to the Minister of Health. This has increased the profile of, and coordinated action for, the areas of health, linking mortality audit data to action, and the prevention of avoidable deaths.<sup>94</sup> A number of initiatives have been established to curb injury and violence, including the implementation of the Firearms Control Act in 2004; reduction in the number of illegal firearms; intensified enforcement efforts for arresting people driving under the influence of alcohol or drugs; and, in 2009, the formation of the Department of Women, Children and Persons with Disability which established the National Council Against Gender-Based Violence in 2012.<sup>94</sup>

### 1.7.3.3 The paradox remains

Yet, a recurring theme is the paradox that remains, wherein improvements in health are not being made, and health outcomes are persistently poor, despite high levels of health expenditure, vast local and overseas donor investments in health and welfare, and numerous supporting policies and programmes.<sup>82,95</sup> The poor health outcomes over the past two decades include insufficient progress for health MDG 5 (improving maternal health) and 6 (combating HIV/AIDS, malaria and other diseases), and decreases for MDG 4 (reduction of under-5 mortality); high levels of young-child undernutrition; radical reductions in life expectancy at birth by nearly 20 years since 1994, mainly due to the rapid rise in HIV-related mortality; doubling of the number of new tuberculosis cases during 2001-2006; the emergence of multidrug-resistant (MDR) and extensively drug-resistant (XDR) tuberculosis; delays in scaling up antiretroviral treatment and continued high rates of heterosexual transmission of HIV; rising death rates from cervical and prostate cancer, diabetes, and hypertensive and kidney disease; high proportions of adults suffering from a mental disorder; and unacceptably high rates of interpersonal violence and injury.<sup>95</sup>

The proportion of the global burden of disease borne by the approximately 48 million people of the country around 2000, was disproportionately high.<sup>95</sup> The total disability-adjusted life-years (DALYs) for high-burden conditions was almost equivalent to that of Bangladesh, whose population lives in considerably worse poverty, and is three times larger than South Africa's.<sup>95</sup>

For a number of years now South Africa has been facing a uniquely high and diverse burden of disease, referred to as a quadruple burden of disease,<sup>93,99</sup> or four “colliding epidemics”.<sup>71,94</sup> Poverty-related morbidity, such as infectious and parasitic diseases, malnutrition, and maternal, child and neonatal mortality remain widespread, alongside a massive burden from HIV/AIDS and tuberculosis—mentioned separately from other infectious diseases due to its enormous death, premature death, and disability toll in the country (31% of total disability-adjusted life-years (DALYs) were attributed to HIV/AIDS in 2000<sup>93</sup>). In addition, a very high burden of morbidity and mortality results, concurrently, from non-communicable diseases, and violence and injury.<sup>93,99-101</sup>

This combination of chronic and acute conditions, spanning all age groups, imposes a huge burden on an underdeveloped and weak public health-care system that struggles to serve about 85% of the population without private health cover, while struggling to overcome low staff-patient ratios, low staff morale, poor administrative management, limited infrastructure, and lack of funding.<sup>95</sup>

## **1.8 Study setting 2: Agincourt Health and Demographic Surveillance System**

An international network of 49 HDSSs, coordinated by the International Network for the Continuous Demographic Evaluation of Populations and Their Health (INDEPTH), currently operates in 20 countries, mostly in sub-Saharan Africa and Asia.<sup>102,103</sup> In South Africa, three such INDEPTH sites conduct longitudinal health and demographic studies of demarcated local populations. These are the Agincourt Health and Demographic Surveillance System (Agincourt HDSS), Dikgale, and Africa Centre Demographic Information System. The Agincourt HDSS was established in 1992 and is the longest-running of the three sites; it forms the research foundation of the *MRC/Wits Rural Public Health and Health Transitions Research Unit*.

### **1.8.1 Geographic and population features**

The Agincourt HDSS is a sub-district of the Bushbuckridge district in the Ehlanzeni health district in semi-arid rural north-east South Africa in the Mpumalanga Province, close to the eastern border with Mozambique (Figure 1.3, 1.4 & 5.5). The study site covers 402 km<sup>2</sup> and has 27 villages with both traditional and elected civic leadership. The position of the traditional leader or chief is inherited, whereas a number of Community Development Forums are aligned with local government, and consist of elected representatives. The baseline census of 1992



counted 57,609 individuals in 8,896 households in 20 villages. These numbers increased to a de jure population of 90,036 people in 16,000 households in 27 villages in 2011. The de facto population, (permanent residents in the site, i.e. not the temporary migrants who reside in the area for less than six months per year), was 66,876 in 2011. The area is relatively densely populated at 174 persons per km<sup>2</sup>.<sup>104</sup>

## **1.8.2 Socio-economic features**

Despite recent development initiatives, including electrification and tarring of the larger gravel roads, infrastructure remains limited. The area has no sanitation system other than pit latrines, and running water is restricted to erratic communal standpipes. By 2012, there was no piped water to dwellings, and electricity was affordable to only a minority of the population. Vehicle ownership is rare and minibus-taxis are the predominant form of transport. Plots of land are generally too small and rainfall too low to support and sustain subsistence farming.<sup>105</sup>

Generally, households supplement purchased maize and other foods with crops grown at home and wild foods collected from the surrounding environment. There are high unemployment levels in the area, contributing to the temporary labour migration of both men and women to bigger towns and cities in the province, as well as adjacent provinces.<sup>105</sup> The country's national non-contributory social grant system, in particular the old-age and child-support grants, and income from labour migration, are vital sources of household income.<sup>104</sup>

## **1.8.3 Health and health service features**

### **1.8.3.1 Mortality indicators of health, and causes of death**

Life expectancy at birth for both males and females dramatically declined from 1994 to 2004 by 13-14 years, and increased thereafter to reach 55.7 years in males and 64.4 years in females in 2009. Adult mortality doubled in men, and almost tripled in women between 1994 and 2004, but declined for men in 2009 to 505.9 deaths per 1000 population and for women to 382.4 deaths. Infant mortality increased from 24.7 in 1994 to 39.0 per 1000 live births in 2004, and remained at this level in 2009. Under-5 mortality was approximately 38 per 1000 live births for both boys and girls in 1994, and increased to 51 for girls and 45 for boys in 2009.<sup>104</sup>

Cause-of-death data for the period 2002-2005 broadly reflects a profile similar to that of the national population, with causes of death being attributable to four identifiable broad groups –

HIV/AIDS and tuberculosis; other communicable and poverty-related conditions; NCDs; and injuries, presenting simultaneously. Mortality rates in the study site have increased substantially since the mid-1990s to 2002-05. All-cause mortality increased by 87% over this period, mainly due to a six-fold increase in age-standardized death rates from infectious and parasitic disease, from 74 to 446 per 100,000 person years; but also due to a modest increase in NCD age-standardized death rates from 197 to 227 per 100,000 person years. Age-specific NCD mortality rates increased significantly in adults 30 years or older. Age-standardized death rates for injuries remained constant over time at the fairly high level of about 70 per 100,000 person years.<sup>63</sup>

### **1.8.3.2 Local health services**

Two health centres and five primary health care clinics are located within the study site. Three district hospitals are outside the site, between 25 and 60 km away. Kahn *et al* reports that primary care management of NCDs and their associated risk factors is limited in rural settings, while chronic infectious disease demands considerable resources from local health services, and dominates service development.<sup>104</sup> The authors suggest, along with others,<sup>63</sup> integrated approaches to community-oriented provision of long-term care in the face of rising mortality and risk from NCDs alongside pervasive levels of HIV and tuberculosis infection, and related morbidity.<sup>104</sup>

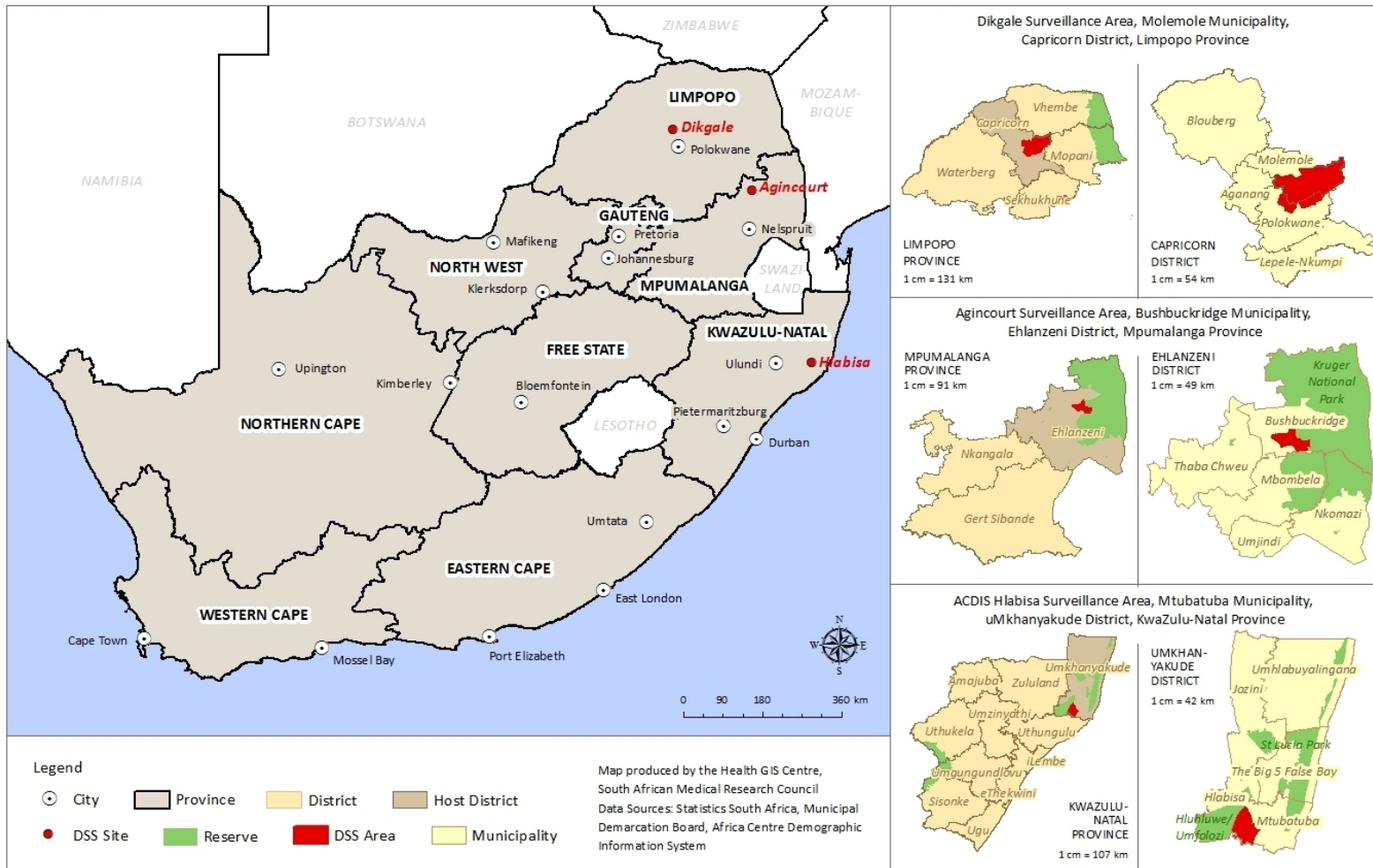


Figure 1.3: Provincial delineations of South Africa, with geographic indications and insets of three INDEPTH surveillance sites.

Source: Map produced by the Health GIS Centre, South African Medical Research Council. Administrative boundaries sourced from the Municipal Demarcation Board, 2011; Settlement level data sourced from Statistics South Africa, 2011.

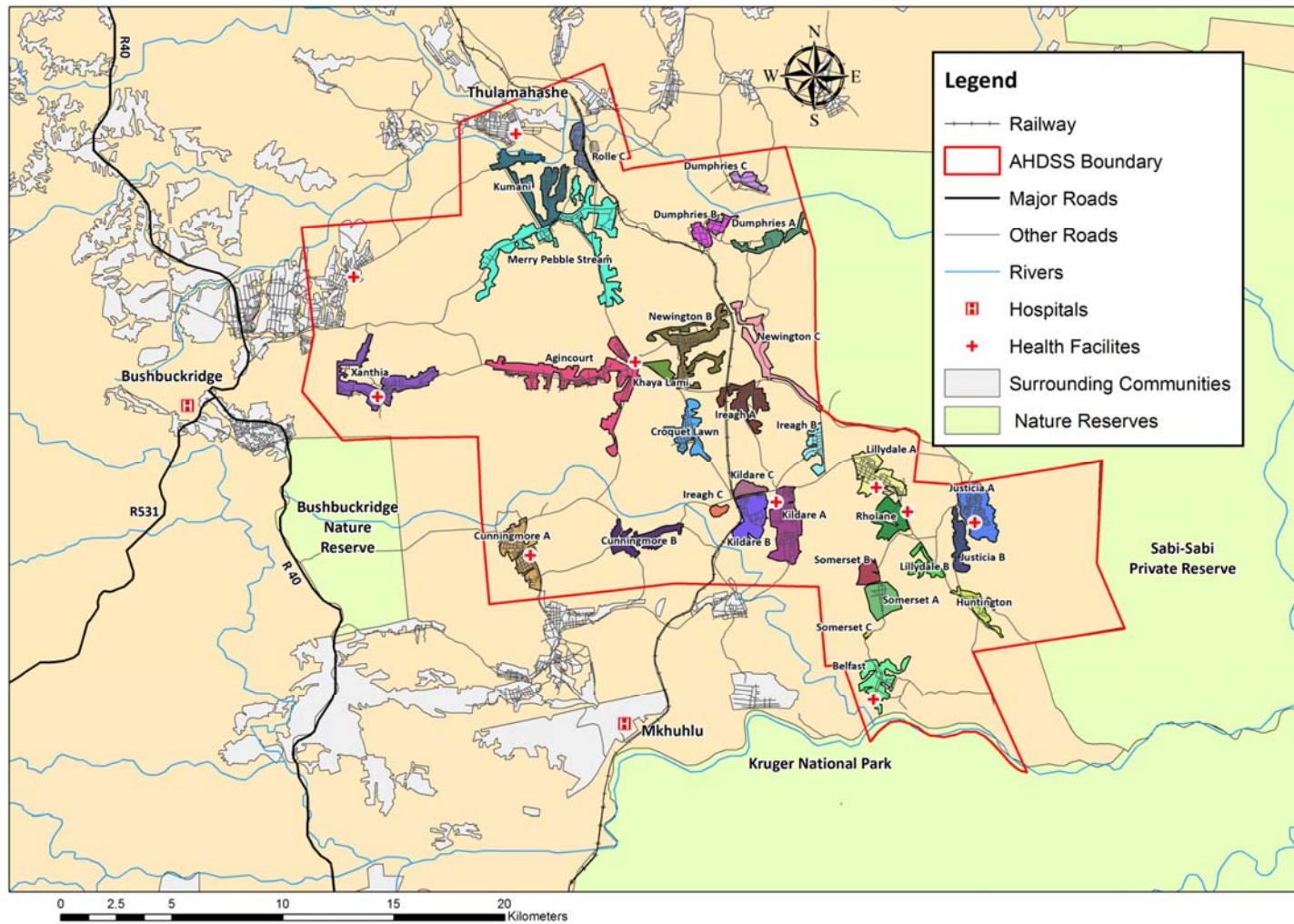


Figure 1.4: Map indicating the location of the villages in the Agincourt HDSS site (with red boundary) in north-east South Africa.  
 Source: Map created by Paul Mee, GIS Unit, MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt).

## **1.9 Availability, utility and quality of mortality data in South Africa**

While an effective CRVS system remains the ideal data source for continuous and universal mortality measurements,<sup>5</sup> providing critical evidence for public health planning and prioritizing,<sup>8</sup> few countries with the greatest need for vital-event and cause-of-death data have the capacity and resources to obtain and/or maintain these systems.<sup>5</sup> This statement is relevant for many countries in Africa: for the period 1995-2004, for example, 7% of the population in Africa lived in countries with >90% completeness in death registration. For the 55-year period between 1950 and 2005, 6 of 46 WHO member countries in Africa, representing 8% of the continent's population, reported cause-of-death statistics at least once to the WHO.<sup>9</sup> For the more recent 1990 – 2003 period, 4 of 46 WHO member countries reported civil registration cause-of-death data to the WHO.<sup>8</sup>

In a continent often reported as lacking the basic data to infer levels and trends of mortality, the 1990s in South Africa, under a new, democratic dispensation, brought about major political and social transformations. One of these was a commitment to transforming the civil registration and vital statistics systems into an information system that effectively serves civil record-keeping and public-health planning.<sup>106</sup> Attempts have also been made to collect mortality data from households in population censuses and selected national household sample surveys. At the same time, the INDEPTH Agincourt HDSS was established in rural north-east South Africa, which paved the way for another two field sites in other provinces (Figure 1.3), providing an opportunity for systematic and regular monitoring of vital events in these geographic areas.

### **1.9.1 Collection of death and cause-of-death data in South Africa**

South Africa's national CRVS system and three INDEPTH study sites are the only sources of non-sampled mortality data for all ages and all causes, with large continuous data sets that include causes of death. Deaths occurring in the INDEPTH site populations are captured and processed into the individual INDEPTH surveillance systems, a surveillance process and operational system separate and distinct from the civil responsibility of reporting a death into the government's national vital registration system. The longest-running of these study sites, with over 20 years of mortality data, is the Agincourt HDSS, referred to in Section 1.5, and mapped in Figure 1.4.

### 1.9.1.1 Registration of deaths into the civil registration system

The Births and Deaths Registration Act, 1992 (Act no. 51 of 1992) provides the legislative mandate for compulsory notification and registration of all births and deaths in the country.<sup>107</sup> A notice of death must be given by an informant within 72 hours of the death. A new death notification form, the BI-1663, was promulgated in 1998, replacing the BI-7 form. With inputs from public health and mortality researchers, changes were introduced in the form with a view to enhance death registration, collect locally relevant data for public health planning, and achieve compliance with WHO standards for the certification of causes of death.<sup>106,108</sup> Information required for registering the death is collected on the first page. Selected socio-demographic and health data (such as smoking and pregnancy status) of the deceased, as well as details of the cause of death, are collected on the second page. In an effort to emphasize confidentiality of the cause and personal details on page 2, and to improve the quality of the details provided, the completed page 2 is required to be sealed in an envelope, stapled to the first page, and used for statistical purposes only, not for the vital registration process.<sup>106</sup>

In March 2014, regulations came into effect that introduced another new death notification form, the DHA-1663.<sup>109</sup> The main difference between the BI-1663 and DHA-1663 relates to the causes of perinatal deaths. The new DHA-1663 form has a separate section to record both the maternal factors and infant factors resulting in perinatal deaths.<sup>50</sup>

The death notification and registration process to be followed after a death or stillbirth is graphically presented in Figure 1.5. In cases of natural deaths, the act requires the medical practitioner to complete form BI-1663 or DHA-1663. In the case of a stillbirth, the death notification form also makes provision for a registered professional nurse to complete the form. If no medical practitioner or professional nurse is available, within a reasonable distance, to certify the death, as may happen for example in remote rural areas, a Death Report, form BI-1680, or DHA-1680 (the latter form was regulated in 2014), may be completed by a tribal leader or headman, to certify the death and to provide a description of the circumstances that led to the death.<sup>110,111</sup> The Death Report (BI-1680/DHA-1680) can only be issued by a person authorised by the Department of Home Affairs to do so. The Death Report is then sent to the nearest office of the Department of Home Affairs where an official checks the form, and once satisfied that the death had a natural cause, and that there was no medical practitioner to certify the death, the official will transfer the information to the BI-1663/ DHA-1663 form.

On receipt of these forms by the Department of Home Affairs, the Department will issue the informant with a death certificate (Form DHA-5) or a proof of notice of death (Form DHA-1577), and a burial order (From DHA-14A).<sup>109</sup> The death is then entered into the national civil registration system, and, in the case of South African citizens and permanent residents whose births had been registered, the National Population Register will be updated.<sup>110,112</sup>

If the medical practitioner is not in a position to certify that the deceased died exclusively as a result of natural causes, a forensic pathologist or district surgeon must certify in Section D of form BI-1663/Section C of Form DHA-1663 that a post-mortem examination has been conducted, and must then certify whether the cause of death was natural, unnatural or under investigation. Unnatural deaths, in terms of the Inquests Act of 1959,<sup>113</sup> are subject to medico-legal investigation. A forensic autopsy is required to establish the cause of death, and an inquest must be held. The results of the inquest are then sent to the Department of Home Affairs, where the final death certificate is issued.<sup>110</sup>

#### **1.9.1.2 Cause-of-death information from death notification forms**

The death notification forms are collected from the Department of Home Affairs by Stats SA for capturing, processing, and the publication of selected socio-demographic variables and causes of death. At Stats SA, labels of unique identifiers are pasted on each form, and socio-demographic variables and causes of death are captured and analysed.<sup>114</sup> Following the example of the international form of the medical certificate of cause of death recommended by the World Health Assembly,<sup>115</sup> the cause of death is entered in two parts. A sequence of conditions is entered in Part 1, beginning with the immediate cause of death on line (a), which is due to the condition on line (b), which is due to the condition in line (c), which, in turn, is due to the underlying cause of death on line (d). Other significant conditions contributing to death, but not resulting in the underlying cause, must be entered in Part 2.

Since 1997, nosologists at Stats SA code the causes according to a ‘what you see is what you code’ or non-interpretive principle,<sup>50</sup> and according to ICD-10.<sup>115</sup> For the 2011 cause-of-death data, 4-character coding was employed where sufficient details were provided, and 3-character coding elsewhere.<sup>50</sup> Underlying causes are derived automatically, using the Automated Classification of Medical Entities software (ACME), the current version of which is ACME 2011. The programme applies the WHO rules for the selection of the underlying cause of death, and provides an editing user-interface for cases that need to be resolved manually, such as causes

which are considered by the system as rare.<sup>50,114</sup> During the 2011 data processing period both ACME and an additional software programme, IRIS, were used for comparing the results from each.

Information obtained from death notification forms is captured and analysed on an ongoing basis by Stats SA. Stats SA then produces death and cause-of-death statistics for South Africa on an annual basis.

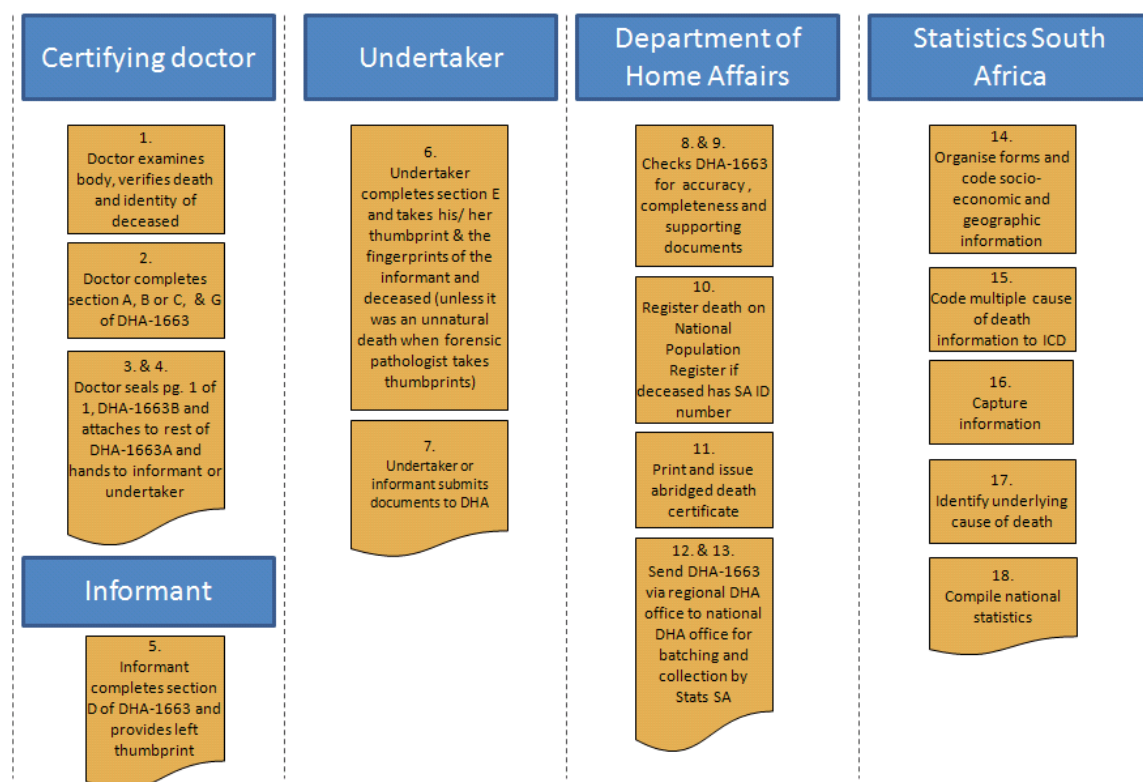


Figure 1.5: The death notification, certification, and registration process to be followed after a death or stillbirth

Source: Statistics South Africa<sup>116</sup>

## 1.9.2 Collection of death and cause-of-death data in Agincourt HDSS

### 1.9.2.1 Recording of deaths in the Agincourt HDSS

The HDSS population includes all persons resident in the study site and, hence, no sampling is required. The population also includes people linked as temporary migrants to the households. Mortality, fertility and migration data are based on a comprehensive registration system, which began with a baseline enumeration of the whole population in 1992. This has been followed by



a routine update of vital events, which involves repeat returns to all households in the population. Four updates were done between 1992 and 1998, and annual updates have occurred from 1999 to present. Routinely measured variables include births, deaths, inward and outward migrations, health-seeking practices, and education.<sup>105</sup>

A trained fieldworker visits each household unit in the study site, and interviews the most knowledgeable respondent available after verbal consent has been obtained. Information regarding individuals is checked and updated for all household members. Existing data are verified, and all new vital events experienced by each member of the household since the previous update are recorded, including migration details, changes in union status, and deaths. Data quality checks are done by the duplicate surveying of a random sample of households and rigorous self-, cross- and quality controller-checking of census forms at field and office levels.<sup>105</sup>

#### **1.9.2.2 Cause-of-death information from Agincourt HDSS verbal autopsy questionnaires**

Full verbal autopsies are conducted on every death recorded during annual census updates, using a locally-validated instrument. To assist in the process of establishing the most probable cause of death, an interview by a specially trained fieldworker is conducted 1-11 months after a death, with the closest caregiver of the deceased in the primary language of the caregiver. Fieldworker feedback, on-site supervision, and the regular review of completed verbal autopsy questionnaires are conducted with a view to improving the quality of the collected information.<sup>104,105</sup>

These questionnaires are independently evaluated by two medical doctors who assign a probable immediate, contributory, and underlying cause to each death. Equivalent diagnoses from these clinicians are accepted as the probable cause of death. Where these causes differ, the two clinicians discuss the case. If consensus is not achieved, a third clinician, blind to earlier findings, assesses the details. If the third assessment is in agreement with one other, that cause is accepted as the probable cause of death. If no agreement, the cause is coded as “ill-defined”. Additionally, final diagnoses are compared with hospital reference diagnoses, where available. ICD-10 conventions are followed in the attribution of causes of death.<sup>105</sup> More recently, cause-of-death data have also been analysed using probabilistic modelling with the InterVA software.<sup>117,118</sup>

Information on deaths—obtained from annual updates of resident status and vital events for every field-site member, and causes of deaths as obtained via verbal autopsy interviews—are captured and analysed on an ongoing basis by Agincourt HDSS staff. Mortality statistics are made

available regularly through online data repositories, or when requested, by tailored data extraction from the field-site database.

### **1.10 Rationale for the research**

The importance of valid and reliable mortality data as a key input for health decision-making, development programmes, and research agendas, is widely acknowledged (Section 1.1). Despite this, few low- and middle-income countries have such data, and even fewer have assessed the quality of their mortality data to enable an understanding of local data quality.<sup>8,9</sup> Although periodic, independent evaluations of mortality statistics by countries themselves are particularly useful to identify biases in local data and make locally-relevant recommendations to improve mortality data,<sup>9-11</sup> such an evaluation of national mortality statistics has not yet been conducted for South Africa. Neither has a review been conducted of the sources of mortality data in South Africa since that conducted by Bourne,<sup>12</sup> who covered CRVS mortality data sources from 1910 to 1992.

Two WHO comparative assessments rated the quality of South Africa's 1996 mortality data as low.<sup>8,9</sup> With the transition to democracy during the 1990s, governance and public services, including civil registration, have undergone major legislative, policy, and practical transformations.<sup>106</sup> Key events facilitated improvements in the coverage of civil registration and vital statistics, including the passing of the Births and Deaths Registration Act of 1992 that abolished differential civil registration based on race and rural/urban residence.<sup>119</sup> A range of actions were furthermore initiated to improve birth and death registration practices and the quality of vital data. These included the introduction of a new death notification form that complies with WHO standards; the establishment of provincial task teams to assist in implementing the new form; the distribution of guidelines on birth and death registration; and issuing letters about relevant guidelines and new procedures to all registered doctors.<sup>106</sup> In addition, targeted capacity strengthening was undertaken at Stats SA to improve the compilation and production of annual vital statistics for the country.<sup>119</sup>

It is of considerable interest to ascertain whether or not these events and investments have had an impact on the availability, coverage, and quality of more recent CRVS data.

A small number of small-scale studies have examined the quality of cause-of-death data from civil registration in South Africa, using medical records to evaluate the accuracy of certified

cause attribution.<sup>120-123</sup> However, problems have been identified with the quality of the information on medical records,<sup>120,123</sup> pointing to the need to ascertain whether other mortality data sources may serve as suitable material for evaluating and validating the quality of certified causes of death. These small-scale studies involved urban deaths only, while no study to-date has assessed the quality of civil registration cause-of-death data from a rural area in South Africa. In addition, despite the critical need of district health managers to know and manage the health care needs of the district population they serve, mortality profiles at this level are biased and come with strong warnings of invalid data.<sup>13</sup> These gaps in knowledge call for research to start filling the void.

From a wider perspective, given the dire lack of cause-of-death data in Africa,<sup>8,9</sup> South African cause-of-death data have previously been used widely by research and international development agencies as the basis for estimating cause-specific mortality in many African countries.<sup>124,125</sup> This adds to the critical and wider importance of understanding the quality of mortality data in South Africa.

Hence, there are important local and international reasons to assess the quality of recent civil registration mortality data as the most extensive source of mortality data in the country; to identify, document, and review alternative mortality data sources; and to explore whether and how the quality of civil registration cause-of-death data can be assessed in a comparative study to give insight into the quality and utility of certified cause-of-death data.

### **1.11 Aim and research questions**

The primary aim of the research was to assess the quality of mortality data collected and processed in the national CRVS system of South Africa. Recognizing the time lapse since the previous review of mortality data in the country, and acknowledging the value of alternative sources of mortality data to complement and evaluate the quality of civil registration data, a secondary aim was to identify, document and review alternative mortality data sources, and the availability of data from these sources, in the country.

For the realization of these aims, the following research questions are investigated:

1. *What mortality data sources are available in democratic South Africa, and what are the limitations, strengths, uses, and public availability of data from these sources?*
2. *What criteria can be used for evaluating national death and cause-of-death statistics from civil registration in South Africa?*
3. *What is the quality of national mortality statistics from civil registration for the period 1997 – 2007?*
4. *Would a linkage study between the CRVS and the Agincourt HDSS mortality data collections be possible, and if so, would such linkage be of good quality and useful?*
5. *What can the mortality data collected in Agincourt HDSS indicate about the completeness of death registration and factors associated with registration?*
6. *What can the verbal autopsy data collected in Agincourt HDSS indicate about the quality and utility of cause-of-death attribution in the civil registration system?*
7. *Are the matching study results useful for developing adjustment factors for cause-of-death attribution at the district level?*

## **1.12 Outline of thesis**

*Chapter 1* describes the importance and utility of mortality data for decision-making in public health and measuring indicators of population health, and highlights different sources of mortality data. Given the strong focus on mortality data from CRVS systems throughout the thesis, their historical development, attributes, and advantages for mortality measurement are described. As further background to the content chapters, selected features of the study settings referred to in the thesis, i.e. South Africa and the Agincourt HDSS, are provided. This is followed by the rationale, aims and research questions to be addressed in the thesis.

Mortality data sources for South Africa were identified and the data collection instruments, code lists, relevant legislation, and publications of such data were reviewed. Emanating from this review, *Chapter 2* presents an assessment and critical appraisal of the characteristics, public availability, uses, strengths, and limitations of different mortality data sources, and provides recommendations for improving mortality measurement (research question 1).

Wide recognition is given in the literature to the value of national civil registrations as a preferred source of national and sub-national mortality data. A concise overview of the international development of the recording of vital events has been given in Section 1.3.1, and *Chapter 3* provides an overview of the development of vital-event recording in South Africa. South Africa's overview outlines the many challenges faced during the 1900s in the country—the transformation of the CRVS system during the 1990s after democratization, serving to contextualize official mortality data recording and compilation, and the rutted road to its improvement. Seeking suitable criteria to evaluate the quality of national CRVS mortality data for South Africa, a summary is then provided of the data attributes and criteria developed over time for such evaluation at the country level (research question 2). Four data attributes and nine criteria, used successfully in similar country assessments elsewhere,<sup>10,126</sup> are used to evaluate and form an opinion of the quality of local civil registration mortality data for South Africa (research question 3).

*Chapter 3* reveals rather large differences in the estimates of completeness of national death registration in South Africa. Additionally, cause-of-death information from national civil registration is identified as a serious problem area in *Chapter 3*. Given the importance of completeness of death registration and the validity of cause-of-death attribution, *Chapters 4* and *5* describe a sub-study aiming to link and compare mortality data from a) Stats SA's database of national civil registration data, and b) the Agincourt HDSS surveillance database, for the same individuals who died during the period 2006-2009 (research question 4), with the purpose to: a) quantify the level of completeness of death registration in the civil registration system (research question 5), and b) assess the level of agreement in cause attribution between these two data sources (research question 6). Given the critical need among district managers to know the cause-of-death profile of their district, *Chapter 5* ends with an exploration of the utility of applying adjustment factors from the Agincourt HDSS/Stats SA linkage study to the cause-specific mortality fractions of the surrounding district (research question 7).

The thesis concludes in *Chapter 6* with a synopsis of findings, conclusions and recommendations related to the research questions. In *Chapter 7* the thesis findings are consolidated in the context of the latest directions in the literature.



## 2 Chapter 2: Identification, review and documentation of post-apartheid mortality data sources for South Africa

---

### 2.1 Introduction

Accurate mortality data by age, sex, and cause are essential for planning, monitoring and assessing public health programmes and policies.<sup>9,14</sup> In countries with fully-functional statistical systems, the necessary information for such data is usually derived from well-developed CRVS systems that include medically-certified causes of death, as well as reliable population counts, by various jurisdictions, from regular censuses or a population register.

Such information, however, is often not available in countries with poorly-developed statistical systems where international comparative studies<sup>8,9</sup> have indicated challenges, such as incomplete registration of births and deaths; misreporting of the age at death; limited certification of causes of death by a physician; large proportions of causes of death recorded as ill-defined or non-specific causes; and coverage and age-reporting errors in population numbers that are to serve as denominators.<sup>14</sup> Mahapatra *et al* found in 2007 that only about 30% of the world's population lived in areas with over 90% completeness of death registration,<sup>9</sup> while Obermeyer *et al* reported in 2010 that in most African countries, less than a quarter of deaths had been registered.<sup>127</sup> Of the countries in Africa and South-East Asia, Setel *et al* reported in 2007 that only 2% have had complete death registration, and half recorded no cause-of-death data.<sup>5</sup>

In countries with such challenges, alternative data sources and methods are necessary for the collection of mortality and other vital data to provide national and sub-national estimates of the levels and causes of mortality. A number of such data sources, techniques, and methods for collecting mortality data are referred to in Section 1.2. The methods include population censuses, household sample surveys, and demographic surveillance in sentinel sites. Hill *et al* report that these data collection strategies have been developed as “interim substitutes” for complete civil registration, largely as cost-effective alternatives for deriving demographic estimates in poor countries where civil registration is lacking.<sup>14</sup>

The potential contributions of these methods to measure births, child mortality, adult mortality, and causes of death have been summarised by Hill *et al*, and are shown in Table 2.1. Strengths and weaknesses are indicated for each approach to data collection. Particular strengths are indicated for data from civil registration. Other approaches do not meet all the data needs, and

often require adjustment, or to be supplemented with other assessment methods, in order to produce satisfactory estimates.<sup>14</sup> These opinions have been echoed by several other reports.<sup>2-4,6,15,18,19,43,128,129</sup>

Table 2.1: Contributions of different approaches to measuring key mortality and population health indicators

	Level of estimate	Civil registration system	Demographic surveillance sites	Sample registration systems	Population censuses	Household sample surveys
Births	National	Yes	No	Yes	Maybe*	Yes
	Differentials	Yes	Limited	Limited	Maybe*	Limited
Child mortality	National	Yes	No	Yes	Yes†	Yes
	Differentials	Yes	Limited	Limited	Yes†	Limited
Adult mortality	National	Yes	No	Yes	Maybe*‡	Weak¶
	Differentials	Yes	Limited	Limited	Perhaps*‡	No
Cause of death	All	Yes	Yes§	Yes§	Maybe§	Yes§**

\* With assessment and possible adjustment; methods do not always work.  
† For a recent period by indirect methods.  
‡ For an inter-censal period.  
¶ Methods measuring parental survival or sibling history.  
§ With verbal autopsy.  
\*\* For child deaths identified by a full birth history.

Source: Table populated with information from Hill *et al*<sup>14</sup>

## 2.2 Under-developed CRVS systems in most African countries, but transformation in South Africa

While CRVS systems remain under-developed in most African countries,<sup>4,5,9</sup> the transition to democracy during the 1990s in South Africa brought about major transformations in legislation, governance, and public services and practice.<sup>106,119</sup> These included transformation of the civil registration processes, and vital statistics compilation and production, to advance mortality data availability and utility (compare key events and initiatives under “Transformation during the 1990s affecting civil registration” in Section 2.4).

In addition, mortality data items have been included in population censuses and national household surveys – the October Household Surveys programme in 1993 marking the beginning of annual national data collection for demographic, development and poverty indicators. A number of other survey, surveillance and facility-based reporting initiatives were also initiated, which include data items with potential for mortality analysis (Section 2.4).



### **2.3 A next review of mortality data sources in South Africa**

With the previous review of mortality data sources for South Africa, conducted twenty years ago, and only including official sources from the CRVS system,<sup>12</sup> and with the number of post-apartheid initiatives that now provide opportunities for mortality measurement, it is timely to review and document what mortality data sources are currently available, and to assess the strengths, limitations and public availability of the data. Against this background, Section 2.4 contains a published review that aims to answer the first research question of the thesis:

*Research question 1:* What mortality data sources are available in democratic South Africa, and what are the limitations, strengths, uses, and public availability of data from these sources?

The article reports on the study design of the review; documents mortality data sources available in democratic South Africa; and assesses and critically appraises the characteristics, public availability, uses, strengths, and limitations of these data sources. Data limitations and strengths are discussed in the context of their use for particular mortality indicators and analyses, and recommendations are made for improving data quality and reliability of mortality measurement in South Africa. Finally, a short overview discussion and conclusion about the chapter is provided in Section 2.5.

### **2.4 Paper 1: “Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa”**

# Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa

Jané Joubert<sup>1,2\*</sup>, Chalapati Rao<sup>1</sup>, Debbie Bradshaw<sup>2</sup>,  
Rob E. Dorrington<sup>3</sup>, Theo Vos<sup>1</sup> and Alan D. Lopez<sup>1</sup>

<sup>1</sup>School of Population Health, University of Queensland, Herston, QLD, Australia; <sup>2</sup>Burden of Disease Research Unit, South African Medical Research Council, Tygerberg, South Africa; <sup>3</sup>Centre for Actuarial Research, University of Cape Town, Cape Town, South Africa

The value of good-quality mortality data for public health is widely acknowledged. While effective civil registration systems remains the ‘gold standard’ source for continuous mortality measurement, less than 25% of deaths are registered in most African countries. Alternative data collection systems can provide mortality data to complement those from civil registration, given an understanding of data source characteristics and data quality. We aim to document mortality data sources in post-democracy South Africa; to report on availability, limitations, strengths, and possible complementary uses of the data; and to make recommendations for improved data for mortality measurement. Civil registration and alternative mortality data collection systems, data availability, and complementary uses were assessed by reviewing blank questionnaires, death notification forms, death data capture sheets, and patient cards; legislation; electronic data archives and databases; and related information in scientific journals, research reports, statistical releases, government reports and books. Recent transformation has enhanced civil registration and official mortality data availability. Additionally, a range of mortality data items are available in three population censuses, three demographic surveillance systems, and a number of national surveys, mortality audits, and disease notification programmes. Child and adult mortality items were found in all national data sources, and maternal mortality items in most. Detailed cause-of-death data are available from civil registration and demographic surveillance. In a continent often reported as lacking the basic data to infer levels, patterns and trends of mortality, there is evidence of substantial improvement in South Africa in the availability of data for mortality assessment. Mortality data sources are many and varied, providing opportunity for comparing results and improved public health planning. However, more can and must be done to improve mortality measurement by improving data quality, triangulating data, and expanding analytic capacity. Cause data, in particular, must be improved.

Keywords: *mortality data sources; South Africa; civil registration; census; surveys; surveillance; data availability*

Received: 2 August 2012; Revised: 18 November 2012; Accepted: 21 November 2012; Published: 27 December 2012

The importance of mortality data for public health and development is widely acknowledged and publicly recognised since the 1600s in Graunt’s Bills of Mortality (1). Globally, the importance of this continues to be reflected in a number of initiatives, such as the inclusion of mortality as a component of the Human Development Index (HDI) (2) and the expression of Millennium Development Goals (MDGs) 4, 5, and 6 in terms of mortality (3, 4). Although global, these indicators and targets firstly need to be measured at

national and sub-national levels. However, the required data for their measurement are often not readily available, particularly in developing countries and the world’s poorest populations. In 2007, the *Lancet* “Who Counts?” series confirmed that few countries in greatest need of vital-event and cause-of-death data have the capacity to obtain these (5).

Civil registration systems, population censuses, and surveys are highlighted as principal sources for measuring mortality (6–8). Recent reports also include sample

vital registration and demographic surveillance systems (9). Civil registration with high coverage and accurate medical certification and coding of the cause of death remains the ‘gold standard’ source of continuous mortality data. However, only approximately 30% of the world’s population live in areas with >90% completeness of death registration (10). In most African countries, for example, less than one-quarter of deaths are registered (11). Only 2% of the countries in Africa and South-East Asia have complete death registration data and half of the countries in these regions record no cause-of-death data (5). Other challenges include incomplete birth registration, misreported age at death, delays of death records under medico-legal investigation, and delays in releasing the data. For cause-of-death data, limitations include omission of the cause on the death certificate, inclusion of a cause, but not certified by a physician, inclusion of a physician-certified cause, but recorded as ‘ill-defined’ or ‘undetermined’, or a physician-certified cause, but inadvertently misdiagnosed or advertently euphemised or misclassified.

In such settings, mortality researchers rely on available alternative data collection systems such as censuses, surveys, and sample and small-area systems (9). Censuses provide essential data on population denominators and the opportunity to collect national mortality data through direct recall of household events in defined time periods, or indirect measures by collecting information on parental, spousal, sibling or child survival. Accurate information can provide estimates of childhood and adult mortality by population sub-groups and small areas. Potential weaknesses include: recall bias, event omission due to population undercount, household collapse, or mother’s death, question limitations due to the length of the questionnaire, inaccurate age reporting, and the relative infrequency of enumerations at 5- or 10-year intervals. Household surveys have played a significant role over the past four decades in assessing childhood mortality through birth histories in areas with inadequate civil registration (12). Such surveys can also inform adult mortality assessment through questions on parental, sibling and spousal survival, and questions about household deaths in a specified period. Household surveys have the advantage that they can be carefully monitored for data quality, but can also be compromised by insufficient fieldwork training and supervision. Further limitations include the inability to make estimates for areas below provincial/state level, recall bias, and missing births and deaths due to discord- or death-related household disintegration or zero-survivor families (11).

Health and demographic surveillance sites (HDSSs) in defined geographic area, such as those co-ordinated by the International Network for the Demographic Evaluation of Populations and their Health in Develop-

ing Countries (INDEPTH), constitute another approach to collecting mortality data and fills an important gap in mortality assessment in many low- and middle-income countries (13). Births and deaths information are critical in each HDSS, actively identified through regular visits by trained fieldwork staff to all households in the HDSS, recording events in the period since the previous visit. With most deaths in these sites not occurring in health facilities, the cause of death is typically assessed via a verbal autopsy instrument and interviews with close kin or carers of the deceased. This is commonly followed by physician assessment and consensus acknowledgement of the probable cause of death. As a promising alternative, computerised procedures have been developed with Bayesian probabilistic modelling over the past decade, leading to the InterVA suite of models and culminating with the recent, freely available InterVA-4 model (14). Though typically intensive and thorough in pursuing vital data, and usually achieving complete, or close to complete, death recording, HDSSs are generally restricted to small geographic areas and their populations are not necessarily representative of the national population. The lay report of the circumstances leading to death, often reported a few months after the event, can be a further limitation.

While civil registration systems remain underdeveloped in most African countries, recent political and public services transformation in South Africa, along with determined efforts by researchers, have focussed on enhancing the civil registration system and advancing mortality data availability from vital statistics compilation (5, 15). Additionally, mortality data items have been included in local censuses, national surveys, HDSS data collections, and condition-specific registry, disease notification and mortality audit programmes. Each of these sources has relative strengths and limitations, and an understanding of the characteristics of these sources and the quality of the data produced by them is important to guide the utilisation of their data and exploit potential complementary properties. As South African cause-specific mortality patterns have been used to model mortality in many sub-Saharan Africa countries (16, 17), a closer look at the availability and quality of such data may be useful beyond its local value.

This article aims to review civil registration and alternative data collection systems for mortality estimation in post-democracy South Africa. Our objectives are to document these data sources, report on public availability and possible complementary uses of data from these sources, discuss data limitations and strengths in the context of their use for particular mortality indicators and analyses, and make recommendations for improving data quality and the reliability of mortality estimation.

## Study design

Background information of mortality data sources, data availability, and possible complementary uses of mortality data were assessed by reviewing blank survey and census questionnaires, death notification forms, death data capture sheets, and patient treatment and clinic/hospital cards; legislation; electronic data archives and databases and web-based data repositories; and related information in scientific journals, research reports, statistical releases, government reports, working papers, and books.

In the context of the Africa Programme on Accelerated Improvement of Civil Registration and Vital Statistics (APAI-CRVS) (18), and the global importance attached to civil registration systems for mortality assessment (19), the review includes a particular focus on the development of the country's civil registration and vital statistics system over the past century. Additionally, summary background information was extracted and data availability assessed for three post-democracy population censuses, a number of national household surveys, one rapid mortality surveillance system, and three HDSSs. Tuberculosis, cancer, and injuries are major contributors to the country's mortality burden, and key information was obtained of three facility-based surveillance systems related to these causes. Additionally, given the international pressure to provide reliable information about child and maternal health to inform MDGs 4 and 5, and the potential to complement vital registration data, summary information was extracted about three facility-based mortality audit programmes for peri-natal, child and maternal deaths.

## Mortality data sources in South Africa

### *Civil registration and vital statistics systems in South Africa: a century of challenges*

Starting in 1867, a series of laws on birth and death registration were enacted at sub-national level. Half a century later, the Births, Deaths and Marriages Registration Act of 1923 promised the possibility of uniform registration practice across the country (20–23). The act remained in place until replaced by the Births and Deaths Registration Act of 1992 that, unlike the 1923 act, required complete coverage of all people and all geographic areas (24). Between 1923 and 1992, however, the civil registration and vital statistics systems faced numerous challenges. Neither the single, national act of 1923, nor the establishment of the national statistical office in 1914 brought about an inclusive, comprehensive civil registration system. Instead, it was a partial registration system, covering selected segments of an ideologically contrived population, based on 'homeland'/common land, rural/urban, and population-group differentiation.

Over time, geographic and population fragmentation became further entrenched, formalised by legislation, significantly stunting the civil registration system through the under-registration of deaths in large parts of particularly the majority, Black African,<sup>1</sup> population group. Such legislation included the restriction of land leasing and ownership among Black Africans to 'homelands' or designated reserves, mainly in under-developed rural areas with limited economic possibilities (25), and dividing the non-'homeland' or common area into freely accessible (rural) and restricted (urban) areas. This residence and movement control confined the majority of Black Africans to rural residence. With the Bantu Homelands Citizenship Act of 1979, Black Africans were required to become a citizen of one of 10 'homelands' (22, 23, 26). After granting independence to four 'homelands', they became excluded from the South African statistical system and responsible for generating their own vital registration information (21). However, these states were largely incapable of doing this (27) and, in the remaining six 'homelands', several laws and circumstances constrained civil registration (22).

The Population Registration Act of 1950 made provision for the compilation of a manual population register that, counter-intuitively, played a minor role in producing vital statistics (22, 23). In 1972, a computerised population register was initiated, but did not capture the civil details of Black Africans until 1986 (21, 22). The civil registration was also affected by the 1923 act under which death registration for rural-living Black Africans was voluntary (20, 21), implying that during the 1920s approximately 86% of Black Africans (28) were under no obligation to register any death. Medical certificates were required for urban but not rural deaths (20), further inhibiting vital statistics in South Africa.

In addition to these inhibiting effects and the lack of registration infrastructure and resources in most of the 'homeland' areas, reporting of vital events was also probably restrained by the many disruptive effects of forced removal and resettlement (25, 29–32) on people's motivation and means to do so (25, 31, 33). Moreover, legislation, including civil registration legislation, itself impeded vital registration for decades as one act ensured that, for the majority population, rural living was

<sup>1</sup>'Population group' is used as a collective term that may elsewhere be called 'race' or 'ethnic group'. The use of the terms 'Black African', 'Coloured', 'Indian/Asian', and 'White' are not intended to denote biological difference, neither to support a racial or ethnic classification system. Under the Population Registration Act of 1950, South Africans were classified into these groups. The classification was associated with disparities in many spheres of life, including health. To acknowledge this impact, and to help track progress in redressing past inequalities based on the classification, mortality and other data are still classified by these terms, although individuals self-classify. For this reason, and because a historical perspective is presented at times, it is necessary to reference these terms in this article.

compulsory while, at the same time, another ensured that rural death registration was not (21). With completeness of Black African death registration estimated to have ranged from about a quarter of all deaths during the late 1960s to about half in the mid-1980s (34), the vital-event details of this group in particular became severely under-represented in vital statistics, limiting the use and generalisability of civil registration data considerably. (21, 22). Thus, although official mortality statistics have been collated and published since 1910 (20), the above events led to large biases in vital statistics data, and numerous barriers to producing reliable, representative and timely mortality statistics.

#### Transformation during the 1990s affecting civil registration

During the 1990s, however, under a new, democratic dispensation, major and rapid political and social transformation ensued in all spheres of governance, including a commitment to transforming the civil registration and vital statistics systems into an information system that effectively serve civil record-keeping and public health planning. Bah (23) identified three key events during the 1990s which held new promise for vital registration coverage and content: 1) the passing of the Births and Deaths Registration Act of 1992 (24), leaving no scope for optional or differential registration; 2) the adoption of the interim Constitution of South Africa in 1993 (35), ensuring the consolidation of all geographic segments into one geo-political unit and, therewith, the centralisation of the civil registration system under one entity; and 3) the agreement among three key role players – the Department of Health, Department of Home Affairs (DHA), and the official national statistical agency, Statistics South Africa (Stats SA) – to establish, in collaboration with health researchers, a joint technical committee tasked to enhance civil registration and improve the vital statistics system (23, 36). These events spawned a range of initiatives to increase the registration of deaths and improve the quality of vital-event data (23, 37, 38), including the introduction of a new death notification form to bring local data in line with international standards and to achieve compliance with WHO standards for the certification of causes of death (36, 39, 40).

Institutional capacity was strengthened through study tours and visits of key government officials to civil registration/vital statistics offices in Australia, Sweden, and the United States (23). More capacity and initiatives have been developed to enhance coverage of death registration, including the establishment of provincial task teams who developed a strategy to enhance registration, and facilitated the introduction of the new death notification form to ensure its implementation; distribution of certification and ICD code manuals to hospitals

and health professional and academic organisations for staff training; letters to all registered doctors, informing them of the new procedures and relevant guidelines; the development of guidelines by the Department of Health on birth and death registration; training of health workers in all provinces about the importance and process of vital registration; making the necessary forms for birth registration available to mothers at the time of delivery; and assisting mothers to complete and submit the forms to the DHA (21, 36, 41).

#### Capturing the effects of transformation in completeness estimation

These specific efforts, along with the political and social transformations mentioned previously, are likely to have contributed to the increasing levels of completeness of registration for both adult and child deaths over the past two decades. Estimation of completeness of registration during the 1980–90s was a very complicated task. However, meticulous, comprehensive research since the late 1980s has carefully sought to understand the merit of different local mortality data sources and the applicability of different indirect methods in the South African mortality data context over time (27, 42–44). Application of the Bennett and Horiuchi's Synthetic Extinct Generations method (45), for example, to deaths from Stats SA's vital registration database and the Population Register, relative to population estimates from the ASSA600 AIDS and Demographic Model of the Actuarial Society of South Africa, estimated that the level of adult death registration improved from 85% to about 90% for the period 1996–2000 (43).

For the period 1996–2006, the estimated level of death registration improved from 43 to 89% for infants; from 44 to 78% for children under 5 years; and from 43 to 57% for children aged 1–4 years (46). These childhood estimates were derived using a multi-stage method described in Darikwa and Dorrington (46), using registered death data from civil registration; data from the 2007 Community Survey (child deaths over the past 12 months as reported by households, children ever born/children surviving data, and data on the survival of the last child born to women aged 12–49 years); 2001 Census (reported household deaths); and data from previous research (47–50) based on the 1998 South Africa Demographic and Health Survey (SADHS) and 1996 Census. Completeness levels between 1996 and 2006 were determined by assuming that the completeness of death registration follows a logistic trend over time; that completeness in any age group did not decrease over the period; and that the trend in completeness is smooth (i.e. fluctuations in the data are not due to fluctuations in completeness). A logistic curve was then fitted (46, 51).

The completeness in infant death registration has increased particularly rapidly since 2001, most probably

resulting from the government's intensified efforts to register births and deaths, particularly in facilities, along with many infant deaths occurring before leaving these facilities. This bodes well for more accurate estimates of this important indicator, in particular the prospect of deriving infant mortality directly from vital registration data with minimal adjustment (46).

#### Mortality data and data availability from civil registration

Death notification forms are administrated by DHA. These forms are then processed by Stats SA to capture the mortality and selected socio-demographic and health data (52). From this, cause-specific mortality statistics are produced by Stats SA, coded to three-digit codes according to the tenth revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10) (53). Anonymous unit record cause data for 1997–2009 are available electronically upon request, subject to a data user's agreement. To protect the identity of the deceased, certain fields such as the date of death and place of death or residence, are not publicly available. Mortality and cause-of-death data for 2006–2008 are additionally freely available in web-based data repositories, and users can analyse and tabulate a number of variables with socio-demographic (e.g. age and sex), health status (e.g. smoking and pregnancy status), and cause data (e.g. immediate and underlying cause) with online and available software Nesstar and SuperWEB (54). Reports (online or hard copy) from 1997 to 2009 offer a variety of tabulations for all-cause and cause-specific mortality, both nationally and provincially (55). The latest reports additionally offer tabulated numbers of death and selected cause patterns by district municipality (55). Over the past decade, the production of official mortality statistics has improved considerably, and mortality data are available annually, reporting on deaths that occurred in the calendar year 2 years prior to publication.

#### National surveillance, censuses and surveys

##### Rapid Mortality Surveillance System using national data from the Population Register

The 1990s, still suffering the effects of the pre-democratic civil registration and vital statistics system, presented with a substantial time lag in the production and release of mortality data. These delays were particularly problematic amidst the rapid unfolding of an enormous HIV/AIDS epidemic and the perceived significant mortality changes in the population. An urgent need thus arose for more up-to-date mortality statistics and the continuous monitoring of more recent mortality trends. In response, the South African Medical Research Council (MRC) in collaboration with the University of Cape Town, negotiated access to death data by age and sex as recorded on the National Population Register maintained by the DHA. A project was set up at the MRC in 1999 to

capture and monitor trends in these data as rapidly as possible. The MRC database is updated monthly with death data provided electronically by the DHA, allowing mortality by age and sex to be monitored within a few months after the date of death (56, 57).

As for the vital statistics system of Stats SA, the source of the Rapid Mortality Surveillance System (RMS) is death notifications submitted to the DHA. However, there is a difference between the numbers of deaths captured by these two systems. Stats SA captures all deaths notified to the DHA, while the RMS only captures those deaths notified to the DHA which have been recorded onto the National Population Register, i.e. only the deaths of individuals with a South African birth or identity certificate (as only people with these certificates are on the Population Register). The RMS, therefore, captures fewer deaths compared to Stats SA's vital statistics system, on average about 12% less for the years 2002–2009, but more than sufficient numbers to serve the purpose for which it was developed (on average 493,000 deaths for 2000–2011) (57).

While mortality reports currently are being published with a time lag of approximately two years, the RMS remains useful for providing information about deaths within months after occurrence. Additionally, the RMS is useful for tracking changes in mortality due to the roll out of interventions such as programmes to prevent mother-to-child transmission of HIV, and provision of antiretroviral therapy.

The RMS data are received and stored by the SA MRC for continuous rapid assessment of changing trends in the deaths by age and sex. The availability of the data has been negotiated with the purpose of rapidly assessing and informing about changes as assessed by experienced mortality researchers. To inform research and policy action adequately, the data needs to be interpreted taking into account the prevailing levels of completeness of death registration, the extent of birth and death registration into the population register, and levels of population growth. Findings from the rapid mortality surveillance system are regularly reported in publicly available papers, reports, and conference presentations (43, 56–59).

##### Post-democracy population censuses: 1996, 2001 and 2011

Post-unification (1910) and pre-democracy (1994), 14 population censuses with variable coverage have been conducted in South Africa, the first in 1911. Since democracy, three censuses that covered the total South African population have been conducted, respectively, in 1996, 2001, and 2011. For all three censuses, a post-enumeration survey was undertaken to determine the degree of under- or over-count in the population. For the 1996 and 2001 censuses, the population was undercounted by an estimated 10.7% (60) and 17.6%

(61), respectively. Results from the 2011 census have not been published yet. More information is available at the Population Statistics Section of the Stats SA website (62).

#### Mortality data from censuses and data availability

Information about the types of census data collected to measure mortality in different interest groups, and the years for which such data were collected, are shown in Table 1. Census data tabulations at national and sub-national level are available on request from Stats SA. Children ever born/children surviving (CEB/CS) and parental survival data from a 10% sample of the household unit record data for the 1996 and 2001 censuses are available together with selected socio-demographic variables in web-based data repositories via Nesstar and SuperWEB (54). For 2001, deaths reported by households are also available in the 10% sample. Metadata for the census variables are available on Stats SA's 1996 and 2001 census web pages (62). Additional data input is required to calculate adult and child mortality rates from these variables. Stats SA commissioned the Centre for Actuarial Research to analyse and evaluate the mortality data collected in the 2001 census. The resultant detailed report (44) contains essential information for users on the quality of child and adult mortality data collected.

#### Post-democracy national surveys

Brief background information of five national surveys is individually given below after which mortality data

from these surveys and its availability for public use are discussed collectively. Details about the enumerated number of households and participants, and the types of mortality data collected, are in Tables 1 and 2.

#### October Household Surveys: 1993–1999

The establishment of the October Household Surveys (OHS) programme in 1993 marked the beginning of the national collection of demographic information on an annual basis. The OHS was a cross-sectional sample survey undertaken by Stats SA from 1993 to 1999, aiming to collect individual and household information that covered a range of development and poverty indicators. The OHS was replaced by the General Household Survey. The surveys were based on a probability sample of a large number of households, targeting residents in private households and workers–hostels countrywide. Fieldworkers visited sampled households and filled the survey questionnaire during face-to-face interviews (63). See Tables 1 and 2 and the Stats SA (64) and University of Cape Town's DataFirst (65) websites for more information.

#### General Household Survey: 2002–2012

The General Household Survey (GHS) has been conducted annually by Stats SA from 2002 to 2011 and was in the field until September 2012 for the next round. The GHS was instituted to monitor development indicators and development programmes on a regular basis.

**Table 1.** Post-democracy data sources for mortality analysis in South Africa by enumeration years

Data sources	Enumeration year(s)			
	Child mortality	Adult mortality	Maternal mortality	Causes of death
Vital Registration (VR)	1997–current*	1997–current*	1997–current*	1997–current* <sup>††</sup>
Rapid Mortality Surveillance System (RMS)	1998–current*	1998–current*	–	1998–current <sup>‡‡</sup>
Population census	1996, <sup>†</sup> 2001, <sup>†</sup> 2011 <sup>†</sup>	1996, <sup>  </sup> 2001, <sup>  ,¶</sup> 2011 <sup>  ,¶</sup>	2001 & 2011 <sup>**</sup>	2001 & 2011 <sup>‡‡</sup>
Demographic Surveillance Sites (DSS): > Agincourt	1992–current*	1992–current*	1992–current*	1991–current <sup>§§</sup>
> ACDIS	2000–current*	2000–current*	2000–current*	2000–current <sup>§§</sup>
> Dikgale	1996–current*	1996–current*	1996–current*	2011–current <sup>§§</sup>
Community Survey (CS)	2007 <sup>‡</sup>	2007 <sup>‡</sup>	2007 <sup>‡</sup>	2007 <sup>‡, ‡‡</sup>
October Household Survey (OHS)	1993–1999 <sup>‡</sup>	1993–1999 <sup>‡</sup>	–	1993–1998 <sup>‡, ‡‡</sup>
General Household Survey (GHS)	2002 <sup>‡</sup>	2002–2011 <sup>‡</sup>	–	–
Demographic & Health Survey (DHS)	1998; <sup>‡</sup> 2003 <sup>‡</sup>	1998 & 2003 <sup>‡</sup>	1998 & 2003 <sup>‡</sup>	1998 & 2003 <sup>‡, ‡‡</sup>
National Income Dynamics Study (NIDS)	2008 <sup>‡</sup>	2008 <sup>‡</sup>	–	2008 <sup>‡, ‡‡</sup>

Source: Table created by authors from vital registration and survey information as follows: VR: Stats SA, 2012 (54), Stats SA, various years (55); Census 1996: Stats SA, 2012 (88); Census 2001: Stats SA, 2012 (54); OHS: National Research Foundation (63); GHS and CS: Stats SA web-based Nesstar information and Stats SA electronic reports (54, 69, 89); DHS: Department of Health *et al.*, 2002 (47), Department of Health *et al.*, 2007 (68); NIDS: Moultrie & Dorrington, 2009 (90).

Notes: \*Direct estimation from routine surveillance; <sup>†</sup>children ever born/children surviving (CEB/CS); <sup>‡</sup>See Table 2; <sup>||</sup>parental survival; <sup>¶</sup>deaths in the household; <sup>\*\*</sup>deaths in the household plus pregnancy/delivery-related question; <sup>††</sup>cause obtained from medical certificate of cause of death on death notification form (BI-1663), or headman reporting on death report (BI-1680); <sup>‡‡</sup>censuses and surveys are not traditional ways to collect cause-of-death data. For the censuses, surveys and RMS, causes were broadly indicated as natural/unnatural, pregnancy/delivery-related, or accident/violence-related causes; <sup>§§</sup>cause ascertained via information from a verbal autopsy instrument.

**Table 2.** National surveys measuring mortality, by year of survey, number of households and persons enumerated, and different methods of mortality measurement

Year	Number of households	Number of persons	Deaths in the household	Parental survival	Sibling survival	Spousal survival	Full birth histories	Summary data on births, deaths of previous births, and surviving children
<b>October Household Survey (OHS)</b>								
1993	30,233	136,468	✓ 12 months					✓ W egb 15–49 <sup>†</sup>
1994	30,279	132,469	✓ 12 months					✓ W egb <55 <sup>†</sup>
1995	29,700	130,787	✓ 22 months	✓				✓ W egb <55 <sup>†</sup>
1996	15,917	72,889	✓ 22 months	✓				✓ W egb <55 <sup>†</sup>
1997	29,810	140,015	✓ 22 months	✓	✓ (Sisterhd)*	✓		✓ W egb <sup>†</sup>
1998	18,981	82,262	✓ 22 months	✓	✓ (Sisterhd)*	✓		✓ W egb <sup>†</sup>
1999	26,164	106,650	✓ 12 months					✓ (W gb12mo) <sup>‡</sup>
<b>General Household Survey (GHS)</b>								
2002	26,243	102,461		✓				✓ (W 12–49)
2003 to 2011	Varied: 24,333 to 29,236	Varied: 94,263 to 109,824		✓				
<b>Community Survey (CS)</b>								
2007	246,618	949,105	✓ 12 months	✓				✓ (W 12–49)
<b>Demographic and Health Survey (DHS)</b>								
1998	12,540	17,500	✓ 12 months	✓	✓		✓	✓ (W 15–49)
2003	7,756	18,274		✓	✓		✓	✓ (W 15–49)
<b>National Income Dynamics Study (NIDS)</b>								
2008	7,305	28,255	✓ 24 months	✓			✓	✓ (W 15–49)

Source: Table created by authors from the following information on surveys: OHS: National Research Foundation (63); GHS and CS: Stats SA web-based Nesstar information and Stats SA electronic reports (54, 69, 89); DHS: Department of Health *et al.*, 2002 (47), Department of Health *et al.*, 2007 (68); NIDS: Moultrie & Dorrington, 2009 (90), Leibbrandt *et al.*, 2009 (71).  
 Notes: \*sisterhood method; <sup>†</sup>W egb: women who have ever given birth; <sup>‡</sup>W gb12mo: all women who have given birth in the last 12 months.



The survey aims to measure multiple facets of the living conditions of the country's households, and the quality of service delivery in selected service sectors. The GHS is a cross-sectional survey, based on a representative sample drawn from the total population. The target population is private households and residents in workers–hostels. Using probability-proportional-to-size principles, a multi-stage, stratified random sample is drawn. Households are visited by fieldwork teams and an extensive questionnaire is filled by enumerators during face-to-face interviews (66, 67). Further information is available in Tables 1 and 2, and at the Stats SA (64) and DataFirst websites (65).

#### South Africa Demographic and Health Survey: 1998 and 2003

Post-democracy, two national Demographic and Health Surveys (DHSs) were conducted collaboratively by the Department of Health, SA MRC, and OrcMacro. The 1998 SADHS employed a two-stage sample based on 1996 census demarcations and stratified according to the nine provinces, each subsequently stratified by urban/non-urban residence (47). The 2003 survey sample, based on the enumeration areas created during the 2001 census, was designed as a nationally representative sample of households. Stratification took place according to the provinces and subsequently by urban/non-urban residence (68). Eligible women were prompted for full birth histories in both surveys. Tables 1 and 2 highlight more information about the types of mortality data collected. More information about the 1998 and 2003 surveys is available in the final full reports (47, 68).

#### Community Survey: 2007

The Community Survey (CS), conducted by Stats SA, was a large-scale nationally representative inter-census household survey conducted in 2007, designed to provide information on the trends of selected demographic, social, and socio-economic profiles of the population. The sampling procedure included a two-stage stratified random sampling process, the first involving the selection of enumerator areas within each municipality, and the second the selection of dwelling units within enumerator areas. Enumerators visited the selected sampled dwelling units and completed questionnaires during face-to-face interviews with study participants (69). The realised sample was adjusted to replicate the national population in a way that the data are consistent internally and with other censuses and surveys (70). Tables 1 and 2 and the Stats SA (64) and DataFirst (65) websites hold more information.

#### National Income Dynamics Study, Wave 1: 2008

The National Income Dynamics Study (NIDS) was South Africa's first national panel study to document

the dynamic structure of households and changes in the incomes, expenditures, assets, access to services, education, health, and well-being of household members. The target population was private households in all provinces and residents in workers–hostels, convents, and monasteries. Households were sampled with a stratified, two-stage cluster sample design, randomly selecting 400 primary sampling units in the first stage from Stats SA's 2003 master sample. In each primary sampling unit, two dwelling-unit clusters were selected. The first fieldwork wave commenced in February 2008. Information was collected on all household members, both resident and non-resident. A household questionnaire and an individual questionnaire for each adult and child in the household were administered via face-to-face interviews in each household (71). See Tables 1 and 2 and the NIDS methodology report (71) for more information.

#### Mortality data collected in national surveys and data availability

Mortality was assessed for different age or interest groups across the surveys, and mortality items sometimes varied within surveys across time (Tables 1–2). Items to assess mortality via both direct and indirect measures were included in these surveys. Assessments of deaths in the household, parental survival, and summary birth histories were included most frequently across time and surveys. When a death was reported, selected further information about the death and deceased were collected. The years of enumeration, interest group, and mortality data items used are shown in Tables 1 and 2.

Unit record data from the OHS can be requested for use in the DataFirst Research Data Centre at the University of Cape Town. The Research Data Centre makes data, statistical analysis software, and trained staff available, free of charge, for this purpose (65). OHS data and metadata are available on compact disc for a fee from Stats SA (72). Unit record data from the GHS and CS are available on disc from Stats SA via request. Stats SA's web-based data repositories (54) contain information about mortality as collected in the GHS and CS. From the GHS, parental survival data for 2002–2010 can be found via SuperWEB, and for 2002–2008 via Nesstar (54). From the CS, parental survival data are available via Nesstar and SuperWEB, and summary birth histories and death-in-the-household data via Nesstar (54). Users can tabulate CS death-in-the-household data by selected demographic variables (e.g. age and sex) and sub-national entities (e.g. province and district council). Stats SA also responds to special requests for tabulations from the Stats SA surveys without requiring a data user's agreement. Due to the poor quality of the data, the 2003 DHS failed to provide reliable estimates of adult and maternal mortality from the parental and sibling survival questions (68). The birth history data, however, were used to

estimate levels and trends in infant and child mortality, and apart from a chapter presenting these results (68), unit record data from the 1998 and 2003 DHS surveys are available per request and by signing a data user's agreement, from the Department of Health and SA MRC. For NIDS, data and supportive documentation for Wave 1 are available via DataFirst servers upon completion of an online form and agreement to the terms of data use (71). Publicly available datasets from any of these surveys contain only non-confidential data.

Extensive resources and information related to these surveys, including questionnaires, reports, metadata, code lists, public data downloads, and microdata-request forms, are available from the University of Cape Town's DataFirst webpage (65).

### *Small-area Demographic and Health Surveillance*

Three INDEPTH HDSSs collect longitudinal health and demographic data in three rural surveillance areas. The Agincourt HDSS in the Bushbuckridge district of Mpumalanga province has collected data since 1992 (73); Dikgale HDSS in the Mankweng district of Limpopo, since 1996 (74); and the Africa Centre Demographic Information System (ACDIS) in the Umkhanyakude district in KwaZulu-Natal, since 2000 (75). Agincourt had a population of approximately 90,000 people in 2011 (73); Dikgale, approximately 8,000 in 2008 (74); and ACDIS, approximately 85,000 in 2008 (75).

### *Mortality data collected in and available from HDSSs*

Mortality information and a range of other socio-demographic and health information are collected through annual censuses and updates of vital events in each household in the site. Verbal autopsies, a well-established method in the absence of routine death registration, are used for classifying causes of death from population-based inquiries (76–78), and are conducted by specially trained fieldworkers who interview a close relative or caregiver of the deceased. Efforts to refine the approach, have led to international standards for verbal autopsy and strengthening standardised interpretation of verbal autopsy data through the InterVA tool, recently culminating in the launch of the InterVA-4 model (14, 79, 80). At Agincourt and ACDIS, the probable cause of death has been attributed via subsequent physician assessment of the verbal autopsy information (73, 81, 82). However, more recently all three sites have been utilising the automated InterVA tool for probabilistic verbal autopsy interpretation and probable cause attribution (83–85) (e-mail communication from Dr. Chifundo Kanjala and Prof. Marianne Alberts, 4–5 April 2012).

Mortality and population data from ACDIS and Agincourt are available through data products, data downloads, and accompanying supportive documenta-

tion at the HDSS' websites (86, 87). Public access to ACDIS and Agincourt data is available via links to downloadable datasets comprising an approximate 1%- and 10%-sample, respectively, of the full datasets (86, 87). These sample datasets can be used for teaching, familiarising potential users with the structure and availability of data, or developing selected analyses before requesting the full dataset. Unit record data that are not publically available can be requested from senior site staff at ACDIS and Agincourt, accompanied by a motivation, analysis plan, and data user's agreement (86, 87). The INDEPTH Network is committed to the principles and practice of data sharing, and has launched the iSHARE portal aiming to make data from the HDSSs publicly available (13).

### *Selected facility-based reporting systems*

Apart from South Africa's routine notification system currently reporting incidence and deaths from 33 notifiable medical conditions to local, provincial, and/or national health departments, the country's particularly sizeable burden of disease from cancer, injury, and tuberculosis is reflected in facility-based surveillance systems to help assess the extent and impact of these conditions. Table 3 provides key information about these surveillance systems. Recognising the current international pressure to provide reliable information about maternal and child health to monitor MDGs 4 and 5, and acknowledging the shortcomings in reporting such mortality via civil registration, Table 3 refers additionally to three facility-based structured mortality audits for peri-natal, child and maternal deaths. The value of other national reporting systems, including the South African Birth Defects Surveillance System (SABDSS), South African Dialysis and Transplantation Registry (SADTR), Surveillance of Work-Related and Respiratory Diseases in South Africa (SORDSA), as well as injury-reporting of the Mine Health and Safety Inspectorate, National Transport Information System, and the South African Police Services, are acknowledged but not described here.

### *Secondary data source: ASSA AIDS and Demographic Model*

Despite having improved vital registration data and nationally representative HIV prevalence data, these data sources do not provide decision makers with a direct measure of the mortality impact of the country's extensive HIV/AIDS epidemic (56, 100). Mathematical models have hence become necessary, and local actuarial researchers have developed the ASSA AIDS and Demographic Model (101) to estimate such impact. The model has been calibrated to empirical data sources, including vital registration, census, and survey data adjusted for biases (100, 102). As time passed and more relevant empirical data became available, updated revisions of the model were released. A number of mortality and

**Table 3.** Selected facility-based data sources that may complement vital registration mortality data

Condition- or age-specific data sources		
Programme	Enumeration years	Selected key information about source
Confidential Enquiry into Maternal Deaths (CEMD)	1998–current	<ul style="list-style-type: none"> <li>● Facility-based, structured reporting</li> <li>● Compulsory reporting after maternal death has been made a notifiable death</li> <li>● Systematic investigation of event, cause and modifiable factors</li> <li>● Despite partial coverage, data useful for highlighting main problems and opportunities in addressing maternal mortality</li> <li>● Resulted in the publication of two sets of national guidelines*</li> </ul>
Peri-natal Problem Identification Programme (PIIP)	2000–current	<ul style="list-style-type: none"> <li>● Facility-based, structured clinical mortality audit of peri-natal deaths</li> <li>● Voluntary participation; compulsory in some provinces</li> <li>● Data set relates to nearly 3,000,000 births and 108,469 deaths</li> <li>● During 2008–09, 275 facilities participated, representing about 963,000 births, i.e. approximately 52% of all facility births for 2008–09</li> <li>● Data not nationally representative, but standardised data collection ensures comparable over time and participating facilities; generate recommendations for better peri-natal care, improved clinical practice; and prioritisation of clinical and public health research.</li> </ul>
Child Healthcare Problem Identification Programme (Child PIP)	2005–current	<ul style="list-style-type: none"> <li>● Facility-based, structured clinical mortality audit of paediatric deaths</li> <li>● Voluntary participation</li> <li>● 2005–2009 data related to 19,295 deaths of 343,408 admissions in 101 hospitals in all nine provinces, representing just under 30% of all hospitals</li> <li>● Not nationally representative data, but standardised collection ensures comparable data;</li> <li>● Recommendations address key health functions, i.e. policy, management and administration, clinical practice, and education.</li> </ul>
National Cancer Registry (NCR)	1986–current	<ul style="list-style-type: none"> <li>● Passive pathology-based surveillance system, with pathology reports confirming a histological cancer diagnosis submitted by selected pathology laboratories</li> <li>● Voluntary participation</li> <li>● Data obtained from 79 laboratories in 2001</li> <li>● Average of 70,000 cancer cases annually, incl. at least 50,000 new cases</li> </ul>
National Injury Mortality Surveillance System (NIMSS)	1999–current	<ul style="list-style-type: none"> <li>● Active collation and centralisation of routinely-kept data of all non-natural deaths entering the forensic medico-legal system at participating mortuaries</li> <li>● Voluntary participation</li> <li>● Data collection and compilation designed in accordance to particular shortcomings in the national registration system regarding non-natural deaths</li> <li>● Systematic information is collected about the incidence and causes of non-natural deaths and demographic characteristics of the deceased</li> <li>● 2001–2008: full coverage in a number of large cities</li> </ul>

Table 3 (Continued)

Condition- or age-specific data sources		
Programme	Enumeration years	Selected key information about source
National Tuberculosis Registry (NTR) and Electronic Tuberculosis Register (ETR.Net)	1995–current	<ul style="list-style-type: none"> <li>● Facility-based reporting of case finding, sputum examination, treatment, and outcomes through standardised forms</li> <li>● Compulsory reporting – TB is notifiable in terms of the National Health Act</li> <li>● Tuberculosis Register (GW 20/11) introduced in 1995</li> <li>● Suspect Register, Laboratory Register, Patient Treatment Card, Clinic/Hospital Card contribute to register</li> <li>● Priority reporting by all facilities to DoH within 24 hours for multidrug-resistant (MDR) and extensively drug-resistant (XDR) TB</li> <li>● Electronic TB Register (ETR.Net) implemented in 2004/5 – facilitates standardised recording, reporting and evaluation of programmes</li> <li>● Information flow: a) Data on case finding, smear conversion and treatment outcomes captured at facility level from patient and facility records; b) collated at sub/district level in electronic TB register; c) data exported into district health information system and transmitted to provincial and national level.</li> </ul>

Source: Table created by authors from the following: CEMD: National Committee CEMD, 2008 (91); PPIP: Pattinson (ed), 2011 (92), Bradshaw *et al.*, 2008 (93); Child PIP: Stephen *et al.*, 2008 (94), Bradshaw *et al.*, 2008 (93); NCR: Albrecht, 2006 (95), Mqoqi *et al.*, 2003 (96); NIMSS: Matzopoulos, 2002 (97); TB: Dept of Health, 2004 (98), Dept of Health, 2007 (99).

Note: \*National Maternity Guidelines for District Hospitals and Clinics, and Essential Steps in the Management of Common Causes of Maternal Deaths in South Africa.

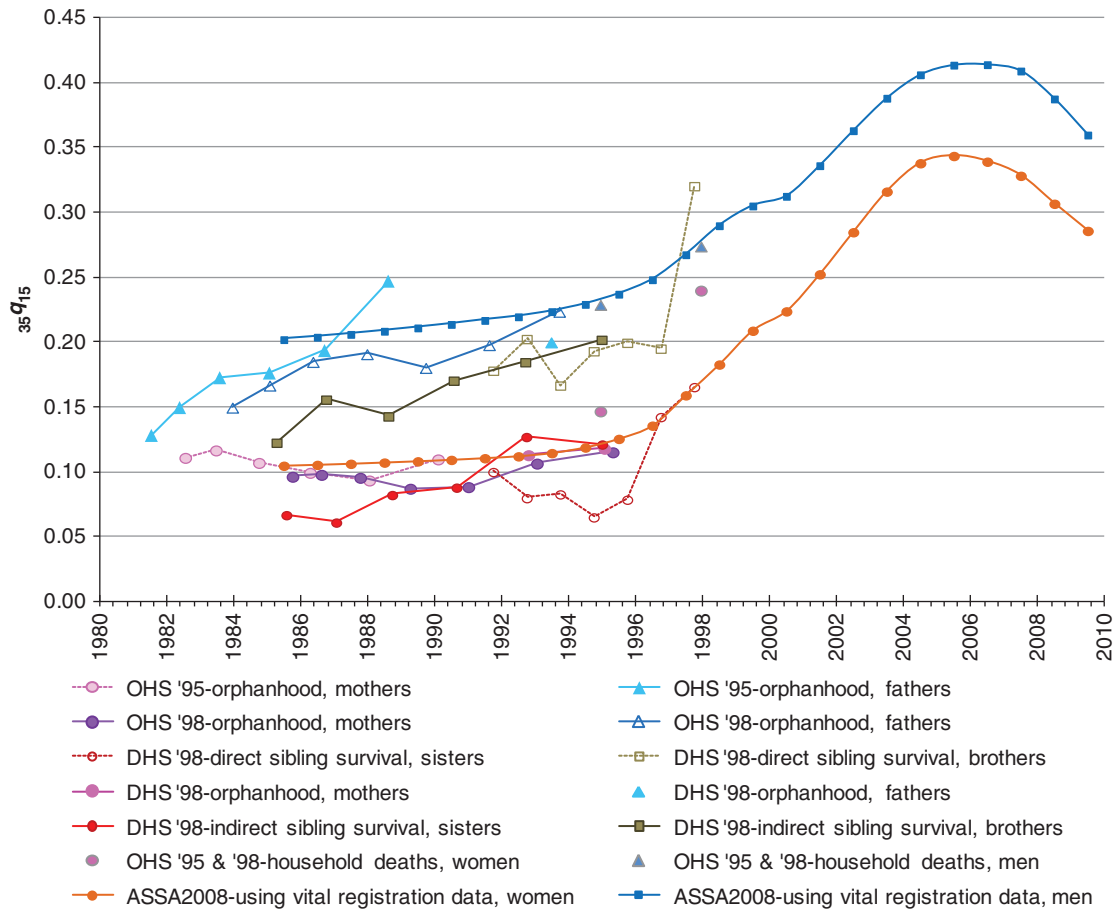
population indicators are available from the models and are widely used as a basis for health policy and planning by both government and the public health research community in South Africa (103–110). While these models are of much practical use, they should be considered ‘interim’ measures until complete vital registration and improved medical certification of causes of death are achieved. Upon online registration, mortality indicators such as  ${}_5q_0$  (under-5 mortality) and  ${}_{45}q_{15}$  (adult mortality) are freely available at the website of the Actuarial Society of South Africa (101).

#### Adult mortality measures from selected sources

As this article aims to review mortality data sources and not results from these sources, Fig. 1 is merely provided to indicate the variety of data sources available for estimating adult mortality, specifically the probability of dying between ages 15 and 50 ( ${}_{35}q_{15}$ ). Estimates of  ${}_{35}q_{15}$  were derived from using both direct and indirect methods, as indicated in Fig. 1. A fairly consistent trend of increasing mortality for most of the 1990s and early 2000s is produced by the different data sources and methods, with a levelling and decline in mortality in more recent years.

#### Discussion

This review demonstrates a rich and varied list of mortality data sources in South Africa. However, it is important to be aware of the strengths and limitations of the different sources and the quality of their data to ensure suitable and strategic utilisation thereof. Different mortality indicators are required for different purposes, and at varied levels of population aggregation. For instance, reliable measures of peri-natal and under-5 mortality at the health-district level are of critical importance in planning and providing for maternal and child health services. Therefore, it is necessary to have robust measures of these indicators at this level, or even at sub-district level. Mortality rates for specific conditions such as tuberculosis, HIV/AIDS, injuries, cardiovascular conditions, neoplasms, and respiratory disease – the high-burden conditions in South Africa – should ideally be measured at magisterial level and at least at health-district level to inform planning for prevention, detection and treatment optimally. Mortality from maternal conditions and other less-prevalent non-communicable diseases, in contrast, can at best be effectively monitored for differences and change at the provincial level, given their relatively infrequent occurrence.



*Fig. 1.* The probability of dying between ages 15 and 50 ( $_{35}q_{15}$ ) from different data sources. Source: OHS and DHS estimates extracted from Figure 6.2 in Dorrington RE, Moultrie TA, Timæus IM, 2004 (44). ASSA2008 estimates calculated from ASSA life table values which are based on vital registration data (101).

**Under-5 mortality**

The under-5 mortality rate (U5MR) is a key indicator of child health and overall development. While its measurement at national level is important for monitoring countries’ progress towards the targets of MDG 4, timely and accurate measurement at sub-national levels are critical for evaluating and prioritising child health care needs and services. Although vital registration is the optimal data source for this, the under-reporting of stillbirths, live births and childhood deaths in South Africa results in under-estimates of child mortality (43, 46, 57). Furthermore, in the context of rapid epidemiological change, the current 2-year reporting delay reduces the utility of the rates. Data from complete birth histories collected in DHSs, are generally a key source for measuring U5MR trends (111, 112), but do not permit estimates lower than provincial/state level. In addition, data quality problems in both the 2001 Census and 2003 DHS have rendered it impossible to derive reliable estimates of under-5 mortality from these sources (113). Census-based summary birth histories may yield estimates at the health-district level, but apart from recall

and omission biases, such estimates are limited by their availability only once in 5 or 10 years, and for a reference period of several years preceding the census. Finally, the existing audit programmes for child (Child PIP) and peri-natal (PPIP) events are rich in their content, but biased in that only facility-based events are recorded, and participation in most provinces continues to be voluntarily.

For as long as birth and death registration are incomplete, a strategy is required that would distinguish and integrate useful, quality data from different well-administered sources towards deriving robust data on the levels and determinants of U5MRs at district level. Research in Indonesia, for example, has demonstrated the low-cost, time-efficient potential to adapt the DHS model into a ‘mini-DHS’ to collect data and provide robust under-5 mortality measures at the district level, allowing researchers to point out significant differentials at this level, thus assisting health-district officials to plan for a locally-appropriate response towards achieving national targets for MDG4 (114). Alternatively, South Africa needs to further strengthen the efforts towards complete

birth and death registration, a process that has significantly progressed in a short period (57, 115). Particular efforts for children aged 1–4 years are needed (46). The APAI-CRVS (18) initiative and the recommendations of the Health Data Advisory and Co-ordination Committee (109) show great potential for further stimulating vital registration towards completeness.

### Adult mortality levels and causes

Overshadowed by a focus on child health for many decades, it has taken a severe epidemic to modify the neglect of adult mortality in sub-Saharan Africa. This neglect was partly due to the lack of reliable empirical data to measure adult mortality in the region. For most of the past century, South Africa was no exception (41, 116). During the 1980s and 1990s, however, meticulous research efforts started putting together pieces from the disjointed vital registration puzzle (42, 117–119). As alternative data sources became available in tandem with improved civil registration and vital statistics practices, researchers were in a position to triangulate and interrogate different sources and started having a better handle on estimating adult mortality levels (18, 34, 43, 44, 56, 120–124). Differences in adult mortality estimates are shown in these publications, indicating data limitations such as event omission and recall bias in data from censuses and surveys, age misreporting, violation of selected assumptions in indirect methods, and uncertainty about the level of completeness of death registration.

More challenging has been deriving cause-specific mortality estimates. The vital registration system is likely the optimal source to calculate cause-specific estimates from, but a number of problems limit its utilisation, including an incomplete national cause profile due to incomplete death registration, and an urban bias in registration. For reported deaths, limitations of cause data include incomplete medical certification of the cause(s) of death, relatively high proportions of deaths in the ill-defined natural and undetermined unnatural categories, and misclassification of causes of death (104, 110, 117, 119, 125, 126). Cause limitations are exacerbated by the continued practice that traditional headmen, on the basis of relatives' information about the deaths, are allowed to certify deaths from natural causes. This may affect up to 10% of primarily rural registered deaths (127). While it is a welcome practice in terms of improving completeness of death reporting, it is not ideal for cause-of-death data.

### Tuberculosis

Alternative sources of causes of adult deaths could be useful for mortality estimates from specific causes. Tuberculosis reporting is compulsory in terms of the National Health Act, and the Register operated by the

National Tuberculosis Control Program could play a complementary role, both as a tool to keep track of deaths at the health-district level, and, where data could be matched via linking variables, as a means of assessing agreement of cause attribution between the Tuberculosis Register and vital registration. In addition, this may be useful in assessing the completeness of tuberculosis reporting on the death notification form. The close association between HIV and tuberculosis calls for appropriate strategies to cross-reference data from these two sources for verifying event occurrence and capturing suitable additional data to guide programme improvement to prevent or curb mortality. Similarly, towards optimally informing MDG5, teasing out differing estimates of maternal mortality (109, 128) may gain from linking death records from the Confidential Enquiry into Maternal Deaths and vital registration databases, and triangulation with census data.

### Cancer

Cancer has a considerable impact on the country's disease burden as the 4th leading category of cause of death in 2000 (104). One in 12 cancer causes, accounting for almost 2,400 cases, were ill-defined and could not be attributed to a site-specific cancer (129), thereby diminishing the utility of the information. It may hence be useful to link records from the Cancer Registry and vital registration databases in a capture–recapture design towards reducing ill-defined cancer diagnoses in the vital registration database. Compared to vital registration data, such registers or audits generally have a considerable advantage in terms of disease control in that they have the potential to measure cause-specific incidence, prevalence, treatment, and case fatality at the health-district or at least provincial level, thus being able to point out differentials at this level, and assisting health-district and provincial officials to identify potential patient load and priorities for locally-appropriate health services.

### Injuries

For injuries, there is a system problem because the South African death notification form does not include a field for the apparent manner of death (homicide, suicide or accident). All deaths from injuries are certified as 'unnatural' deaths, and must undergo a medico-legal investigation at a state mortuary. However, some forensic pathologists consider that, in terms of the Inquest Act, they cannot indicate the circumstance of the death on the death certificate. Thus, the external cause (e.g. burn, firearm discharge, or fall) of many injury deaths is undetermined. While the National Transport Information System records information for selected motor vehicle collision deaths, the Mine Health and Safety Inspectorate records fatal mining injuries, and the South African Police Services violence-related injury, mortality

from other external causes is not monitored by any agency (130). The National Injury Mortality Surveillance System (NIMSS) data are filling this gap by providing more comprehensive external cause information which would be valuable in the design and evaluation of injury control programs, but are limited by the lack of full-country coverage and an urban bias (104, 107). For as long as the civil registration data do not include the external cause of injury deaths and NIMSS data do not include all injury deaths, the response to this large cause of premature death and disability can neither be comprehensive nor adequate. It may therefore be worthwhile to adapt the death notification form to include the external cause of injuries, apparent manner of death, scene of injury (e.g. private house, or street/highway), and district of injury (which may differ to the district of death). The value of these data items for injury prevention and safety and peace promotion speak for themselves.

#### HIV/AIDS

Finally, HIV/AIDS was estimated the single largest cause of both death and years of life lost (YLLs) in 2000, respectively, accounting for 30% of total deaths, and 39% of total YLLs (104). Despite its enormous impact on mortality and premature death, HIV is not notifiable in South Africa, and no register or audit are assigned to capture details of suspected or confirmed HIV cases. A number of studies have found HIV/AIDS under-certified in both adult and paediatric deaths (110, 126, 131–134) and although these reports are valuable in alerting data users to problems with the accuracy of cause attribution, they should also be seen as valuable in alerting certifying officers, coders, and researchers to indicator conditions, alternate terminology, and euphemisms that are used to indicate HIV as a possible cause.

A national initiative to improve the quality of medical certification should emphasise the importance of appropriate recording of HIV on death notifications, particularly in the new political climate of acceptance of the role of HIV in causing AIDS (135, 136), towards accurately informing local responses and reliably reporting progress on Target 6A of MDG6 (i.e. have halted by 2015 and begun to reverse the spread of HIV/AIDS). This initiative should be monitored by a medical record review in a representative sample of death notifications to ascertain the veracity of certification and coding practices. Additionally, the HDSSs have built considerable relationships of trust in their communities, and matching HDSS and vital registration records may generate valuable knowledge of the extent of HIV/AIDS misclassification in registered rural deaths. While substantial problems of accuracy have been identified with physician-assigned causes of death in national vital registration data (110, 126), and even in deaths that occurred in

tertiary health facilities (39, 132, 133), local HDSS studies using verbal autopsy data have shown successful detection and a substantial presence of HIV-related mortality with closely comparable findings between physician- and InterVA-assessments. (84, 85)

#### *Challenges and opportunities for mortality measurement*

While our review suggests that there are a number of potentially useful data sources on mortality, some of which could be used complementary, it is also clear that data use and analysis based on these collections have been restricted by limitations. At times, mortality data collection has been poor and mortality levels could not be derived due to the extent of missing or illogical data in selected surveys and censuses. For some earlier data sources, the quality of the data was unassessed and the data unused. This may have resulted from a lack of knowledge on how to assess data quality issues, limited capacity to apply selected methods of mortality estimation, prolonged time periods before data become available for public use, financial costs to obtain data, or bureaucratic processes that hinder data access and use. Additionally, sample sizes varied across years for surveys – at times substantially, and not necessarily congruent with national population numbers; the age ranges of respondents for the same data items at times differ across surveys, or across years within a survey; and changing administrative borders and place names have sometimes affected mortality reporting and measurement.

Recognising these challenges presents an important step towards improving mortality measurement, from the planning of enquiry/reporting systems through data collection, processing, and compilation, to depositing data in the public domain for independent evaluation and analysis. Stats SA has greatly improved availability of national mortality data over the past 20 years, has reduced public use waiting time, and has collaborated with strategic partners to improve completeness of vital registration. Mortality measurement will further gain from creating opportunities for wider public knowledge about the importance and public health uses of reliable and valid mortality data; further improvements in completeness of registration and timeliness of data availability; adequate, on-going training of certifiers and coders in cause attribution; and strategic strengthening of analytical capacity at Stats SA and research and academic institutions.

#### Conclusion

In a continent often reported as lacking the basic data to infer levels and trends of all-cause and cause-specific mortality, this article has identified a number of data sources in South Africa that, after critical review and adjustment, could yield valuable policy insights into

mortality change over the past two decades. Data sources with mortality items are many and varied, offering a promising scenario for improved population health planning from an evidence base informed by multiple sources. However, it is clear that more can and must be gained for mortality measurement by tackling three key issues: data quality, data triangulation, and analytic capacity.

Data quality in surveys and censuses can be improved by demanding nothing less than excellent fieldworker training and excellent quality control measures in the field. For improved quality in vital statistics, further focussed advances in completeness of death registration, and, in particular, a strong, co-ordinated national response towards improved coverage and accuracy of medical certification of causes of death is recommended. The latter is simply critical. Moreover, with studies pointing to problems in physician-certified causes, (137, 138) such causes should not be taken as automatically having content validity, and the possibility of routinely comparing a sample of death certificates with hospital records/doctors' notes/clinic or day hospital cards, should be pursued.

A focussed agenda is recommended towards data triangulation and contestability via linkage and validation studies that will allow drawing on complementary properties of different sources and, in particular, will assist in completeness estimation and improve our understanding of the accuracy in cause-of-death attribution. Such improved understanding holds clear gains for improved mortality estimates, enhanced resource and service distribution, and, eventually, better meeting the health needs of the population.

However, data quality assessment and triangulation, like other aspects of mortality measurement, require sufficient competent analytic capacity. As analytic capacity has not been expanded upon compared to mortality experiences in the population, nor the increase in national mortality data collection and availability, the expansion and strengthening of analytic capacity is a critical, overarching recommendation.

Achieving these will not be easy, and a co-ordinated research agenda for mortality data collection, evaluation, comparison, analysis, and use, along with an operational agenda for quality assurance and analytical capacity strengthening, are recommended. These should be generated and backed-up by adequate and independent human and fiscal resources. For the future, it will be important to adopt a strategic approach to data collection, streamlined by lessons from past experience, and enhanced by successes and innovative modes of data collection elsewhere.

## Acknowledgements

The research was carried out while the first author was holding a University of Queensland Research Scholarship and Endeavour

International Postgraduate Research Scholarship at the University of Queensland.

## Author contributions

JJ and CR conceptualised the article. JJ undertook the review and wrote the first draft with substantial inputs from CR. All co-authors gave expert inputs and contributed to the critical review of subsequent drafts. JJ produced the tables and graph with expert inputs from CR (tables) and RD (graph).

## Conflict of interest and funding

No funding was received to conduct the review. The authors have no conflict of interest.

## References

1. Graunt J. Natural and political observations made upon the bills of mortality. In: Willcox WF, ed. A reprint of economic tracts, 1939. Baltimore: Johns Hopkins Press; 1662.
2. UNDP (1990). Human Development Report 1990. New York, NY: United Nations Development Programme.
3. United Nations (2000). United Nations Millennium Declaration. Resolution adopted by the General Assembly Fifty-Fifth Session Agenda Item 60 (b). A/RES/55/2. New York, NY: United Nations.
4. United Nations (2011). The Millennium Development Goals Report 2011. New York, NY: United Nations. Available from: [http://www.un.org/millenniumgoals/pdf/\(2011\\_E\)%20MDG%20Report%202011\\_Book%20LR.pdf](http://www.un.org/millenniumgoals/pdf/(2011_E)%20MDG%20Report%202011_Book%20LR.pdf) [cited 7 February 2012].
5. Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S, et al. A scandal of invisibility: making everyone count by counting everyone. *Lancet* 2007; 370: 1569–77.
6. Shryock HS, Siegel JS, Stockwell EG. The methods and materials of demography, Condensed edition. San Diego, CA: Academic Press; 1976.
7. United Nations Secretariat. Approaches to the collection of mortality data in the context of data needs. In: United Nations, ed. Data bases for mortality measurement: Proceedings of the United Nations/World Health Organization Working Group on Data Bases for Measurement of Levels, Trends, and Differentials in Mortality; Bangkok; 1981; ST/ESA/Series/A/84. New York: United Nations; 1984. pp. 33–44.
8. Rowland DT. Demographic methods and concepts. New York, NY: Oxford University Press; 2006.
9. Hill K, Lopez AD, Shibuya K, Jha P. Interim measures for meeting needs for health sector data: births, deaths, and causes of death. *Lancet* 2007; 370: 1726–35.
10. Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, Rao C, et al. Civil registration systems and vital statistics: successes and missed opportunities. *Lancet* 2007; 370: 1653–63.
11. Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD, et al. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974–2006. *PLoS Med* 2010; 7: e1000260.
12. Child Mortality Coordination Group. Tracking progress towards the Millennium Development Goals: reaching consensus on child mortality levels and trends. *Bull World Health Organ* 2006; 84: 225–32.
13. Sankoh O, Byass P. The INDEPTH Network: filling vital gaps in global epidemiology. *Int J Epidemiol* 2012; 41: 579–88.



14. Byass P, Chandramohan D, Clark SJ, Ambruso L, Fottrell E, Graham WJ, et al. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Glob Health Action* 2012; 5: 1–8.
15. Bradshaw D, Groenewald P, Bourne DE, Mahomed H, Nojilana B, Daniels J, et al. Making COD statistics useful for public health at local level in the city of Cape Town. *Bull World Health Organ* 2006; 84: 211–7.
16. Mathers C, Lopez A, Stein C, Fat DM, Rao C, Inoue M, et al. Deaths and disease burden by cause: global burden of disease estimates for 2001 by World Bank country groups. Disease Control Priorities Project, Working Paper No. 18. Annex A, Table 6. New York, NY; 2004, revised 2005. Available from: <http://www.dcp2.org/file/33/wp18.pdf> [cited 11 Mar 2012].
17. Mathers C, Lopez A, Murray C. The burden of disease and mortality by condition: data, methods, and results for 2001. In: Lopez A, Mathers C, Ezzati M, Jamison D, Murray CJ, eds. *Global burden of disease and risk factors*. New York, NY: Oxford University Press; 2006. p. 100–6, Table 3A.3.
18. United Nations, African Union Commission, African Development Bank. Africa programme on accelerated improvement of civil registration and vital statistics (APAI-CRVS). New York: Economic Commission for Africa, United Nations; 2011.
19. Health Metrics Network [Internet]. Geneva: Health Metrics Network, WHO; 2012. Available from: [http://www.who.int/healthmetrics/move\\_it/en/index.html](http://www.who.int/healthmetrics/move_it/en/index.html) [cited 12 July 2012].
20. Bourne DE. Sources of South African mortality data, 1910–1992. *S Afr Med J* 1995; 85: 1020–2.
21. Bah S. The making and unmaking of a national but stratified vital statistics system in the Republic of South Africa and the new making of a more comprehensive vital statistics system. *South Afr J Demogr* 1999; 7: 45–50.
22. Khalfani AK, Zuberi T, Bah S, Lehohla P. Population statistics. In: Zuberi T, Sibanda A, Udjo EO, eds. *The Demography of South Africa*. Armonk, New York: M.E. Sharpe; 2005.
23. Bah S. Multiple forces working in unison: the case of rapid improvement of vital statistics in South Africa post-1996. *World Health Popul* 2009; 11: 50–9.
24. Republic of South Africa. Births and Deaths Registration Act, 1992 (No. 51 of 1992). In: *Government Gazette No. 13953*. Cape Town: Government Printer; 1992.
25. Freund B. Forced resettlement and the political economy of South Africa. *Rev Afr Polit Econ* 1984; 49–63.
26. Zulu EM, Sibanda A. Racial differences in household structure. In: Zuberi T, Sibanda A, Udjo EO, eds. *The demography of South Africa*. Armonk, New York: M.E. Sharpe; 2005.
27. Dorrington RE, Bradshaw D, Wegner T. Estimates of the level and shape of mortality rates in South Africa around 1985 and 1990 derived by applying indirect demographic techniques to reported deaths. MRC Technical Report. Cape Town: South African Medical Research Council; 1999.
28. Kok P, O'Donovan M, Bouare O, Van Zyl J. Post-apartheid patterns of internal migration in South Africa. Pretoria: Human Sciences Research Council Publishers; 2003.
29. Kaufman C. Reproductive control in South Africa. New York; 1997. Available from: <http://www.popcouncil.org/pdfs/wp/97.pdf> [cited: 8 June 2012].
30. Murray C. Class, gender and the household: the developmental cycle in Southern Africa. *Dev Change* 1987; 18: 235–49.
31. Murray C. Migrant labour and changing family structure in the rural periphery of southern Africa. *J South Afr Stud* 1980; 6: 139–56.
32. Platzky L, Walker C, Surplus People Project (South Africa). *The surplus people: forced removals in South Africa*. Johannesburg: Ravan Press; 1985.
33. Sharp J. Relocation and the problem of survival in Qwaqwa: a report from the field. *Soc Dynam* 1982; 8: 11–29.
34. Moultrie TA, Dorrington RE. Used for ill; used for good: a century of collecting data on race in South Africa. *Ethn Racial Stud* 2011; 35: 1447–65.
35. Republic of South Africa. Constitution of the Republic of South Africa (Act 200 of 1993). Cape Town: Government Printer; 1993 (Assented to: 25 Jan 1994; Commencement: 27 April 1994).
36. Groenewald P, Pieterse D. Getting the facts of death right. *S Afr Med J* 2007; 97: 1056–8.
37. Bradshaw D, Kielkowski D, Sitas F. New birth and death registration forms - A foundation for the future, a challenge for health workers? *S Afr Med J* 1998; 88: 971–4.
38. National Health Information System. Evaluation of the new death notification form (BI 1663). Pretoria, South Africa: Department of Health; 1999.
39. Nojilana B, Groenewald P, Bradshaw B, Reagon G. Quality of cause of death certification at an academic hospital in Cape Town, South Africa. *S Afr Med J* 2009; 99: 648–52.
40. Burger EH, Van der Merwe L, Volmink J. Errors in the completion of the death notification form. *S Afr Med J* 2007; 97: 1077–81.
41. Rao C, Bradshaw D, Mathers CD. Improving death registration and statistics in developing countries: lessons from sub-Saharan Africa. *South Afr J Demogr* 2004; 9: 81–99.
42. Dorrington RE. African mortality rates - an initial estimate. *Trans of the Actuarial Society of South Africa* 1989; 7: 169–202.
43. Dorrington RE, Bourne D, Bradshaw D, Laubscher R, Timeaus IM. The impact of HIV/AIDS on adult mortality in South Africa. Tygerberg, Cape Town; 2001. Available from: <http://www.mrc.ac.za/bod/complete.pdf> [cited 12 September 2010].
44. Dorrington RE, Moultrie TA, Timeaus IM. Estimation of mortality using the South African Census 2001 data. Cape Town; 2004. Available from: [http://www.commerce.uct.ac.za/Research\\_Units/CARE/Monographs/Monographs/Mono11.pdf](http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono11.pdf) [cited 11 August 2010].
45. Bennett NG, Horiuchi S. Mortality estimation from registered deaths in less developed countries. *Demography* 1984; 21: 217–33.
46. Darikwa TB, Dorrington RE. The level and trends of child mortality in South Africa, 1996–2006. *J Afr Population Stud* 2011; 25: 158–72.
47. Department of Health (South Africa), South African Medical Research Council, Macro International Inc. South Africa demographic and health survey 1998. Full Report. Pretoria; 2002. Available from: <http://www.measuredhs.com/publications/publication-FR131-DHS-Final-Reports.cfm> [cited 18 July 2012].
48. Dorrington R, Moultrie T, Timeaus I, Nannan N. Estimates of provincial fertility and mortality in South Africa, 1985–1996. *South Afr J Demogr* 2004; 9: 25–57.
49. Nannan N, Bradshaw D, Timeaus I, Dorrington R. The impact of HIV/AIDS on infant and child mortality in South Africa (poster). 13th International AIDS Conference; Durban, South Africa; 9–14 July 2000.
50. Udjo EO. An examination of recent census and survey data on mortality within the context of HIV/AIDS. In: Zuberi T, Sibanda A, Udjo E, eds. *The demography of South Africa*. Armonk, New York: M.E. Sharpe; 2005.
51. Arriaga EE. Population analysis with microcomputers, volume 1. Presentation of techniques. Washington, USA: United States Census Bureau; 1994.

52. Statistics South Africa. Mortality and causes of death in South Africa, 2007: findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa; 2009.
53. World Health Organization. ICD-10: International statistical classification of diseases and related health problems, Tenth revision. 2nd ed. Geneva: World Health Organization; 2004.
54. Statistics South Africa [Internet]. Pretoria: Statistics South Africa; c2007. Available from: <http://interactive.statssa.gov.za:8282/webview/> and <http://interactive.statssa.gov.za/superweb/login.do> [cited 10 Nov 2012].
55. Statistics South Africa [Internet]. Pretoria: Statistics South Africa; various years. Available from: <http://www.statssa.gov.za/publications/statspastfuture.asp?PPN=P0309.3&SCH=5097> [cited 8 February 2012].
56. Bradshaw D, Laubscher R, Dorrington RE, Bourne DE, Timæus IM. Unabated rise in number of adult deaths in South Africa. *S Afr Med J* 2004; 94: 278–9.
57. Bradshaw D, Dorrington R, Laubscher R. Rapid mortality surveillance report 2011. Cape Town: South African Medical Research Council; 2012.
58. Brody L, Bradshaw D, Bourne D, Laubscher R. Western Cape Province Mortality 1997–2002. Analysis of empirical cause of death data collected by Statistics South Africa from death notifications. Cape Town; 2007. Available from: <http://www.mrc.ac.za/bod/wcmortality2002.pdf> [cited 9 March 2012].
59. Dorrington RE, Bradshaw B, Laubscher R, Bourne D. The impact of HIV/AIDS on adult mortality in South Africa and the impact of antiretroviral therapy on that impact. Population Association of America (PAA) 2008 Annual Meeting, New Orleans, LA. 17–19 April 2008.
60. Statistics South Africa. The people of South Africa: population census 1996. Calculating the undercount in Census '96. Report No. 03-01-18 (1996). Pretoria; 1998. Available from: <http://www.statssa.gov.za/census01/Census96/HTML/default.htm> [cited 18 July 2012].
61. Statistics South Africa. Census 2001: how the count was done. Pretoria; 2003. Available from: <http://www.statssa.gov.za/publications/Report-03-02-02/Report-03-02-02.pdf> [cited 8 June 2012].
62. Statistics South Africa [Internet]. Pretoria: Statistics South Africa; c2005. Available from: <http://www.statssa.gov.za/publications/populationstats.asp> [cited 10 Nov 2012].
63. National Research Foundation. South African data archive: data catalogue. Pretoria: National Research Foundation; 2001.
64. Statistics South Africa [Internet]. Pretoria: Statistics South Africa; c2012. Available from: <http://www.statssa.gov.za/default.asp> [cited 10 Nov 2012].
65. University of Cape Town [Internet]. Cape Town (South Africa): University of Cape Town. Available from: <http://www.datafirst.uct.ac.za/catalogue3/index.php/catalog> [cited 10 Nov 2012].
66. Statistics South Africa. General Household Survey 2009. Statistical release P0318. Pretoria; 2010. Available from: <http://www.statssa.gov.za/publications/P0318/P0318June2009.pdf> [cited 8 September 2010].
67. Statistics South Africa. General Household Survey 2009: metadata. Report No. P0318. Pretoria; 2010. Available from: <http://interactive.statssa.gov.za:8282/metadata/surveys/GHS/2009/General%20Household%20Survey%202009%20metadata.pdf> [cited 8 September 2010].
68. Department of Health (South Africa), South African Medical Research Council, Macro International Inc. South Africa Demographic and Health Survey 2003. Full report. Pretoria; 2007. Available from: <http://www.measuredhs.com/publications/publication-FR206-DHS-Final-Reports.cfm> [cited 18 July 2010].
69. Statistics South Africa. Community Survey, 2007. Basic results: municipalities. statistical release P0301.1. Pretoria; 2008. Available from: <http://www.statssa.gov.za/Publications/P03011/P030112007.pdf> [cited 6 September 2010].
70. Statistics South Africa. Community Survey 2007: methodology, processes and highlights of key results. Report No. 03-01-20 (2007). Pretoria; 2007. Available from: <http://www.statssa.gov.za/publications/Report-03-01-20/Report-03-01-20.pdf> [cited 6 September 2010].
71. Leibbrandt M, Woolard I, De Villiers L. National Income Dynamics Study (NIDS). Methodology: Report on NIDS Wave 1. Technical Paper no. 1. Cape Town; 2009. Available from: <http://www.nids.uct.ac.za/home/index.php?Nids-Documentation/technical-papers.html> [cited 8 March 2011].
72. Statistics South Africa. October Household Survey 1999. Statistical Release P0317. Pretoria; 2000. Available from: <http://www.statssa.gov.za/publications/statsdownload.asp?PPN=P0317&SCH=854> [cited 10 March 2012].
73. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt Health and Socio-Demographic Surveillance System. *Int J Epidemiol* 2012; 41: 988–1001.
74. Cook I, Alberts M, Burger S, Byass P. All-cause mortality trends in Dikgale, rural South Africa, 1996–2003. *Scand J Public Health* 2008; 36: 753–60.
75. Tanser F, Hosegood V, Barnighausen T, Herbst K, Nyirenda M, Muhwava W, et al. Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *Int J Epidemiol* 2008; 37: 956–62.
76. Chandramohan D, Maude GH, Rodrigues LC, Hayes RJ. Verbal autopsies for adult deaths: their development and validation in a multicentre study. *Trop Med Int Health* 1998; 3: 436–46.
77. Fauveau V. Assessing probable causes of death without death registration or certificates: a new science? *Bull World Health Organ* 2006; 84: 246–7.
78. Snow RW, Winstanley MT, Marsh VM, Newton C, Waruiru C, Mwangi I, et al. Childhood deaths in Africa: uses and limitations of verbal autopsies. *Lancet* 1992; 340: 351–5.
79. Baiden F, Bawah A, Biai S, Binka F, Boerma T, Byass P, et al. Setting international standards for verbal autopsy. *Bull World Health Organ* 2007; 85: 570–1.
80. Fottrell E, Byass P. Verbal autopsy: methods in transition. *Epidemiol Rev* 2010; 32: 38–55.
81. Tollman SM, Kahn K, Sartorius B, Collinson MA, Clark SJ, Garenne ML. Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study. *Lancet* 2008; 372: 893–901.
82. Hosegood V, Vanneste A-M, Timæus IM. Levels and causes of adult mortality in rural South Africa: the impact of AIDS. *AIDS* 2004; 18: 663–71.
83. Byass P, Kahn K, Fottrell E, Collinson MA, Tollman SM. Moving from data on deaths to public health policy in Agincourt, South Africa: approaches to analysing and understanding verbal autopsy findings. *PLoS Med* 2010; 7: e1000325.
84. Byass P, Kahn K, Fottrell E, Mee P, Collinson M, Tollman S. Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. *Popul Health Metr* 2011; 9: 46.
85. Herbst A, Mafojane T, Newell M-L. Verbal autopsy-based cause-specific mortality trends in rural KwaZulu-Natal, South Africa, 2000–2009. *Popul Health Metr* 2011; 9: 47.
86. Africa Centre for Health and Population Studies [Internet]. Somkhele, Mtubatuba (South Africa): Africa Centre for

- Health and Population Studies, University of KwaZulu-Natal. Available from: <http://www.africacentre.ac.za/Default.aspx?tabid=69> [cited 10 November 2012].
87. Agincourt MRC/Wits Rural Public Health and Health Transitions Research Unit [Internet]. Agincourt, Bushbuckridge (South Africa): Agincourt MRC/Wits Rural Public Health and Health Transitions Research Unit, University of the Witwatersrand. Available from: <http://www.agincourt.co.za/index.php/data/> [cited 10 Nov 2012].
  88. Statistics South Africa [Internet]. Pretoria: Statistics South Africa; 2012. Available from: <http://www.statssa.gov.za/census01/Census96/HTML/default.htm> [cited 7 February 2012].
  89. Statistics South Africa [Internet]. Pretoria: Statistics South Africa; 2002–2010. Available from: <http://www.statssa.gov.za/publications/statspastfuture.asp?PPN=P0318&SCH=4912> [cited 12 February 2012].
  90. Moultrie T, Dorrington RE, National Income Dynamics Study (NIDS). Demography: Analysis of the NIDS Wave 1 dataset. Discussion Paper no. 9. Cape Town; 2009. Available from: <http://www.nids.uct.ac.za/home/index.php?Nids-Documentation/discussion-papers.html> [cited 8 March 2012].
  91. National Committee for Confidential Enquiry into Maternal Deaths. Saving mothers 2005–2007: fourth report on confidential enquiries into maternal deaths in South Africa. Pretoria; 2008. Available from: [http://www.doh.gov.za/docs/reports/2011/saving\\_b.pdf](http://www.doh.gov.za/docs/reports/2011/saving_b.pdf) [cited 10 February 2012].
  92. Pattinson RC, ed. Saving babies 2008–2009: seventh report on perinatal care in South Africa. Pretoria; 2011. Available from: <http://www.ppip.co.za/downloads/Saving%20Babies%202008-9.pdf> [cited 12 February 2012].
  93. Bradshaw D, Chopra M, Kerber K, Lawn J, Moodley J, Pattinson R, et al. Every death counts: saving the lives of mothers, babies and children in South Africa. Cape Town: South African Medical Research Council; 2008.
  94. Stephen C, Bamford L, Patrick M, Wittenberg D, eds. Saving children 2009: five years of data. A sixth survey of child healthcare in South Africa. Pretoria: Tshepesa Press, MRC, CDC; 2011.
  95. Albrecht C. Overview of the South African cancer research environment as a basis for discussions concerning the activation of CARISA (Cancer Research Initiative of South Africa). Report prepared by Dr Carl Albrecht, independent medical research consultant; 2006. Available from: <http://www.sahealthinfo.org/cancer/overviewdocument.pdf> [cited 12 February 2012].
  96. Mqoqi N, Sitas F, Halkett J. Disease registries. In: Ijumba P, ed. South African health review 2002. Durban: Health Systems Trust; 2003 pp. 389–410.
  97. Matzopoulos R. A profile of fatal injuries in South Africa. Third annual report of the National Injury Mortality Surveillance System (NIMSS). Cape Town; 2002. Available from: <http://www.sahealthinfo.org/violence/> [cited 31 August 2004].
  98. Department of Health (South Africa). The South African National Tuberculosis Control Programme: practical Guidelines. Pretoria; 2004. Available from: <http://www.kznhealth.gov.za/chrp/documents/Guidelines/Guidelines%20National/Tuberculosis/SA%20TB%20Guidelines%202004.pdf> [cited 15 February 2012].
  99. Department of Health (South Africa). Tuberculosis strategic plan for South Africa, 2007–2011. Pretoria; 2007. Available from: <http://www.info.gov.za/view/DownloadFileAction?id=72544> [cited 15 February 2012].
  100. Johnson LF, Dorrington RE. Modelling the demographic impact of HIV/AIDS in South Africa and the likely impact of interventions. *Demogr Res* 2006; 14: 541–74.
  101. Actuarial Society of South Africa: AIDS Committee [Internet]. Cape Town, South Africa: Actuarial Society of South Africa; 2011. Available from: <http://aids.actuarialsociety.org.za/ASSA2008-Model-3480.htm> [cited 10 November 2012].
  102. Dorrington RE. The ASSA2000 suite of models. Somerset West, South Africa: Actuarial Society of South Africa Convention; 2000.
  103. Department of Health (South Africa). Operational plan for comprehensive HIV and AIDS care, management and treatment for South Africa. Pretoria; 2003. Available from: <http://www.info.gov.za/otherdocs/2003/aidsplan/report.pdf> [cited 21 February 2012].
  104. Bradshaw D, Groenewald P, Laubscher R, Nannan N, Nojilana B, Norman R, et al. Initial burden of disease estimates for South Africa, 2000. *S Afr Med J* 2003; 93: 682–8.
  105. Nattrass N. The moral economy of AIDS in South Africa. Cambridge: Cambridge University Press; 2004.
  106. Dorrington RE, Johnson L, Bradshaw D, Daniel T. The demographic impact of HIV/AIDS in South Africa. National and provincial indicators for 2006; 2006. Available from: <http://www.mrc.ac.za/bod/DemographicImpactHIVIndicators.pdf> [cited 19 November 2012].
  107. Norman R, Matzopoulos R, Groenewald P, Bradshaw D. The high burden of injuries in South Africa. *Bull World Health Organ* 2007; 85: 695–702.
  108. Johnson LF, Dorrington RE, Bradshaw D, Pillay-Van Wyk V, Rehle TM. Sexual behaviour patterns in South Africa and their association with the spread of HIV: insights from a mathematical model. *Demogr Res* 2009; 21: 289–340.
  109. Department of Health (South Africa). Health Data Advisory and Co-ordination Committee (HDACC) report. Pretoria; 2012. Available from: <http://www.doh.gov.za/docs/reports/2012/hdaccreport.pdf> [cited 2 August 2012].
  110. Birnbaum JK, Murray CJ, Lozano R. Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organ* 2011; 89: 278–85.
  111. Sullivan M, Bicego GT, Rutstein SO. Assessment of the quality of data used for the direct estimation of infant and child mortality in the Demographic and Health Surveys. An assessment of DHS-I data quality. Demographic and Health Surveys Methodological Report, No 1. Colombia, MD: Institute for Resource Development/Macro Systems; 1990. pp. 113–43.
  112. Korenromp E, Arnold F, Williams B, Nahlen B, Snow RW. Monitoring trends in under-5 mortality rates through national birth history surveys. *Int J Epidemiol* 2004; 33: 1293–301.
  113. Bradshaw D, Dorrington RE. Child mortality in South Africa - We have lost touch. *S Afr Med J* 2007; 97: 582–3.
  114. Pardosi JF, Adair T, Rao C, Kosen S, Tarigan IU. Measuring Subnational Under-5 Mortality: lessons from a survey in the Eastern Indonesian District of Ende. *Asia Pac J Public Health* 2011: 1–11.
  115. Lopez AD, AbouZahr C, Shibuya K, Gollogly L. Keeping count: births, deaths, and causes of death. *Lancet* 2007; 370: 1744–6.
  116. Bradshaw D, Timaeus IM. Levels and trends of adult mortality. In: Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK, Hofman KJ, et al., eds. Disease and mortality in sub-Saharan Africa, 2nd ed. Washington, DC: The World Bank; 2006 pp. 31–42.
  117. Botha JL, Bradshaw D. African vital statistics - A black hole? *S Afr Med J* 1985; 67: 977–81.
  118. Dorrington RE. Estimates of the level and shape of mortality rates in South Africa around 1985 and 1990 derived by applying indirect demographic techniques to report deaths. Cape Town: University of Cape Town; 1988.
  119. Bradshaw D, Dorrington RE, Sitas F. The level of mortality in South Africa in 1985 - What does it tell us about health? *S Afr Med J* 1992; 82: 237–40.

120. Timaeus IM. Impact of the HIV epidemic on mortality in sub-Saharan Africa: evidence from national surveys and censuses. *AIDS* 1998; 12(Suppl 1): S15–27.
121. Dorrington RE, Timaeus IM, Moultrie TA. Death distribution methods for estimating adult mortality: sensitivity analysis with simulated data errors, revisited. Abstract submitted to the 2007 UAPS Conference. UAPS Conference; Arusha, Tanzania; 10–14 December 2007.
122. Udjo EO. A re-look at recent statistics on mortality in the context of HIV/AIDS with particular reference to South Africa. *Curr HIV Res* 2008; 6: 143–51.
123. Dorrington RE, Moultrie TA. A careful look at “A re-look at recent statistics on mortality in the context of HIV/AIDS with particular reference to South Africa”. *Curr HIV Res* 2008; 6: 279–85.
124. Timaeus IM, Jasseh M. Adult mortality in Sub-Saharan Africa: evidence from demographic and health surveys. *Demography* 2004; 41: 757–72.
125. Bradshaw D, Schneider M, Dorrington RE, Bourne DE, Laubscher R. South African cause-of-death profile in transition – 1996 and future trends. *S Afr Med J* 2002; 92: 618–23.
126. Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D. Identifying deaths from AIDS in South Africa. *AIDS* 2005; 19: 193–201.
127. Bradshaw D, Pillay-Van Wyk V, Laubscher R, Nojilana B, Groenewald P, Nannan N, et al. Cause of death statistics for South Africa: challenges and possibilities for improvement. Cape Town; 2010. Available from: [http://www.mrc.ac.za/bod/cause\\_death\\_statsSA.pdf](http://www.mrc.ac.za/bod/cause_death_statsSA.pdf) [cited 14 April 2012].
128. Dorrington RE, Bradshaw D. Maternal mortality in South Africa: lessons from a case study in the use of deaths reported by households in censuses and surveys. *J Popul Res* 2011; 28: 49–73.
129. Norman R, Bradshaw D, Schneider M, Pieterse D, Groenewald P. Revised burden of disease estimates for the comparative risk factor assessment, South Africa 2000. Methodological Note. Cape Town: South African Medical Research Council; 2006.
130. Donson H, ed. A profile of fatal injuries in South Africa, 2008. Annual Report for South Africa based on the National Injury Mortality Surveillance System (NIMSS). Cape Town: Medical Research Council; 2010.
131. Yudkin PL, Burger EH, Bradshaw D, Groenewald P, Ward AM, Volmink J. Deaths caused by HIV disease under-reported in South Africa. *AIDS* 2009; 23: 1600–2.
132. Grandin W, Westwood T, Lagerdien K, King MS. Deaths at Red Cross Children’s Hospital, Cape Town 1999–2003—a study of death notification forms. *S Afr Med J* 2006; 96: 964–8.
133. Westwood AT. Childhood deaths due to HIV—the role of the new death certificate. *S Afr Med J* 2000; 90: 877.
134. Burger EH, Groenewald P, Bradshaw D, Ward AM, Yudkin PL, Volmink J. Validation study of cause of death statistics in Cape Town, South Africa, found poor agreement. *J Clin Epidemiol* 2012; 65: 309–16.
135. Abdool Karim SS, Churchyard GJ, Abdool Karim Q, Lawn SD. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. *Lancet* 2009; 374: 921–33.
136. The Lancet. South Africa’s AIDS response: the next 5 years. *Lancet* 2012; 379: 1365.
137. Johansson LA, Westerling R, Rosenberg HM. Methodology of studies evaluating death certificate accuracy were flawed. *J Clin Epidemiol* 2006; 59: 125–31.
138. Pattaraarchachai J, Rao C, Polprasert W, Porapakkham Y, Pao-in W, Singwerathum N, et al. Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Popul Health Metr* 2010; 8: 12.

---

**\*Jané Joubert**

Edith Cavell Building  
 School of Population Health  
 University of Queensland  
 Herston Road, Herston  
 QLD, 4006, Australia  
 Email: [janetta.joubert@uqconnect.edu.au](mailto:janetta.joubert@uqconnect.edu.au)

## 2.5 Concluding remarks

International reviews of mortality data have painted a bleak picture of absent or deficient CRVS systems, limited alternative mortality data sources, incomplete death registration, and poor availability of cause-of-death data in most countries in Africa.<sup>3,5,8,9,43</sup> This mortality data source review, however, generates optimism for South Africa, showing a number of nationally representative mortality data sources, small-area universal data sources, and more, with a variety of mortality data items that could either be used as complementary to one another, or, where suitable, as verification or validation of the CRVS data. Indeed, the review suggests a data-rich and fortunate situation for South Africa in the context of the opinion by Hill and colleagues who report that “state-of-the-art health information systems consist of two complementary parts, first a universal and effective civil registration system, and second, a variable range of information from censuses and sample surveys in addition to civil registration systems”.<sup>14</sup>

Although there is indeed potential for South Africa’s health information system to benefit from two complementary mortality data systems, it is important to note that the South African review also found that data use and analysis based on these data collections have been challenged by a number of restrictions, including data quality problems; challenges related to timeliness and representativeness of data; and very limited analytic capacity for mortality measurement.

In order to coarsely capture the local availability and suitability of data for mortality measurement in South Africa, the table by Hill *et al*<sup>14</sup> has been adapted to reflect the findings from this review of South African mortality data sources. Table 2.2 below has been populated with information about the different data sources identified for South Africa. It is clear that the availability and suitability, in a broad sense, of data are not ideal or satisfactory for most alternative sources, particularly with regard to detail in cause-of-death data, representativeness or universality, and continuity of data collection.

The CRVS system appears to be the ideal approach for measuring key mortality indicators for South Africa at both the national and sub-national levels (Table 2.2). However, the broad criteria of availability and suitability of mortality data are inadequate to provide a comprehensive understanding of mortality data quality. In this light, suitable evaluation criteria with which to evaluate the quality of mortality data from civil registration in South Africa, will be sought in Chapter 3.

Table 2.2: Contributions, broadly in terms of availability and suitability for mortality measurement, of different approaches to measuring key mortality and population health indicators in democratic South Africa

	Level of estimate	Civil registration and vital statistics	Demographic surveillance sites	Rapid Mortality Surveillance System	Population censuses	Household sample surveys: cross-sectional	Household sample panel study	Audits: for child & perinatal, & maternal §§ events	Disease-specific registers	ASSA AIDS and Demographic Model
<b>Births</b>	National	Yes	No, but universal in site	N/A	Yes,* but once in 5 or 10 years	Yes*†	Yes*†	Facility-based	N/A	Yes
	Differentials	Yes	Limited to study site	N/A	Yes,* but once in 5 or 10 years	Yes*†	Yes*†	Facility-based	N/A	Yes
<b>Child mortality</b>	National	Yes	No, but universal in site	Not publically available, to be adjusted ‡‡	Yes,† but once in 5 or 10 years	Yes*†	Yes*†	Facility-based	N/A	Yes
	Differentials	Yes	Limited to study site	Not publically available, to be adjusted ‡‡	Yes,† but once in 5 or 10 years	Limited*†	Limited*†	Facility-based	N/A	Yes
<b>Adult mortality</b>	National	Yes	No, but universal in site	Not publically available, to be adjusted ‡‡	Limited, may be indirect ‡ ¶	Limited ¶	Limited ¶	N/A	N/A	Yes
	Differentials	Yes	Limited to study site	Not publically available, to be adjusted ‡‡	Limited, may be indirect ‡ ¶	Limited ¶	Limited ¶	N/A	N/A	Yes
<b>Maternal mortality</b>	National	Yes	No, but universal in site	Not publically available, broadly indicated ¶¶	Broadly indicated ††	Broadly indicated ††	No	Facility-based §§	N/A	No
	Differentials	Yes	Limited to study site	Not publically available, broadly indicated ¶¶	Broadly indicated ††	Broadly indicated ††	No	Facility-based §§	N/A	No
<b>Cause of death</b>	National	Yes	No, but universal in site	Not publically available, broadly indicated ¶¶	Broadly indicated ¶¶	Yes, limited** ¶¶	Broadly indicated ¶¶	Facility-based	No, facility-based	No
	Differentials	Yes	Limited to study site §	Not publically available, broadly indicated ¶¶	Broadly indicated ¶¶	Yes, limited** ¶¶	Broadly indicated ¶¶	Facility-based	Facility-based, cause-specific	No

\* With technical assessment of the reliability of the data, and possible adjustment, provided the method worked well, and quality control was good in the field.  
† For a recent or reference period by indirect methods; may be affected by recall and omission biases.  
‡ For an inter-censal period.  
¶ Methods measuring parental survival or sibling history and/or deaths in the household, may be affected by recall and omission bias.  
§ With verbal autopsy.  
\*\* For child deaths identified by a full birth history.  
†† Pregnancy/delivery-related question asked once death in household was indicated.  
‡‡ The Rapid Mortality Surveillance System capture only deaths of persons with a South African birth certificate or identity document, i.e. not all deaths are in this data source.  
¶¶ Causes of death are broadly indicated as natural/unnatural, or pregnancy/delivery-related, or accident/violence-related causes.  
§§ Refers to the Confidential Enquiry into Maternal Deaths.

Source: Adapted from Hill *et al* <sup>14</sup> to accommodate data sources from democratic South Africa





### 3 Chapter 3: Civil registration and vital statistics in South Africa: history, current status, and the quality of national mortality statistics

---

#### 3.1 Introduction

Based on the identified strengths and limitations of various mortality data sources in the data source review in Chapter 2, and the general criteria of availability and suitability of the data generated by these data sources, the CRVS system appears to be the preferred approach for measuring key mortality indicators for South Africa at both the national and sub-national level. This augurs well with the global importance attached to civil registration and vital statistics systems, and the wide recognition of civil registration as a preferred source of mortality data (compare Section 1.2.3).

A historical overview of the development of vital-event recording in South Africa, and the current status of the CRVS system will be provided in Section 3.2. Recognizing the insufficiency of using only the general criteria of data availability and suitability in forming an opinion of the quality of mortality data, suitable evaluation criteria for assessing the quality of country-level mortality data from CRVS systems will be sought and summarised in Section 3.3 and 3.4, with the aim of finding an answer to research question 2:

Research question 2: *What criteria can be used for evaluating national death and cause-of-death statistics from civil registration in South Africa?*

Different criteria for different purposes have been found in different settings, in different types of literature (Section 3.3). Utilizing those criteria considered most suitable for assessing the quality of national CRVS mortality data, South Africa's national statistics will be assessed in seeking an answer to the thesis' third research question:

Research question 3: *What is the quality of national mortality statistics from civil registration for the period 1997 – 2007?*

#### 3.2 CRVS in South Africa: a concise overview

There is overlap between Section 3.2.1 below and selected paragraphs in the published review article incorporated in Section 2.4.<sup>130</sup> A more extensive discussion in Section 3.2.1 was deemed

important in the lead-up to the published evaluation of CRVS data in Section 3.5,<sup>131</sup> as the latter allowed very limited space for illuminating the factors that shaped the historical development of the CRVS system.

### **3.2.1 Historical development**

Evidence exists of population enumerations in South Africa since the mid-1600s via diaries, journals, record rolls and colonial registers, mainly to identify voting and taxable populations.<sup>72</sup> In 1867, the first law governing the collection of statistics on births and deaths, Act 16 of 1867, was passed in one of the four colonies that existed at the time.<sup>72,132</sup> The remaining colonies followed from 1880 to 1906, each passing a different series of laws in a decentralized vital registration system.<sup>12,132</sup> In 1910, the Union of South Africa was constituted as a legislative union of the four colonies.

#### **3.2.1.1 Challenges facing CRVS systems over the 1900s**

Shortly thereafter, the modern history of the country's vital statistics system was introduced with the establishment of the national statistical office for South Africa, under the terms of Act 38 of 1914.<sup>132</sup> Though this resulted in the centralisation of vital registration, different laws and practices in the four colonies were superseded by uniform registration practice only after the Births, Deaths and Marriages Registration Act of 1923 was enforced in 1924.<sup>119,132</sup> The act was comprehensive, laid the foundation for the country's civil registration system, and remained in place for the following seven decades until replaced by a similarly comprehensive act in 1992, the Births and Deaths Registration Act 51 of 1992.<sup>12,107,119</sup> Between 1923 and 1992, however, the civil registration and vital statistics systems faced numerous challenges.

Unification in 1910 and a single national vital registration act in 1923 did not bring about an inclusive, successful civil registration system. Instead, at times it emulated a partial registration system with ideologically-engineered divisions based on 'homeland'/common-land, rural/urban, and population-group differentiation, with significant stunting of the system, prompted by legislation that led to under-registration in large parts of the majority population group, Africans. Shortly after unification, geographic segmentation loomed again with the passing of the Natives Land Act of 1913, restricting the leasing and ownership of land for Africans to designated reserves or 'homelands' that made up less than one-tenth of the land mass. These 'homelands'

were scattered over isolated geographic areas in the eastern half of the country, mainly in under-developed rural areas with very limited economic possibilities.<sup>133</sup>

Intensifying the segregationist and ‘homeland’ ideology, the Natives (Urban Areas) Acts of 1923 and 1945 divided South Africa into areas that were freely accessible (rural areas) and restricted (urban areas) for Africans. Under the 1936 Representation of Natives Act, only documented African workers with a ‘pass’ were allowed in the common (non-‘homeland’) areas of the country, thereby requiring the majority to remain in rural areas.<sup>132</sup> These acts strictly controlled the movement of Africans between urban and rural areas, and required their resettlement to the rural ‘homeland’ areas when unemployed, or employed but without a work permit or ‘pass’. The forced removal of Africans was not restricted to urban areas, but hundreds of thousands living in rural, non-‘homeland’ areas were also relocated to the ‘homelands’.<sup>133</sup>

Under the 1923 Births, Deaths and Marriages Registration Act, death registration for rural-living Africans was voluntary.<sup>12,132</sup> During the 1920’s, this implied that about 86% of Africans,<sup>134</sup> then making up nearly 68% of the total population,<sup>72</sup> were under no obligation to register a death. Moreover, while all deaths in urban areas had to be certified by a medical practitioner, or an inquest be held, medical certification was not ordinarily required for rural deaths.<sup>12</sup> While one act had the effect that rural living was compulsory for the majority population, another simultaneously ensured that registration was optional for them.<sup>132</sup> With 86% of Africans living in rural areas in the early 1920s, and the proportion still 61% in 1991<sup>134</sup> —and with this group increasing from 67% of the total population in 1911 to 79% in 1991<sup>72</sup>—it is clear how the civil registration act in itself impeded death registration for most of the majority population group.

From the middle of the 1900s, the performance of the civil registration system was further impacted by differentiation and segregation legislation, in particular with the passing of the Population Registration Act of 1950 and a number of acts furthering the ‘homeland’ ideology. The former act made provision for the compilation of a manual population register that, counter-intuitively, played a minor role in producing vital statistics. Birth, death and marriage certificates were produced by the office of the population register, but Khalfani *et al* report that the register was mainly used as a political tool to record every South African’s population group and to produce population group-based identity documents.<sup>72</sup> Under the act, all persons in the register was assigned to a population group and given an identity number that included a population-group identifier, thereby providing the legal basis for the population-group differentiation upon which future differential policy and practice would be based.<sup>119</sup>

Further fragmentation of the civil registration system was introduced with Act 81 of 1963, requiring vital registration of Africans to be handled at an office and administration system different to that for other population groups. For rural-living Africans, the registration of births and deaths remained optional under this act. Details of births and deaths were kept separately in a births register and a death register, respectively, and no link was made between these registers and the population register.<sup>132</sup>

Elaborating on segregation mechanisms, the Bantu Homelands Citizenship Act of 1979 required every African to become a citizen of one of the ‘homelands’.<sup>72,135</sup> With the granting of independence to four ‘homelands’—Transkei in 1976, Bophuthatswana in 1977, Venda in 1979, and Ciskei in 1981 (collectively known as the TBVC areas)—they became excluded from the South African statistical system, and were responsible for their own generation of vital registration information.<sup>132</sup> However, these states were largely incapable of doing this.<sup>136</sup> Although the remaining six ‘homeland’ areas chose against independence, several laws and circumstances hindered optimal civil registration in these areas.<sup>72</sup>

As well as the inhibiting effects of lacking infrastructure and resources in the largely underdeveloped ‘homeland’ areas, the reporting of vital events was likely restrained by the effects of forced resettlement on people’s motivation and means to do so. Forced resettlements included urban-to-‘homeland’ removals under the 1945 Urban Areas Act, intra-urban removals under the Group Areas Act of 1950, and rural-to-‘homeland’ removals under “agricultural betterment” or “strategic and infrastructural” schemes.<sup>133,137</sup> An estimated 2 million people were affected by such removals,<sup>133</sup> and it is very likely that their reporting of vital events was influenced by extreme demographic instability from vast-scale movements across significant political and social boundaries;<sup>138</sup> resultant complex and destructive social processes, such as family disruption due to the often sudden breakdown of conjugal relationships, rapid turnover of household membership,<sup>138</sup> and the dispersion of children to grandparents or other kin;<sup>139</sup> episodes of inter-ethnic conflict; threats by police and ‘homeland’ authorities during forced relocation;<sup>140</sup> and the struggle for livelihoods amidst extreme levels of unemployment and poor housing conditions in frequently marginalized, remote but often suddenly densely-populated areas with very limited economic possibilities.<sup>133,139,141</sup>

Hence, over time, the details of Africans in particular have become severely under-represented in the manual population register of the time.<sup>72,132</sup> In 1972, a computerised population register was initiated, but only captured the details of Coloureds, Indians/Asians and Whites.<sup>72,132</sup> It was

not until 1986, with the abolishment of restrictions on movement and residence, and the issuing of uniform identity documents across population groups, that civil details of Africans were also entered in the computerised population register.<sup>132</sup>

### **3.2.1.2 Transformation affecting CRVS since the 1990s**

During the 1990s, with the demise of the apartheid ideology and the dawn of the new democratic dispensation, major and rapid political and social transformation ensued in all spheres of governance, including a commitment to transforming the civil registration and vital statistics systems into an information system that could effectively serve civil record-keeping and reliably inform public-health planning.

Three key events during the 1990s held new promise for the coverage and content of vital registration and the population register in South Africa: (i) the passing of the Births and Deaths Registration Act, 1992 (Act 51 of 1992)<sup>107</sup>, leaving no scope for optional or differential registration; (ii) the adoption of the interim Constitution of South Africa (Act 200 of 1993),<sup>142</sup> ensuring the consolidation of the geographic segments of the country into one geo-political unit and, so, the centralisation of the civil registration system under one system of governance; and (iii) the agreement among three main role players, the Department of Health, the Department of Home Affairs, and the official national statistical agency, Stats SA, to establish in collaboration with health researchers a joint technical committee with the explicit purpose of enhancing civil registration and vital statistics. These events spawned a range of initiatives (compare Section 2.4 and 3.2.2) to increase the registration of births and deaths, and improve the quality of vital-event data in the country.<sup>4,106,119,143</sup>

### **3.2.2 Current status of CRVS in South Africa**

At present, the civil registration system remains centralised, with the registration of live births, stillbirths, deaths and marriages remaining under the mandate of the Department of Home Affairs; divorces under that of the Department of Justice and Constitutional Development; and the compilation and dissemination of vital statistics under the mandate of Stats SA. Statistics on recorded live births, marriages and divorces, and deaths and causes of death are published annually.<sup>4</sup>

A number of amendments have been made to the Birth and Death Registration Act of 1992,<sup>107</sup> of which the most recent is the Births and Deaths Registration Amendment Act 18 of 2010,<sup>144</sup> which

commenced on 1 March 2014. The different amendments were proposed and enacted to improve the coverage of vital event registration and enhance the utility of vital data.

A network of vital registration committees at national, provincial, regional, and district levels remains in place and is responsible for maintaining the initiatives that were implemented to improve registration of vital events. Among other things, these committees promote the culture of birth and death registration by organising regular advocacy and publicity campaigns in the form of radio broadcasts, posters, meetings, workshops, and interactive, educational radio programmes. Another continuing initiative is the inclusion of health officials in hospitals and clinics to raise awareness of the importance and advantages of registering births and deaths; to educate mothers-to-be about birth registration procedures during ante-natal visits; and to assist mothers with birth registration during delivery and post-natal visits at the facility itself. Arrangements are in place for either the health officials to forward batches of completed forms to the relevant office of the Department of Home Affairs, or for the Department of Home Affairs to collect the completed forms on a regular basis. Mobile trucks are being used to promote and perform registration of births and deaths at major public events, and to register events in hard-to-reach populations, such as residents in deep-rural areas.<sup>4</sup>

### **3.2.3 The CRVS system in South Africa: concluding remarks**

The overview of the historical development of South Africa's CRVS system highlights decades of movement control, non-compulsory rural death registration, limited or no birth and death registration services in rural areas, and differential administrative mechanisms. These circumstances point to potentially large numbers of unregistered deaths for most of the 1900s, in turn, pointing to severe stunting of the civil registration system, large biases in vital statistics databases, and numerous barriers to producing representative death and cause-of-death statistics.

However, key events and strategic initiatives under democratic rule have removed a number of these barriers and may have contributed to the improved quality of mortality statistics. Eleven years of more recent mortality data have been made available by Stats SA, and Section 3.5 aims to evaluate the quality of national mortality data from civil registration.

Before this evaluation, approaches in the literature to evaluating CRVS mortality data, and the criteria used for such evaluation, will be documented in Section 3.3.

### 3.3 Evaluating CRVS mortality data: evaluation approaches, frameworks, and criteria

#### 3.3.1 Organizations involved in evaluating health information and CRVS systems

A number of efforts have been developed to evaluate the quality of data generated by national health, demographic and development information systems. Efforts to evaluate mortality data include the work by the WHO, Health Metrics Network, the HIS Hub at The University of Queensland in Australia, population and statistics divisions of the United Nations, and advisory panels and expert groups affiliated with these organisations (cf. Section 1.1).

Extensive and valuable work contributing to the assessment of country health information systems has, for example, been done by the Health Metrics Network, with its assessment and monitoring tools.<sup>145,146</sup> However, as reported by Mahapatra *et al.*,<sup>9</sup> their methods have been designed to give a broad overview of all components of a health information system. Several of the Health Metrics Network/WHO framework criteria are based on a wide range of inputs and processes, rather than outputs, ranging from assessing policies and infrastructure for health information resources; financing and human resources in health; organization and management of information systems; country capacity and practices to collect and manage data; to monitoring health and disease programmes.<sup>145,147</sup>

As this thesis has a focus on the availability and quality of mortality data from CRVS systems, it was necessary to narrow down the available literature and concentrate on relevant the frameworks, tools and assessments of CRVS systems as developed by organisations like the United Nations, WHO, and the HIS Hub at The University of Queensland. The Statistical Office of the United Nations assembled official information on the CRVS systems of each country and non-self-governing areas as early as 1950, in order to evaluate the adequacy and international comparability of practices, procedures, definitions, tabulations, and the quality of vital data (including births, deaths, marriages, and divorces), obtained from countries' CRVS systems. Responses were received from over 100 areas, including 65 sovereign countries.<sup>15</sup> A similar study was conducted by the United Nations during the period 1976 – 1979, during which 112 countries/areas responded.<sup>16</sup> These studies provided a wealth of comparative information pertaining to a wide range of input, process and output themes and topics, including legal provisions for civil registration systems, administrative structures of vital statistics systems, tasks of civil registrars and vital-event informants, time and place of vital-event registration, availability of vital statistics, and the time reference for data tabulations.

Extensive and valuable work has been done at the HIS Hub within four themes: (a) building health information systems; (b) developing the health information workforce; (c) strengthening vital statistics and cause-of-death data; and (d) tracking progress toward Millennium Development Goals. Research, assessment frameworks, and tools with a focus on national CRVS systems have been conducted and developed, and successfully applied in various countries, including Bangladesh, India, Indonesia, Japan, Hong Kong, New Zealand, Sri Lanka, Thailand, and Turkey.<sup>36,148</sup> Comprehensive tools and guidelines are available for assessing and strengthening civil registration and vital statistics systems, and enhancing information on causes of death.<sup>149</sup> Similar to the breadth and depth of the Health Metrics Network/WHO assessment items for health information systems,<sup>145</sup> the HIS Hub assessment framework and tools<sup>18,19</sup> cover a broad range of CRVS items in components and sub-components that are mostly input and process-oriented, with three items in one component addressing data access and quality (compare E1 – E3 in output component E in the WHO Assessment Framework<sup>19</sup>).

### **3.3.2 Comparative evaluation of the quality of national mortality data and the criteria used in assessing national mortality data in institutional databases**

With regard to CRVS *mortality data* in particular, the WHO has included an appraisal of the quality of death and cause-of-death data since 1988 in the Introduction of the *World Health Statistics Annual*, supplying summary information on completeness and coverage of death registration; and medical certification, autopsies, and coding procedures of causes of death.<sup>150,151</sup>

Traditionally, the approach and criteria used to evaluate vital statistics were to assess the completeness of vital event recording (the proportion of all deaths that are registered into the civil registration system), and the coverage of such recording (the proportion of the national population that is covered by the civil registration system). These criteria are important and necessary, and have a key bearing on the quality and utility of vital data, but are not sufficient to comprehensively assess the quality and utility of vital data.<sup>17</sup>

In 1990, six plausibility and consistency checks used by the WHO to evaluate country-level cause-of-death data for inclusion in its statistical compilation have been published by Ruzicka and Lopez:<sup>150</sup> (1) using only rubrics in the ICD relevant to mortality—at that stage, either ICD-9 or ICD-8; (2) no incorrect assignment of causes of death with a clear age-sex dependency; (3) consistency between the proportionate distribution of deaths by cause and the estimated mortality level for that country; (4) consistency between the age-sex distribution of deaths for major causes



and that which is expected for each cause; (5) consistency of a current year's data with data from previous years; and (6) a clear indication of the number and proportion of deaths allocated to the ICD chapter labelled 'Symptoms, signs and abnormal clinical and laboratory findings not elsewhere classified' (cf Table 3.1).

Recognizing mortality data as the most widely available and commonly used data for informing health policy and programme development, and health research prioritisation, Mathers *et al*<sup>8</sup> analysed the countries' death registration data that were supplied to the WHO, to determine and compare among countries (1) the extent to which death registration systems cover deaths in the WHO regions. For each country, the authors also assessed (2) the timeliness of available data, (3) completeness of death registration, (4) the coding system used for causes of death, and (5) the extent to which deaths were coded to ill-defined categories.

Mahapatra *et al*,<sup>9</sup> stating that the utility of vital statistics depend on their quality, proposed in *The Lancet* 'Who Counts?' series a comprehensive framework for assessing the quality of vital statistics derived from civil registration. The authors distinguish between *general vital statistics* and *cause-of-death statistics*, proposing 12 and 15 criteria for each category, respectively (Table 3.1). A number of criteria correspond with those suggested by Ruzicka and Lopez and Mathers *et al.* listed above.

From the view point that a thorough understanding of the quality of the data from vital statistics is indispensable for health researchers and policy developers, Phillips *et al*<sup>17</sup> proposed a summary index to measure the performance of vital statistics systems to generate reliable data. The authors created a *Vital Statistics Performance Index*, consisting of six dimensions of vital statistics strength, as found in the country databases developed for the Global Burden of Disease Studies for 2010 and 2013.<sup>3,43</sup> These were: (1) quality of cause of death reporting, (2) quality of age and sex reporting, (3) internal consistency, (4) completeness of death reporting, (5) level of cause-specific detail, and (6) data availability/timeliness. Dimensions (1) – (5) were assessed using empirical indicators, and the sixth was assessed as a function of the combined values of the remaining five.<sup>17</sup> These dimensions correspond well with the criteria suggested above by Ruzicka and Lopez, Mathers *et al* and Mahapatra *et al.*, as demonstrated in Table 3.1.

### 3.3.3 Single-country evaluation of the quality of national mortality data and the criteria used in assessing national mortality data in countries' own databases

Since the institutional approaches during the 1990s by the United Nations and WHO to assessing the adequacy and quality of vital data, it was not until the early 2000s that mortality/population health researchers began to develop more targeted criteria for assessing country-level routine reporting systems for deaths and causes of death.

Building upon the six criteria used by Ruzicka and Lopez in the WHO collection of countries' mortality data,<sup>150</sup> Mahapatra and Rao used nine criteria in their assessment of mortality data in India: (1) content validity of lay reporting systems; (2) adequate coverage and compliance; (3) validity of statistics at sub-national levels of disaggregation; (4) minimal use of residual categories of causes of death; (5) consistency of cause-specific mortality proportions with general mortality levels; (6) absence of incorrect assignment of causes with clear age and sex dependency; (7) no incidence of improbable age and sex distribution by cause; (8) consistency of cause-specific mortality proportion over time; and (9) the timely compilation and publication of statistics.<sup>152</sup> Five of these nine criteria correspond with those suggested by Ruzicka and Lopez<sup>150</sup> (Table 3.1). Mahapatra and Rao<sup>152</sup> added a critical criterion to consider the effect of content validity of cause-of-death data based on verbal autopsy in light of the regular use of this method to collect cause-of-death data in India. Mahapatra and Rao also added criteria related to coverage and sub-national disaggregation, given India's very large total and state populations, and timeliness of statistics to ensure relevance and improve the utility of such statistics.<sup>152</sup>

Four years after the India assessment, Rao *et al*<sup>10</sup> published an evaluation of cause-of-death statistics from another country with a mega-population, China; conceptualizing an assessment framework based on four general data attributes encompassing nine criteria. The first general data attribute, *generalizability*, comprised the criteria (1) coverage and (2) completeness of death registration; *reliability* comprised (3) consistency of cause patterns with general mortality levels and (4) consistency of cause-specific mortality rates over time; *validity* included (5) content validity, (6) the use of ill-defined categories and codes, and (7) incorrect or improbable age or sex dependency; and the fourth attribute, *policy relevance*, covered (8) timeliness, and (9) geographical disaggregation of data. These nine correspond well with each of the criteria listed for other evaluations in Table 3.1, and have been found suitable by França *et al*.<sup>126</sup> to apply for the evaluation of cause-of-death statistics in Brazil. For each attribute and criteria, Rao *et al*<sup>10</sup> provided particularly useful conceptual underpinnings in the China study.

### 3.4 Suitable criteria for evaluating South Africa's national mortality data from CRVS

Seeking to answer research question 2 (*What criteria can be used for evaluating national death and cause-of-death statistics from civil registration in South Africa*), the criteria and dimensions used in the evaluation studies highlighted in Section 3.3 were considered against the historical limitations, recent transformation, and current status of the CRVS system (Section 3.2).

While criteria and data items related to *input* components of a CRVS system (such as a national legal framework, registration infrastructure and resources), and those related to *process* components (including death notification and registration practices, data storage and transmission, training of mortality coders) are certainly important factors in the 'production line' of mortality statistics, they are fairly well established and well-functioning in South Africa<sup>4</sup> (compare Section 3.2.2). On the other hand, concerns about an important *output* element, quality of CRVS mortality data, have been referred to in Section 2.4, pointing to the need to assess such data in an evaluation exercise.

Moreover, mortality statistics from routine civil registration is the only national mortality data source in South Africa with detailed annual cause-of-death information for all ages that could guide important decisions regarding public health research, resource prioritization, and health policy and programmes. It was therefore deemed important that the chosen criteria, when found to perform unsatisfactorily, should be able to identify potential sources of misinformation, bias, error, or inadequacy, for which the grounds could be identified and targeted recommendations made.

Against this background, it is clear that most of the criteria and items used in the broad assessments of health information systems are not suitable for the thesis aim to evaluate the quality of national mortality data from CRVS in South Africa. On the other hand, given substantial agreement in the choice of criteria in the country evaluations for India,<sup>152</sup> China,<sup>10</sup> and Brazil<sup>126</sup> (see Table 3.1), and in light of researchers having been able to assess these criteria successfully in other developing countries with developing statistical systems, it was considered suitable to use the framework and nine criteria, as used in the China and Brazil studies, for assessing national CRVS mortality data for South Africa. The evaluation is presented in a published research article in Section 3.5.

Table 3.1: Criteria for evaluating the quality of country-level mortality data from civil registration systems

Ruzicka & Lopez, 1990 <sup>150</sup> 6 criteria	Mahapatra & Rao, 2001 <sup>152</sup> 9 criteria	Mathers <i>et al</i> , 2005 <sup>8</sup> 5 criteria	Rao <i>et al</i> , 2005 <sup>10</sup> 9 criteria	Franca <i>et al</i> , 2008 <sup>126</sup> 9 criteria	Mahapatra <i>et al</i> , 2007 <sup>9</sup>		Phillips <i>et al</i> , 2014 <sup>17</sup> 6 dimensions
					General vital statistics (12)	Cause-of-death statistics (15)	
Using only rubrics in the ICD relevant to mortality—at the time, ICD-8 or -9		Type of cause-of-death coding used—at the time, ICD-9 or -10					Level of cause-specific detail, measured by number of distinct causes of death in data
No incorrect assignment of causes of death with a clear age-sex dependency	Absence of incorrect assignment of causes with a clear age and sex dependency		Incorrect or improbable age or sex dependency	Incorrect or improbable age or sex dependency		Improbable age or sex classifications	Internal consistency: medically impossible cause assignment for any given age or sex
Consistency between proportionate distribution of deaths by cause and estimated mortality level for the country	Consistency of cause-specific mortality proportions with general mortality levels for the country		Consistency of cause patterns with general levels of mortality	Consistency of cause patterns with general levels of mortality		Consistency between cause of death and general mortality	
Consistency between age-sex distribution of deaths for major causes and that expected for each cause	No incidence of improbable age and sex distribution by cause						
Consistency of a current year's data with data from previous years	Consistency of cause-specific mortality proportions over time		Consistency of cause-specific mortality rates over time	Consistency of cause-specific mortality rates over time	Comparable over time	Comparable over time	
A clear indication of number & proportion of deaths allocated to symptoms, signs and ill-defined conditions	Minimal use of residual categories of causes of death	Use of ill-defined and unspecified categories and codes	Use of ill-defined categories and codes for causes of death	Use of ill-defined categories and codes for causes of death		Use of ill-defined and unspecified categories	Quality of cause-of-death reporting, essentially use of 'garbage' codes <sup>153</sup>
	Content validity of lay reporting systems		Content validity of cause-of-death data	Content validity of cause-of-death data			
	Adequate coverage and compliance	Coverage of the resident population by death registration	Coverage of the resident population by death registration	Coverage of the resident population by death registration	Coverage of resident population by death registration	Coverage of deaths by medical certification	

<b>Ruzicka &amp; Lopez, 1990<sup>150</sup></b> 6 criteria	<b>Mahapatra &amp; Rao, 2001<sup>152</sup></b> 9 criteria	<b>Mathers <i>et al</i>, 2005<sup>8</sup></b> 5 criteria	<b>Rao <i>et al</i>, 2005<sup>10</sup></b> 9 criteria	<b>Franca <i>et al</i>, 2008<sup>126</sup></b> 9 criteria	<b>Mahapatra <i>et al</i>, 2007<sup>9</sup></b>		<b>Phillips <i>et al</i>, 2014<sup>17</sup></b> 6 dimensions
					<b>General vital statistics (12)</b>	<b>Cause-of-death statistics (15)</b>	
	Validity of statistics at sub-national levels of disaggregation		Geographical disaggregation of data	Geographical disaggregation of data	Availability of statistics for small areas	Availability of statistics for small areas	
	Timeliness of compilation and publication of mortality statistics	Timeliness of supplying death registration data to WHO	Timeliness of the release of mortality statistics	Timeliness of the release of mortality statistics	Timeliness in terms of production time	Timeliness in terms of production time	Timeliness and availability of data
		Completeness of death registration	Completeness of death registration	Completeness of death registration	Completeness of death registration	Completeness of deaths with medically-certified cause	Completeness of death reporting
					Missing data: % key variables with no response	Missing data: % of cause-of-death reports with no age/sex data	Quality of age and sex reporting
					Routine tabulations by 5-yr age groups and sex	Routine tabulations by 8 age groups and sex	
					Comparable across space	Comparable across space	
					Regularity of dissemination	Regularity of dissemination	
					Accessibility of data in various formats	Accessibility of data in various formats	
					Accessibility of metadata	Accessibility of metadata	
					Accessibility of user service	Accessibility of user service	
<p>Note: The evaluation studies are listed chronologically from left to right in the columns. Shaded columns are for single-country evaluations, compared to the comparative studies of those countries providing mortality data to the WHO, United Nations and/or GBD study, in clear columns.</p>							

Source: Criteria extracted from research papers cited in the header row of the table.

**3.5 Paper 2: “Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997 – 2007”**

# Evaluating the Quality of National Mortality Statistics from Civil Registration in South Africa, 1997–2007

Jané Joubert<sup>1,2\*</sup>, Chalapati Rao<sup>1</sup>, Debbie Bradshaw<sup>2</sup>, Theo Vos<sup>1,3</sup>, Alan D. Lopez<sup>1,4</sup>

**1** School of Population Health, The University of Queensland, Herston, Brisbane, Queensland, Australia, **2** Burden of Disease Research Unit, South African Medical Research Council, Parow Valley, Western Cape, South Africa, **3** Institute of Health Metrics and Evaluation, University of Washington, Seattle, Washington, United States of America, **4** Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia

## Abstract

**Background:** Two World Health Organization comparative assessments rated the quality of South Africa's 1996 mortality data as low. Since then, focussed initiatives were introduced to improve civil registration and vital statistics. Furthermore, South African cause-of-death data are widely used by research and international development agencies as the basis for making estimates of cause-specific mortality in many African countries. It is hence important to assess the quality of more recent South African data.

**Methods:** We employed nine criteria to evaluate the quality of civil registration mortality data. Four criteria were assessed by analysing 5.38 million deaths that occurred nationally from 1997–2007. For the remaining five criteria, we reviewed relevant legislation, data repositories, and reports to highlight developments which shaped the current status of these criteria.

**Findings:** National mortality statistics from civil registration were rated satisfactory for coverage and completeness of death registration, temporal consistency, age/sex classification, timeliness, and sub-national availability. Epidemiological consistency could not be assessed conclusively as the model lacks the discriminatory power to enable an assessment for South Africa. Selected studies and the extent of ill-defined/non-specific codes suggest substantial shortcomings with single-cause data. The latter criterion and content validity were rated unsatisfactory.

**Conclusion:** In a region marred by mortality data absences and deficiencies, this analysis signifies optimism by revealing considerable progress from a dysfunctional mortality data system to one that offers all-cause mortality data that can be adjusted for demographic and health analysis. Additionally, timely and disaggregated single-cause data are available, certified and coded according to international standards. However, without skillfully estimating adjustments for biases, a considerable confidence gap remains for single-cause data to inform local health planning, or to fill gaps in sparse-data countries on the continent. Improving the accuracy of single-cause data will be a critical contribution to the epidemiologic and population health evidence base in Africa.

**Citation:** Joubert J, Rao C, Bradshaw D, Vos T, Lopez AD (2013) Evaluating the Quality of National Mortality Statistics from Civil Registration in South Africa, 1997–2007. PLoS ONE 8(5): e64592. doi:10.1371/journal.pone.0064592

**Editor:** Tiziana Leone, London School of Economics, United Kingdom

**Received:** December 17, 2012; **Accepted:** April 16, 2013; **Published:** May 27, 2013

**Copyright:** © 2013 Joubert et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** No grant was applied for to conduct the study. However, the research was carried out while the first author was holding a University of Queensland Research Scholarship and Endeavour International Postgraduate Research Scholarship at the University of Queensland. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: janetta.joubert@uqconnect.edu.au

## Introduction

The importance of reliable, valid, comparable mortality data for measuring and improving population health is widely acknowledged, [1,2] yet few low and middle income countries have such data, [1,3] and even fewer have assessed the quality of their mortality data. [3] Periodic evaluations of mortality statistics by countries themselves are particularly useful to identify biases in local data, to shed light on the causes or sources of these biases, and to make locally-relevant recommendations to address these. [3,4] Among the few countries that have carried out such an exercise in the developing world are India, [5] China, [4] and Brazil [6].

A comprehensive country-specific evaluation of national mortality statistics has not yet been conducted for South Africa. However, based on 1996 data, two WHO assessments [1,3] rated South Africa in a group of countries with low quality mortality data, using as criteria completeness of death registration (<70%), use of ill-defined codes (>20% of registered deaths), timeliness and/or using an old revision of the International Statistical Classification of Diseases and Related Health Problems (ICD) or alternate list to code causes of death.

Although death registration was enacted in 1867 and the national statistical office established in 1914, partial coverage driven by variations in civil registration practices for different geographic areas and population groups, resulted in limited utility of national vital statistics for most of the twentieth century. [7–9]

Under the Population Registration Act of 1950, South Africans were classified as ‘Black African’, ‘Coloured’, ‘Indian/Asian’ or ‘White’. This population group classification was associated with disparities in different spheres of life, including civil registration practice, mobility, and residential access, all constraining the coverage of civil registration and completeness of death reporting.

With the emergence of democracy in the early 1990s, governance and public services, including civil registration, underwent major transformation through changing legislation, policy and practice. [10] Three key events played a major role to facilitate improvements in coverage and content of civil registration: (i) the passing of the Births and Deaths Registration Act of 1992, abolishing differential vital registration based on race and rural residence; (ii) the adoption of the Interim Constitution of South Africa in 1993, consolidating the geographically-segmented country into one geo-political unit which enabled the centralisation of the civil registration and vital statistics system (later confirmed in the Final Constitution of 1996); and (iii) a collaboration among strategic role players, i.e. the Department of Health, Department of Home Affairs, Statistics South Africa (StatsSA) and mortality researchers, focussing on enhancing the vital statistics system [8].

A range of initiatives followed these events, towards improving birth and death registration practice and the quality of vital data. These included the introduction of a new death notification form (DNF) that complies with WHO standards; establishment of provincial task teams to assist in implementing the new DNF; distribution of certification and ICD coding manuals for health staff; and guidelines on birth and death registration. [10] In addition, letters about relevant guidelines and new procedures were issued to all registered doctors; and birth registration forms were made available to mothers at the time of delivery, with health workers trained in all provinces to assist mothers to complete and submit these forms. [10] Finally, targeted capacity development was undertaken at StatsSA, to improve compilation and production of annual vital statistics for South Africa [8,10].

Given these initiatives, it is of considerable interest to ascertain whether or not they have had an impact on the quality of more recent civil registration data. Furthermore, South Africa’s cause-specific mortality patterns have previously been used to model mortality in many sub-Saharan Africa countries [11] and continue to be used by research and international development agencies as the basis for making estimates of cause-specific mortality in many African countries since there is very little information on mortality patterns elsewhere. [12] This adds to the critical importance of understanding the quality of South Africa’s mortality data. We therefore aim to evaluate the quality of national mortality statistics from civil registration for the period 1997–2007, to report areas of data strengths and weaknesses, and to make recommendations for improvements in data quality.

## Methods

### Ethics Statement

The mortality data used in this study were obtained from the official statistics agency of South Africa, StatsSA, collecting and providing data under the provisions of the Statistics Act of 1999. [13] To obtain mortality rates, we used population data from a publically-available electronic data source of the Actuarial Society of South Africa (ASSA). [14] Ethical clearance for research involving human participants was not sought as the datasets are anonymous and contain no identifiable information of any study participant. Ethical concerns regarding participant consent and possible negative consequences to study participants have been

taken note of, but are not relevant to the study as the study ‘participants’ are deceased persons. In interactions with collaborators in the host country, however, relevant ethics considerations such as respect for local customs and legal requirements regarding data-use have been upheld.

### Assessment Framework

In the past, completeness of death registration was commonly the only assessment criterion applied to evaluate the quality of national vital statistics. [3] However, as awareness of the usefulness of cause-of-death statistics increased, more assessment criteria were proposed. [15] These criteria have been expanded and used in a framework of which the origin [5,15] and conceptual underpinnings [4,6] have been described elsewhere. To evaluate South Africa’s mortality data, we built on earlier country-specific evaluations, [4–6] employing the general attributes and criteria as defined in the China study [4]:

- a. Generalizability, informed by the criteria (1) coverage and (2) completeness of death registration;
- b. Reliability: (3) epidemiological consistency and (4) temporal consistency;
- c. Validity: (5) content validity, (6) use of ill-defined and non-specific codes, and (7) use of age- and sex-improbable classifications; and
- d. Policy relevance: (8) timeliness, and (9) availability of sub-national data.

Each criterion is rated with three broadly-defined evaluation measures: “satisfactory”, “unsatisfactory” or, where the information is unavailable or insufficient, “unknown”. For differentiating between “satisfactory” and “unsatisfactory”, we employ the thresholds suggested in previous studies [4,6].

### Data Sources

For five criteria (*coverage, completeness, timeliness, sub-national availability and content validity*) information was reviewed in relevant legislation, statistical releases, web-based data repositories, research and government reports, and scholarly journals to inform about developments over time which shaped the current status of these criteria in terms of data adequacy.

For the remaining four criteria (*epidemiological consistency, temporal consistency, age/sex classification, and ill-defined/non-specific codes*) the evaluation draws on a dataset produced by StatsSA with 11 years’ mortality data from DNFs for 5.38 million deaths that occurred nationally from 1 January 1997 to 31 December 2007. [16] This dataset comprises of deaths certified according to the following practices. In cases of natural deaths with access to a medical practitioner, the 1992 Act requires the practitioner to complete a DNF (Form BI-1663). The DNF also makes provision for a registered professional nurse to do so. If neither is available, as may happen for example in remote rural areas, a Death Report (Form BI-1680) must be completed by an authorized traditional leader (headman/chief), member of the police service, or funeral undertaker to certify the death and describe the circumstances that led to the death. [17,18] Unnatural deaths are subject to medico-legal investigation in terms of the Inquests Act of 1959. On receipt of the DNF or Death Report by the Department of Home Affairs, the death is registered into the electronic civil registration system. Hereafter, the forms are collected by StatsSA where trained nosologists code all causes to ICD-10 3-digit codes. [19] Underlying causes are derived automatically with the Automated Classification of Medical Entities software (ACME 2000.05) [20].



## Assessing the General Attributes and Criteria

**Generalisability**, or the extent to which mortality statistics are representative of the population under study, was assessed using the criteria *coverage* and *completeness*. *Coverage* refers to the extent of inclusion of different sectors of the population in the civil registration system, such as geographical sectors (e.g. urban/rural, or sample-based areas); administrative sectors (e.g. provinces, states or districts); or population groups based on country-specific categorizations. *Completeness* refers to the extent to which deaths within the covered population are reported into the civil registration system. For *coverage*, we reviewed and summarised the effect of legislation and policies that mandated and/or constrained geographic, administrative and population coverage of death registration over the past 150 years. Due to unrepresentativeness of the total population and the potential of introducing biases into the data, *coverage* of less than the total population is deemed ‘unsatisfactory’. For *completeness*, published estimates of under-registration of deaths were reviewed. Because of the need to measure the patterns and rates of mortality in a population with the minimum biases, completeness of less than 90% of the covered population is rated ‘unsatisfactory’.

**Reliability** relates to the consistency of mortality data with regard to established epidemiological expectations. For this general attribute, we evaluated two criteria: *epidemiological consistency* and *temporal consistency*. *Epidemiological consistency* of the South African data was evaluated using methods similar to those used in previous country evaluations of national vital registration systems, [4,6] and a variation thereof. Based on the premise that the composition of mortality by cause changes systematically as all-cause mortality decline, [21–23] observed broad patterns of causes of death were compared with expected broad-cause values considering the relationship between the overall level of mortality and the relative contribution of causes to the overall level. The country’s gross domestic product (GDP) is used as a covariate in the model. Such evaluation is based on the theory of the epidemiological transition, according to which declines in all-cause mortality are accompanied by shifts in proportionate mortality: in high-mortality populations, communicable, reproductive and nutritional conditions predominate, whereas chronic and degenerative conditions predominate in low-mortality populations. [21] A historical dataset of international vital registration data was analysed by Salomon and Murray [23] to develop regression models that predict cause-specific compositional mortality by broad cause groups, for given inputs of all-cause mortality by age and sex. The three broad-cause groups are (1) a combined group of communicable diseases, maternal, neonatal and nutritional causes, (2) non-communicable disease, and (3) injuries, as defined in the Global Burden of Disease 1990 study [24].

To assess *epidemiological consistency*, the model predictions by age, sex and broad cause were compared with observed proportions for South Africa. A difference of more than two standard deviations ( $>2$  SD) between observed and predicted proportions suggests unsatisfactory epidemiological consistency of the observed data, unless there are plausible epidemiological reasons for such departures. [4] We used national mortality data by age and sex from civil registration for 2007; population estimates for 2007 from the ASSA2008 AIDS and Demographic Model (ASSA2008) of ASSA; [14] and 2007 GDP estimates from StatsSA [25] to derive model-predicted broad-cause proportionate mortality by age and sex. At first, we compared the broad-cause proportions derived from the cause-of-death models with observed proportionate mortality for South Africa. However, as the compositional cause of death models are based on mortality schedules from countries and time periods not affected by HIV/AIDS, we also compared the

model-based predictions with observed broad-cause proportions after excluding from the observed data the large numbers of death due to HIV/AIDS for 2007 as estimated in preparation for the second South African National Burden of Disease study [26].

*Temporal consistency* was evaluated by examining whether proportionate mortality from 10 leading causes or cause-groups changed in a predictable manner over time in the period 1997 to 2007. This criterion is informed by the proposition that proportionate mortality from different causes changes in a predictable manner over time as overall mortality changes with socio-economic development. [21,23] In the absence of substantial natural disasters, pandemics, or revisions to the classification of diseases, a consistent trend in cause-specific mortality should be observed. Where such impacts occurred, as in the case of the substantive HIV/AIDS epidemics in sub-Saharan Africa, observed cause-specific mortality trends would be expected to reflect increased deaths resulting from the epidemic. We investigated the trajectory over time of malignant neoplasms, ill-defined natural causes, external causes, and infectious and parasitic disease which were among the most commonly-reported categories or groups of disease during the 11-year period. We also examined tuberculosis, lower respiratory infections, diarrhoeal disease, ischaemic heart disease (IHD), stroke, and diabetes, counting among the most commonly-reported communicable and non-communicable single causes for 1997–2007 and ranking among the 10 leading single causes in the South African National Burden of Disease Study, 2000 [27].

For the attribute **validity**, we sought to assess the extent to which mortality data show what they purport to show, and to assess the extent of insufficiently- and inappropriately-attributed causes of death. Three criteria were assessed based on information on DNFs. *Content validity* (criterion 5) was assessed by reviewing local studies that examined the accuracy of cause attribution. Like inaccurate cause attribution, the *use of ill-defined or non-specific codes* (criterion 6) is a large impediment to local usefulness and international comparison of cause data. A proportion larger than 10% of total deaths assigned to ill-defined or non-specific codes was considered unsatisfactory. Aggregated data for 1997–2007, nationally and by province of death occurrence, were analysed to identify the extent of Chapter R codes (Symptoms, signs and ill-defined conditions); three non-specific cancer codes (C76, C80, C97); two major ill-defined cardio-vascular disease (CVD) causes (heart failure (I50) and cardiac arrest (I46)); and injuries of undetermined intent (Y10–Y34). Additionally, to compare the extent of R codes by age, R codes in each age group were calculated as a percentage of total deaths in each age group. Finally, the number of deaths coded to R codes was calculated by province for each year to compare the trajectory of R codes to that of the total number of deaths over the eleven years for each province.

Criterion 7, *use of age- and sex-improbable classification*, is guided by the observation that certain conditions occur primarily in specific age ranges, or cause sex-specific mortality. Departures from anticipated age/sex patterns raise concern about the quality of cause data. The aggregate dataset was examined for departures from 10 sex-specific conditions comprising maternal causes of death and genital tract cancers (Text S1). Age patterns were examined for plausibility and consistency in 27 typically age-dependent causes/cause groups: maternal conditions, perinatal conditions, 16 cancers, cardiovascular disease, and suicide (Text S1). In addition, unadjusted annual age-specific death rates were calculated over the 11-year period for three leading cause groups, i.e. cerebrovascular disease, malignant neoplasms, and IHD, to assess plausibility across age from the raw data. Patterns of age-

and sex-specific rates were examined from the aggregated unadjusted deaths from cerebrovascular disease by province of death occurrence, and nationally, to assess age-consistency across the provinces.

**Policy relevance** was evaluated by assessing *timeliness* of the release of mortality data (criterion 8) and *availability of sub-national data* (criterion 9). These criteria, respectively, are informed by the proposition that out-of-date mortality data are of little relevance for policy and intervention purposes, and that nationally-aggregated data are insufficient to identify local health differentials and needed interventions by health jurisdiction. *Timeliness* was assessed by examining the time gap between the end of the reference period (year of death) and the time of publication of final tabulations. A lag of two years was considered a reasonable threshold. [4] Criterion 9 was evaluated by assessing the public availability of geographically-disaggregated data in paper and electronic reports, online data repositories, and unit record data, at least at provincial level.

## Results

### Criterion 1: Coverage

Death registration, enacted since 1867, was effectively a partial process for most of the 1900s as coverage was constrained by differential registration practice based on geographical segmentation and population segregation imposed by various acts under ‘homeland’ and apartheid policies. [8,28] During the 1990s, the ‘homeland’ ideology was abolished under democratic rule, and the country was geographically unified under one government. With the Births and Deaths Registration Act of 1992, death notification became a national, inclusive legal requirement for all people in all geographic areas. Hence, *coverage* is rated satisfactory.

### Criterion 2: Completeness

Until the 1980s, national completeness of death registration was largely unknown. [29] However, pivotal work has been done by mortality researchers during the 1980s and 1990s, using different methods and data-sources in search of plausible results. [29–33] After consolidating fragmented data sources, and taking into account the heterogeneity of the age structures of the four population groups, distortions introduced into the data by the rapidly- but differentially-spreading HIV/AIDS epidemic, and administrative complications of the data, Dorrington et al. [31] estimated national completeness. Applying Bennett and Horiuchi’s Synthetic Extinct Generations Method (SEG) [34] with deaths from civil registration and the Population Register, relative to population estimates from the ASSA600 AIDS model, [35] large improvements were observed in national completeness of adult deaths, increasing from 73% in 1994 to 89% in 1999/2000. [31] (Fig. 1).

A new opportunity to estimate completeness arose in 2001 with the second all-inclusive census under democracy. [36] Using population counts from the 1996 and 2001 censuses, and inter-censal deaths from civil registration, Dorrington and colleagues [32] estimated completeness with Hill’s Generalized Growth Balance (GGB) [37] method with an adaptation to account for migration. [32] National completeness levels were estimated at 83.4% for males and 84.5% for females. [32] (Fig. 1) In 2007, another nationally-representative data source, the Community Survey, [38] offered further opportunity to assess completeness. Investigating maternal mortality in South Africa, Dorrington and Bradshaw [33] used population counts from the Community Survey and the 2001 census, and deaths from vital registration, and estimated national female completeness for 2001–2007 at

91% with the GGB method, and 89%, using the SEG method (Fig. 1). Furthermore, using the Preston and Hill method [39] with inputs from StatsSA mid-year population estimates and vital registration deaths, StatsSA has estimated completeness to have improved to levels of 87% for males and 79% for females in 2007 [40,41].

Large improvements have been estimated for registration of childhood deaths. For the period 1996–2006, Darikwa and Dorrington [42] estimated infant death registration to have improved from 43% to 89%; registration for children 1–4 years, from 43% to 57%; and for children under 5 years, from 44% to 78% (childhood completeness not presented in Fig. 1). These estimates were derived using a multi-stage method using registered death data from civil registration, data from the 2007 Community Survey (children ever born/children surviving data, data on the survival of the last child born to women aged 12–49 years; and child deaths over the past 12 months as reported by households); 2001 Census (reported household deaths); and previous research which used mortality data based on the 1998 SADHS and 1996 Census [42].

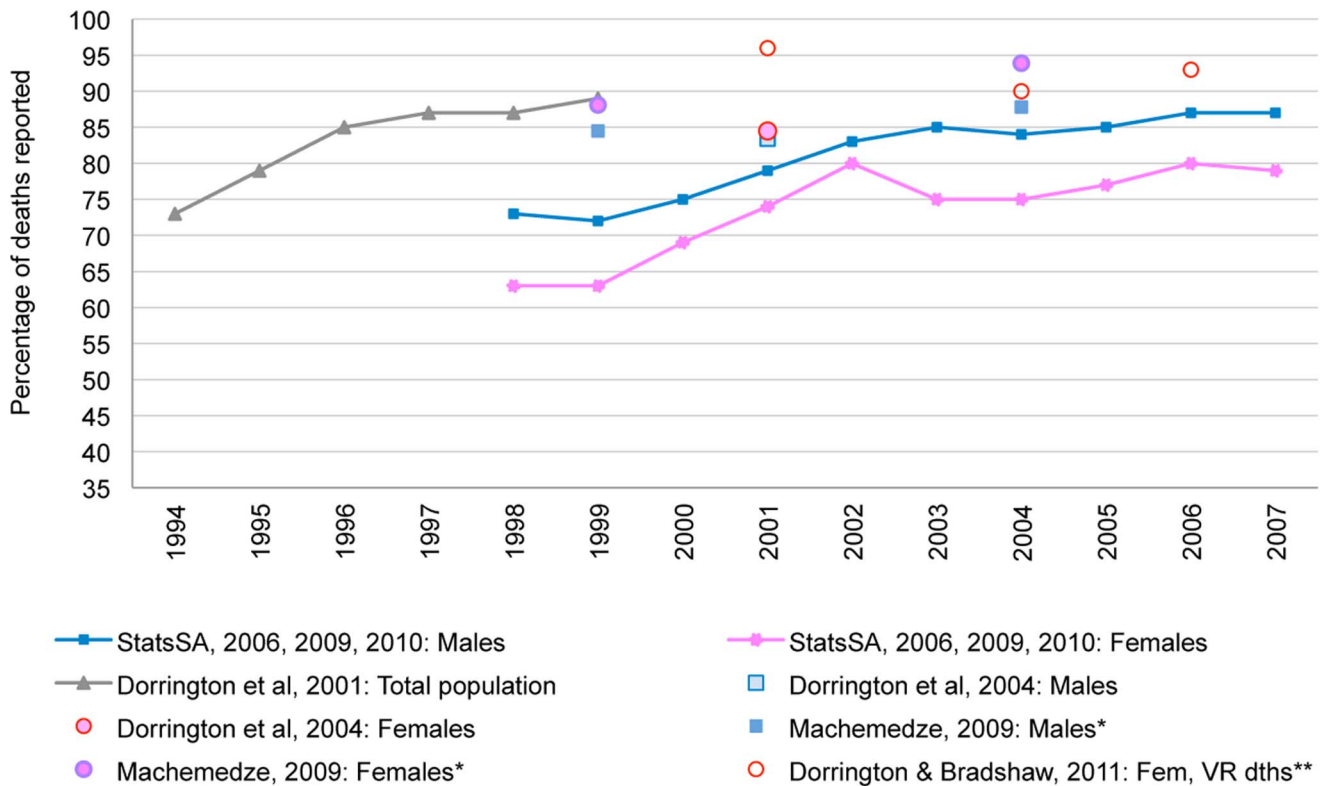
Figure 1 shows a trend of ongoing improvements over time, starting to satisfy the 90% threshold by the early 2000s, while more recent estimates by Dorrington and Bradshaw [33] and Machedmedze [43] exceed 90%. For the Global Burden of Disease Study 2010, Wang and colleagues estimated completeness in South Africa at 95% since 2000, [12] and a satisfactory rating is given for *completeness*.

### Criterion 3: Epidemiological Consistency

Large differences were found between the observed and predicted proportional broad-cause mortality patterns for 2007 (Fig. 2A) before taking HIV/AIDS deaths out of the schedule. Considerably better congruence was found between predicted and observed proportions after removing HIV/AIDS deaths (Fig. 2B). However, close to half of male and 41% of female values still deviated by more than two standard deviations from mean predicted proportions in Fig. 2B. These findings have two interpretations. Firstly, the differences in Figures 2A and 2B indicate that HIV/AIDS mortality is not adequately predicted by the cause-of-death models used in this analysis. Secondly, the persistence of deviations after exclusion of HIV/AIDS deaths, particularly for Group 1 conditions in males, could either be a real phenomenon of higher proportionate mortality from infectious diseases, or an artefact of data quality in terms of cause-of-death assignment. On the basis of differences between the South African epidemiological profile and that represented in the model predictions, as well as the observed residual differences on exclusion of HIV/AIDS deaths, this criterion could not be assessed conclusively as the model does not have the discriminatory power to enable an assessment for South Africa.

### Criterion 4: Temporal Consistency

Temporally consistent trends are shown over the 11-year period with no large annual fluctuations in cause-specific mortality proportions for IHD, diabetes, stroke, malignant neoplasms and ill-defined natural causes (Fig. 3). Rising proportions of selected communicable diseases are consistent over time and in line with expected epidemiological patterns associated with the HIV/AIDS epidemic and the effect of misclassified HIV/AIDS deaths. [44], [45] Given trends that are consistent and concurring with local epidemiological experiences, the data is rated satisfactory for this criterion.



**Figure 1. Estimated completeness of death reporting in South Africa: 1994–2007.** Notes: \*1999 and 2004 are used as midpoints between Census 1996 and Census 2001, and Census 2001 and Community Survey 2007, respectively. \*\*Estimates apply to ages 5–85 years. 2004 was used as midpoint for the period 2001–2007. While the source [33] includes separate GGB and SEG estimates, this graph shows an average of GGB and SEG estimates combined. GGB – Generalized Growth Balance Method; SEG – Synthetic Extinct Generations Method. **Source:** Compiled by the authors from: Stats SA, 2006, [41] 2009, [40] 2010; [17] Dorrington et al., 2001; [31] Dorrington et al., 2004; [32] Machedmedze, 2009; [43] Dorrington & Bradshaw, 2011 [33]. doi:10.1371/journal.pone.0064592.g001

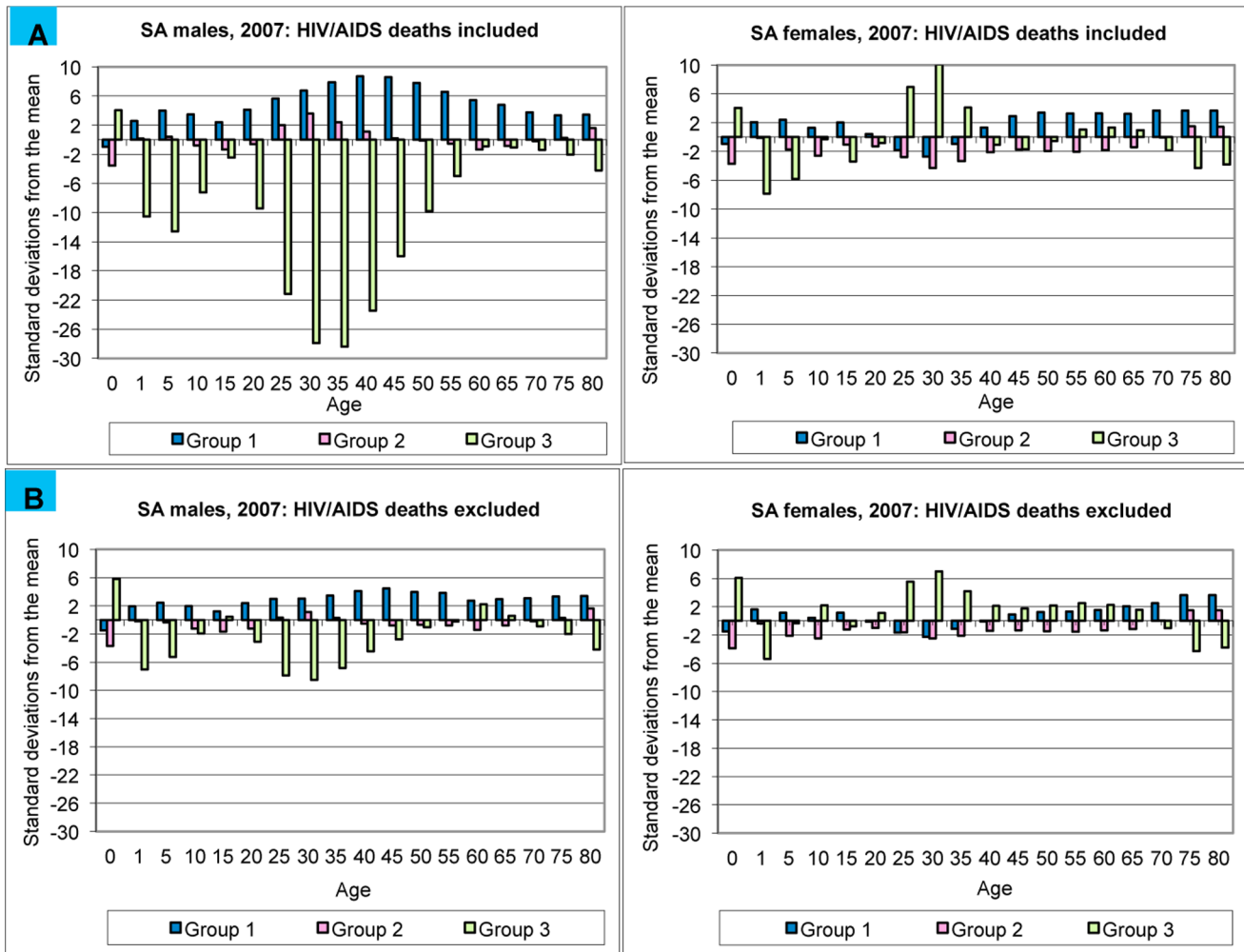
### Criterion 5: Content Validity

The extent to which certification and coding of causes of death are accurate is not routinely assessed, neither has it been studied in a nationally-representative sample of deaths. However, a few studies have shed light on the accuracy of cause attribution. Most of these studies have focussed on the misclassification of HIV/AIDS deaths, and most were limited in size and geographical coverage. Westwood studied death certificates and medical records at a teaching hospital in Cape Town among 500 paediatric deaths, and found that the under-certification of HIV-related deaths was over 11%, and that a further 30% of HIV-related deaths were classified using non-specific terms that could result in inaccurate classification. [46] Grandin *et al.* reported that, although HIV/AIDS accounted for the leading cause of death, causing 32% of all deaths in their study for 1999 to 2003 in a pediatric teaching hospital in Cape Town, the findings are likely under-estimating the role of HIV in causing death in this setting. [47] Comparing DNFs with medical records for 683 deaths in Cape Town, Yudkin *et al.* estimated that 36% of deaths attributable to HIV were classified to other conditions. A further 37% were indicated by euphemisms such as immune suppression or retroviral disease. [48] In a retrospective review of 242 DNFs of deaths occurring in an academic hospital in Cape Town during 2004, Nojilana *et al.* [49] found that the under-reporting of HIV/AIDS was 53%. Analysing vital registration data for 2000–2001, Groenewald *et al.* found 61% of HIV/AIDS deaths classified to other

conditions. [44] Birnbaum *et al.* confirmed substantial misclassification of HIV/AIDS in registered deaths for 1996–2006, suggesting that over 90% of HIV/AIDS deaths were misattributed to other causes. [45] In a recent validation study of 703 deaths in Cape Town, substantial misclassification was found not only for HIV/AIDS, but also for IHD, hypertensive disease, and diabetes. [50] *Content validity* is rated unsatisfactory based on these findings.

### Criterion 6: Use of Ill-defined and Non-specific Causes

Nationally, 12.8% of registered deaths were assigned an R code during 1997–2007, ranging from 12.2% in 1999 to 13.9% in 2007 (Table 1). Typically, about 11% of adult (aged 15–64) deaths had R codes assigned, increasing steadily over the older age groups to double this at ages 80+ (Fig. S1A). Close to two-thirds of R codes were found in persons aged 0–64 years and 36% in ages 65+. Cause-specific proportions from ill-defined and non-specific conditions show substantial differences among the provinces (Table 1). While the number of registered deaths increased dramatically in all nine provinces over time, even doubled in six, the trajectory of Chapter R codes are very different, with much less dramatic increases in absolute numbers (Figures S1B+C). However, with over 10% of causes assigned an R code, and a quarter assigned R- and non-specific codes collectively, this criterion is rated unsatisfactory.



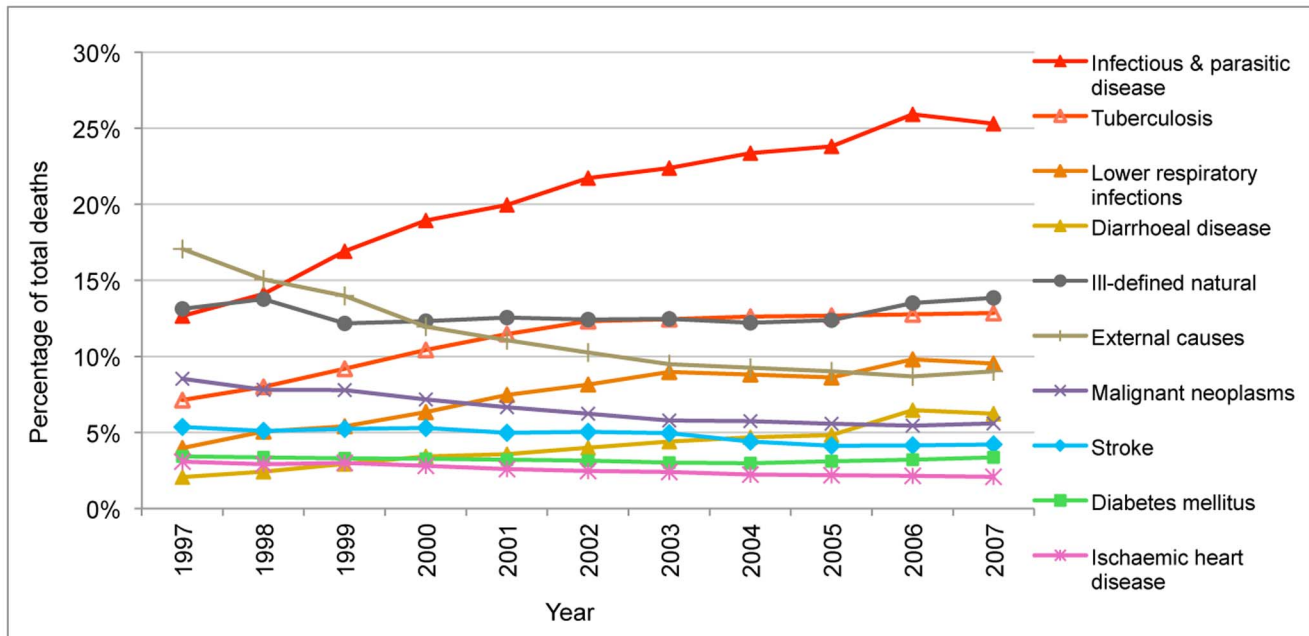
**Figure 2. A. Number of standard deviations by age and sex for South Africa, 2007: HIV/AIDS deaths included.** The number of standard deviation by which observed broad-cause mortality proportions differ from mean predicted proportions when the estimated number of HIV/AIDS deaths are included in the analysis. **B. Number of standard deviations by age and sex for South Africa, 2007: HIV/AIDS deaths excluded.** The number of standard deviation by which observed broad-cause mortality proportions differ from mean predicted proportions when the estimated number of HIV/AIDS deaths are not included in the analysis. **Source:** Mortality data for 2007 from StatsSA vital registration data; [16] HIV/AIDS estimates from Bradshaw et al.; [26] population data from ASSA2008; [14] GDP data from StatsSA [25]. doi:10.1371/journal.pone.0064592.g002

### Criterion 7: Use of Age- and Sex-improbable Classifications

Departures from uniquely sex-specific causes were found in less than one in a million deaths, and from age-dependent causes, in less than one in two thousand deaths (Text S1). Annual age-specific death rates from the raw data for two leading single causes (stroke and IHD) and one leading cause category (malignant neoplasms) reflect the typical increase with increasing age consistently over the 11-year period. Small inconsistencies appear at ages 80 and over for 1997–1999, possibly reflecting age misreporting, problems with cause attribution in older patients with multiple cause presentation, or lower certification and coding accuracy during earlier years of the reference period (Fig. S2A). Age and sex patterns of unadjusted stroke deaths by province appear plausible over age, though starting at relatively young ages (30 years). These patterns, however, are consistent over age for both sexes and across the provinces (Fig. S2B). Criterion 7 is therefore rated satisfactory.

### Criterion 8: Timeliness

Prior to 1994, the extraction and processing of mortality data from DNFs were tedious. After recording a death on the population register, images of the DNFs were put on microfilm, microfilmed images were manually verified against the original DNFs, DNFs were manually indexed and then sent to StatsSA for processing. The verification process, in particular, could take very long. [8] With limited staff training, the classification and coding of causes of death also contributed to delays, [51] and substantial hold-ups followed in the production of mortality statistics during the 1990s. For example, up to November 2001, the latest cause-of-death data available were from 1996, and in December 2001, cause data for 1997–2000 were released. Post-democracy, however, delays have been addressed by concerted multi-stakeholder efforts with a shared common vision to enhance vital statistics; a driving process encompassing motivation and learning from expert opinion; and international donor assistance aimed at staff training in several aspects of vital statistics compilation. [8,17] Over the past decade, further improvements followed with



**Figure 3. Proportion of total deaths due to leading categories and causes of death, 1997–2007.** Source: Vital registration data from StatsSA [16]. doi:10.1371/journal.pone.0064592.g003

advances in electronic data compilation, advanced training of nosologists by Australian and United States experts, and automation of cause attribution. [8] Shorter waiting times and improved regularity followed, such that, since 2007, annual reporting was done consistently on deaths that occurred in the calendar year two years prior to publication. [52] Hence, *timeliness* is rated satisfactory.

#### Criterion 9: Sub-national Availability

South Africa has nine provinces, 46 district municipalities and 231 municipalities. While statistical releases, published in 1999 [53] and 2001, [51] contained limited mortality data by province as the lowest level of disaggregation, recent reports include numerous tabulations by province and district municipality. At provincial level, web-based tables can be created with a number of DNF variables, including underlying cause. The DNF provides information on the place of death occurrence, place of registration of the death, and place of usual residence of the deceased, holding

potential for a wide range of mortality and epidemiological analysis at sub-national level. Anonymous unit-record cause data by province are available electronically upon request, subject to data-use agreements. This criterion is thus rated satisfactory.

#### Discussion

Civil registration data were rated satisfactory for six of nine criteria. These six indicate satisfactory performance in terms of generalisability and policy relevance, and partial satisfaction with reliability and validity of the data. The two criteria rated unsatisfactory, *content validity* and *use of ill-defined and non-specific codes*, signify considerable shortcomings with cause-of-death data. Information on content validity, particularly for infectious disease mortality, would clarify the performance rating for the remaining criterion, i.e. *epidemiological consistency*. Overall, the findings from this evaluation indicate that while there have been substantial improvements in the performance and quality of selected aspects of civil registration mortality data, attention needs to be paid to

**Table 1. Percentage of total deaths assigned selected ill-defined and non-specific codes by province of death occurrence, South Africa, 1997–2007.**

	Western Cape	Eastern Cape	Northern Cape	Free State	KwaZulu-Natal	North West	Gauteng	Mpumalanga	Limpopo	South Africa
Ch. R codes	6.0	17.5	7.5	9.6	15.0	9.9	12.4	8.5	17.9	12.8
Non-spec. cancer	1.5	0.6	0.7	0.5	0.5	0.4	0.8	0.4	0.4	0.6
Ill-def. CVD	3.0	3.0	4.0	4.5	3.5	5.1	3.8	3.4	3.8	3.7
Ill-def. injury	12.3	6.6	6.3	5.4	6.9	5.7	10.4	7.0	4.4	7.6
All four categories	22.8	27.7	18.5	20.0	25.9	21.1	27.4	19.3	26.5	24.7

Notes: Chapter R codes: R00–R99 (Symptoms, signs and ill-defined conditions); Non-specific cancer codes: C76, C80, C97; Ill-defined cardiovascular disease (CVD) causes: heart failure (I50) and cardiac arrest (I46); Ill-defined injury: injuries of undetermined-intent, Y10–Y34.

Ch. – Chapter; Non-spec. – Non-specific; Ill-def. – Ill-defined; CVD – Cardiovascular disease.

doi:10.1371/journal.pone.0064592.t001

cause attribution before the data are deemed optimal for population health assessment and epidemiological research.

Substantial improvement has been observed in the *completeness* of death registration over a short period of less than two decades. There remains a need, though, for focussed attention to improve registration of deaths in children aged 1–4 years. While registration of child deaths is usually lower than that of adult deaths, [54] this can be changed, as demonstrated through the doubling of completeness of infant death registration in the past decade. [42] Further improvements in registering child deaths would be a significant step towards more accurate estimates of under-5 mortality from vital registration, with nominal adjustment. The availability of reliable local measures of under-five mortality is of critical importance for targeted public health interventions to help meet the United Nations Millennium Development Goals 4 & 5 for South Africa.

Despite the overall trend towards higher levels of completeness of adult death registration, the actual estimates continue to reflect differences. The limitations of indirect demographic methods used for estimating completeness are known, [55] largely on account of incompatibility of assumptions underlying the individual methods. For instance, death distribution methods are based on the assumption that the study population is closed to migration. Although adjustments can be made to allow for migration, there is much uncertainty regarding the extent of international migration into South Africa. Additionally, death distribution methods also depend on minimal age misreporting and that completeness is constant by age. While research has found misreporting of older-age mortality data [43] and differential completeness of death registration by age, [33,42] limited expertise and information exist to adjust for the same. The Preston and Hill method [39] applied by StatsSA is also affected by violation of the assumption of a closed and demographically stable population, as well as potential limitations in the use of population growth rates in the method, not its age distribution. [40,41] A further problem with this attempt to estimate the completeness on an annual basis is that the calculation is done relative to population estimates, effectively comparing the deaths against assumed modelled estimates of mortality incorporated in the population estimation. Until reliable in/out migration data are available, the misreporting of age eliminated, and completeness of death reporting constant by age, data inputs for indirect completeness estimation will remain deficient.

One solution to this problem lies in the derivation of direct estimates of completeness, based on dual-record or capture-recapture methods. This will require rigorous mortality data collection in censuses, surveys, and demographic surveillance sites, and appropriate data-linkage with civil registration data. Other likely additional data sources for such analysis include mortality audits (such as the Confidential Enquiry into Maternal Deaths or Child Healthcare Problem Identification Programme) and national disease registers (such as the Electronic Tuberculosis Register).

Our findings in terms of differences between observed and predicted broad-cause patterns were expected when assessing *epidemiological consistency*, since the model is largely based on data from countries which do not reflect high HIV/AIDS mortality. [23] When we attempted to compensate for this model characteristic by excluding the estimated HIV/AIDS deaths, better results for epidemiological consistency were produced, however, considerable incongruence remained, particularly in terms of higher than predicted mortality fractions for infectious diseases in males. It is possible that mortality from infectious diseases other than HIV/AIDS occur at higher frequency in South Africa as compared to the historical data, at similar levels of all-cause mortality. This can be confirmed from studies on content validity of death registration data, for deaths coded to infectious diseases.

Confirmation of coding in a sample of male deaths from infectious diseases could serve as a plausible epidemiological explanation for the differences in proportionate mortality, and could permit a satisfactory rating on this criterion. However, another explanation could be that the HIV/AIDS deaths were underestimated in the preliminary results of the South African National Burden of Disease study in which misclassified HIV/AIDS diagnoses have been estimated based on the excess mortality observed in the distinct age categories, correlated with the increase in HIV prevalence.

The higher than expected proportionate mortality from injuries for women in South Africa are consistent with reports showing that South African women are subject to high levels of intimate partner violence, femicide and rape-homicide. [56–59] More generally, South Africa is known to rank among countries with the highest injury mortality in the world, [60] and these high levels of injury mortality among women were not apparent in the historical data represented by the model, which may explain the differences between observed and model-predicted proportions. It should be borne in mind that these are differences in proportions, and do not reflect a considerably larger number of events. However, these findings give sufficient indication for further exploration, triangulation, and validation of the data to confirm the occurrence of such deaths. The findings also call for sufficient detailing on the DNF to effectively influence the necessary public health action.

In South Africa the epidemiological transition has been radically interrupted, suspended, and likely reversed by the HIV/AIDS epidemic after the transition was steadily on course with all-cause child and adult mortality declining, [14,54,61] and NCDs progressively taking a larger share [26,27,62] during the 1970s and 1980's until the 1990s. In addition, injuries have been a substantial factor for a long while in the country's epidemiological profile, rating among the highest globally. [60] This situation points to fundamental differences between the epidemiological profiles on which the model was based and that of South Africa, which is characterised by the quadruple burden from HIV/AIDS; other infectious/childhood & maternal conditions; NCDs; and injuries. [27] It is therefore likely that the observed departures from predicted patterns are real and that they do not necessarily reflect unsatisfactory quality broad-cause data, but rather point to epidemiological heterogeneity. [63,64] More information on content validity is required to decide whether the differences between the model predictions and observed data are real or an artefact of data quality, and hence we chose to rate the epidemiological consistency of South African data for 2007 as 'unknown'.

Our study reveals that there is considerable uncertainty in unit record cause-of-death data which limits the usability for local health planning and resource prioritisation, or to estimate mortality in other countries on the continent with sparse or no cause-of-death data. Problems with accuracy of the certification and coding of causes of death are widespread, even in countries with good-quality cause-of-death data. [65–67] In South Africa, researchers identified that low quality of medical certification by doctors, certification by non-medical practitioners, and problems in obtaining causes for injury deaths are key challenges to improving data quality. [68] These are valid immediate challenges, but we also recommend that underlying concerns be addressed, including adequate, continued death certification training of medical students and in-service medical certifiers, along with an emphasis on the usefulness of keeping good-quality medical records. Determined efforts are needed to expand civil registration and health services to remote rural areas with adequate staffing. Changes to the DNF are recommended to include fields to adequately record details for external causes (e.g. drowning, or

motor vehicle accident) and apparent manner (e.g. accident, homicide, or suicide) of injury deaths, and the venue (e.g. railway track, private house/yard, construction area) and district of the injury that led to death.

Additionally we recommend examining the feasibility of routine validation of a sample of DNFs with doctors' notes/hospital records/day hospital or clinic cards at major public health facilities such as tertiary hospitals to evaluate data quality from urban areas, as well as linking DSS and vital registration data via capture-recapture methods to assess completeness of death reporting and reliability of reported causes of death in largely rural areas. We also recommend a large-scale validation of cause-of-death data using DNFs and good-quality hospital records to assess both the accuracy of cause attribution and the effect of health interventions. A national validation study may be daunting given severe shortages of human and financial resources, [69] potential challenges in drawing a representative sample, and limitations identified [49,50] in comparing DNFs and medical records. It may be more achievable to start at health-district level: strengthening skills, knowledge and capacity in one health district to be transferred to the next, and building on pioneering work recently undertaken to develop cause-of-death profiles for the 52 health districts in the country [70].

Some criteria in the evaluation framework have the potential to assist in evaluating others. For instance, high scores of content validity would help identify whether differences between observed and modelled cause-specific mortality proportions have true epidemiological explanations, or result from poor data quality, thereby assisting in rating epidemiological consistency. Content validity scores would also help clarify possible departures in temporal consistency and age or sex patterns by cause. For example, mortality from stroke is observed to rise from ages as early as age 30. If these deaths are proved to be accurately attributed to cause, such early stroke mortality can be proved, and serve as information for clinical and public health interventions to reduce this burden.

### Study Limitations

The study has limitations. The intended evaluation regarding *epidemiological consistency* could not be assessed conclusively as the model does not have the discriminatory power to enable an assessment under this criterion for South Africa, and the rating for this criterion remains unknown. The sub-national data made available for our analysis are classified by province of death occurrence. While such data enables an understanding of data quality at the point or area of death, it also limits an understanding of mortality patterns at the place of usual residence of the deceased in those cases where the place of death occurrence differs from that of usual residence. With no national validation studies available, we based our unsatisfactory rating for *content validity* on results from mainly small-scale studies, and studies with a focus on HIV/AIDS. It may therefore be argued that this criterion cannot be evaluated appropriately, and should be rated 'unknown' rather than 'unsatisfactory'.

### Conclusion

Improvements over time, and six satisfactory ratings out nine quality assessment criteria, are a tribute to the focused efforts and investments by strategic players in the death registration system over the past two decades. Performance of criteria assessed in earlier evaluations [1,3] have improved such that it would shift South Africa's mortality statistics from "low" to "medium-high quality", using the evaluation categories of Mahapatra *et al.* [3] However, the

criteria rated 'unsatisfactory' have substantial potential to influence the reliability and validity of mortality data, and point to cause-of-death attribution, in particular, as a vulnerable component in producing good-quality local mortality data.

In previous studies, [1,3,61] the low availability of mortality data, weak quality of all-cause mortality data, and the absence of single-cause data from sub-Saharan African countries often stood out. This analysis symbolizes optimism for a region marred by data absences and deficiencies by showing considerable progress from a system paralyzed by disintegration, disparity and delay to one with potential to offer integrated, inclusive and timely sub-national all-cause mortality statistics that could be adjusted and used for demographic and health analysis. Additionally, cause-of-death data, certified and coded according to international standards in a recent ICD revision, are available by regional and socio-demographic disaggregation. However, a considerable confidence gap remains for single-cause data to be used for health planning without skilfully estimating adjustments for biases. This assessment, and other research, [26,71–74] have identified priority actions that the South African civil registration, health, and education and training authorities need to take to significantly improve confidence in its mortality statistics, in particular, to improve the accuracy of national single-cause data. Improving the accuracy of single-cause data will be a critical contribution to the epidemiologic and population health evidence base in Africa.

### Supporting Information

**Figure S1 A.** Chapter R codes (R00–R99) in each age group as a percentage of total deaths in that age group, South Africa, 1997–2007. Source: StatsSA vital registration data. [16] **B.** Trends in the number of registered deaths by province of death, South Africa, 1997–2007. Source: StatsSA vital registration data. [16] **C.** Trends in the number of deaths assigned an R code (R00–R99) by province of death, South Africa, 1997–2007. Source: StatsSA vital registration data [16]. (XLSX)

**Figure S2 A.** Unadjusted age-specific death rates by year of death for three major cause groups, South Africa, 1997–2007. Source: Mortality data from StatsSA vital registration; [16] population data from ASSA2008. [14] **B.** Unadjusted age- and sex-specific rates for cerebrovascular deaths, by province and nationally, 1997–2007: Log scale. Source: Mortality data from StatsSA vital registration; [16] population data from ASSA2008 [14]. (XLSX)

**Text S1** Sex-specific and age-dependent causes/cause groups. (DOCX)

### Acknowledgments

The authors thank Statistics South Africa for making available the civil registration mortality data used in this analysis, and Dermot Petersen, at the South African Medical Research Council, for technical assistance with finalizing the figures for publication.

### Author Contributions

Conceived and designed the experiments: JJ CR TV AL. Performed the experiments: JJ CR TV. Analyzed the data: JJ DB. Contributed reagents/materials/analysis tools: JJ CR DB TV AL. Wrote the paper: JJ CR DB TV AL. Wrote the first draft of the manuscript: JJ. Extracted the data and prepared the table and graphs: JJ. Provided input and critically reviewed all drafts: CR DB TV AL.

## References

- Mathers CD, Ma Fat D, Inoue M, Rao C, Lopez AD (2005) Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull World Health Organ* 83: 171–177C.
- United Nations Population Division (2006) World Mortality Report 2005. New York: United Nations.
- Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, et al. (2007) Civil registration systems and vital statistics: successes and missed opportunities. *Lancet* 370: 1653–1663.
- Rao C, Lopez AD, Yang G, Begg S, Ma J (2005) Evaluating national cause-of-death statistics: principles and application to the case of China. *Bull World Health Organ* 83: 618–625.
- Mahapatra P, Rao CPV (2001) Cause of death reporting systems in India: a performance analysis. *National Medical Journal of India* 14: 154–162.
- França E, de Abreu DX, Rao C, Lopez AD (2008) Evaluation of cause-of-death statistics for Brazil, 2002–2004. *Int J Epidemiol* 37: 891–901.
- Bah S (1999) The making and unmaking of a national but stratified vital statistics system in the Republic of South Africa and the new making of a more comprehensive vital statistics system. *South Afr J Demogr* 7: 45–50.
- Bah S (2009) Multiple forces working in unison: the case of rapid improvement of vital statistics in South Africa post-1996. *World Health & Population* 11: 50–59.
- Bourne DE (1995) Sources of South African mortality data, 1910–1992. *S Afr Med J* 85: 1020–1022.
- Bradshaw D, Kielkowski D, Sitas F (1998) New birth and death registration forms - A foundation for the future, a challenge for health workers? *S Afr Med J* 88: 971–974.
- Mathers C, Lopez A, Murray C (2006) The burden of disease and mortality by condition: data, methods, and results for 2001. In: Lopez A, Mathers C, Ezzati M, Jamison D, Murray CJ, editors. *Global Burden of Disease and Risk Factors*. New York, NY: Oxford University Press. 100–106, Table 103A.103.
- Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, et al. (2012) Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 380: 2071–2094.
- Republic of South Africa (1999) Statistics Act, 1999 (No. 6 of 1999). In: *Government Gazette No. 19957*. Cape Town: Government Printer.
- Actuarial Society of South Africa: AIDS Committee (c2011) *ASSA2008 AIDS and Demographic Models*. Cape Town, South Africa: Actuarial Society of South Africa.
- Ruzicka LT, Lopez AD (1990) The use of cause-of-death statistics for health situation assessment: national and international experiences. *World Health Stat Q* 43: 249–258.
- Statistics South Africa (1997–2007) Causes of death in South Africa, 1997 to 2007, extracted from death notification forms. Provided by Statistics South Africa, Pretoria, South Africa.
- Statistics South Africa (2010) Mortality and causes of death in South Africa, 2008: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa.
- South Africa Government Services. Application to give notice of death. [cited 25 Sep 2011]; Available from: [http://www.services.gov.za/services/content/Home/ServicesForPeople/TraveloutsideSA/Applicationtogivenoticeofdeath/en\\_ZA](http://www.services.gov.za/services/content/Home/ServicesForPeople/TraveloutsideSA/Applicationtogivenoticeofdeath/en_ZA).
- World Health Organization (2004) *ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth revision*. Geneva: World Health Organization.
- National Center for Health Statistics (2001) *Instruction manual, Part 2c: ICD-10 ACME decision tables for classifying underlying causes of death*. Hyattsville, MD: US Department of Health and Human Services, Public Health Services.
- Omran AR (1971) The Epidemiologic Transition: A Theory of the Epidemiology of Population Change. *The Milbank Memorial Fund quarterly* 49: 509–538.
- Preston S (1976) *Mortality patterns in national populations: with special reference to recorded causes of death*. New York: Academic Press.
- Salomon JA, Murray CJL (2002) The epidemiologic transition revisited: Compositional models for causes of death by age and sex. *Population and Development Review* 28: 205–228.
- Murray CJL, Lopez AD, editors (1996) *The global burden of disease: A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020*. Cambridge, MA: Harvard University Press.
- Statistics South Africa (2009) *Gross Domestic Product. Annual estimates 1993–2008, Regional estimates 2000–2008*. Statistical release P0441. Pretoria: Statistics South Africa. P0441 P0441.
- Bradshaw D, Pillay-Van Wyk V, Laubscher R, Nojilana B, Groenewald P, et al. (2010) Cause of death statistics for South Africa: challenges and possibilities for improvement. Cape Town: South African Medical Research Council.
- Bradshaw D, Groenewald P, Laubscher R, Nannan N, Nojilana B, et al. (2003) Initial burden of disease estimates for South Africa, 2000. *S Afr Med J* 93: 682–688.
- Khalfani AK, Zuberi T, Bah S, Lehohla P (2005) Population statistics. In: Zuberi T, Sibanda A, Udjo EO, editors. *The Demography of South Africa*. Armonk, New York and London, England: M.E. Sharpe.
- Dorrington RE (1989) African mortality rates - an initial estimate. *Transactions of the Actuarial Society of South Africa* 7: 169–202.
- Timaeus IA (1993) Adult mortality. In: Foote KA, Hill KH, Martin LG, editors. *Demographic change in sub-Saharan Africa*. Washington, D.C.: National Academy Press. 218–255.
- Dorrington RE, Bourne D, Bradshaw D, Laubscher R, Timaeus IM (2001) The Impact of HIV/AIDS on Adult Mortality in South Africa. Tygerberg, Cape Town: South African Medical Research Council.
- Dorrington RE, Moultrie TA, Timaeus IM (2004) Estimation of mortality using the South African Census 2001 data. Cape Town: Centre for Actuarial Research, University of Cape Town.
- Dorrington RE, Bradshaw D (2011) Maternal mortality in South Africa: lessons from a case study in the use of deaths reported by households in censuses and surveys. *Journal of Population Research* 28: 49–73.
- Bennett NG, Horiuchi S (1984) Mortality estimation from registered deaths in less developed countries. *Demography* 21: 217–233.
- Actuarial Society of South Africa: AIDS Committee (c2011) *The ASSA600 AIDS model*. Cape Town, South Africa: Actuarial Society of South Africa.
- Statistics South Africa (2003) *Census 2001: How the count was done*. Pretoria: Statistics South Africa.
- Hill KH (1987) Estimating census and death registration completeness. *Asian and Pacific Population Forum* 1: 8–13.
- Statistics South Africa (2008) *Community Survey, 2007. Basic Results: Municipalities*. Statistical release P0301.1. Pretoria: Statistics South Africa.
- Preston SH, Hill K (1980) Estimating completeness of death registration. *Population Studies* 34: 349–366.
- Statistics South Africa (2009) Mortality and causes of death in South Africa, 2007: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa.
- Statistics South Africa (2006) Mortality and causes of death in South Africa, 2003 and 2004: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa.
- Darikwa TB, Dorrington RE (2011) The level and trends of child mortality in South Africa, 1996–2006. *J Afr Population Stud* 25: 158–172.
- Machemedze T (2009) *Old age mortality in South Africa*. MPhil thesis. [Partial fulfillment of course work degree]. Cape Town: University of Cape Town. 74 p.
- Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D (2005) Identifying deaths from AIDS in South Africa. *AIDS* 19: 193–201.
- Birnbaum JK, Murray CJL, Lozano R (2011) Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organ* 89: 278–285.
- Westwood AT (2000) Childhood deaths due to HIV—the role of the new death certificate. *S Afr Med J* 90: 877.
- Grandin W, Westwood T, Lagerdien K, King MS (2006) Deaths at Red Cross Children's Hospital, Cape Town 1999–2003—a study of death notification forms. *S Afr Med J* 96: 964–968.
- Yudkin PL, Burger EH, Bradshaw D, Groenewald P, Ward AM, et al. (2009) Deaths caused by HIV disease under-reported in South Africa. *AIDS* 23: 1600–1602.
- Nojilana B, Groenewald P, Bradshaw D, Reagon G (2009) Quality of cause of death certification at an academic hospital in Cape Town, South Africa. *S Afr Med J* 99: 648–652.
- Burger EH, Groenewald P, Bradshaw D, Ward AM, Yudkin PL, et al. (2012) Validation study of cause of death statistics in Cape Town, South Africa, found poor agreement. *J Clin Epidemiol* 65: 309–316.
- Statistics South Africa (2001) *Advance release of recorded deaths, 1997–2000*. Pretoria, South Africa: Statistics South Africa.
- Statistics South Africa (1997–2010) *Mortality and causes of death in South Africa: Findings from death notification*. Report P0309.3. Pretoria, South Africa: Statistics South Africa.
- Statistics South Africa (1999) *Advance release of recorded deaths, 1996*. Pretoria, South Africa: Statistics South Africa.
- Rajaratnam JK, Marcus JR, Flaxman AD, Wang H, Levin-Rector A, et al. (2010) 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.
- Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD (2010) What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Med* 7: e1000262. Available: <http://dx.doi.org/10.1371/journal.pmed.1000262>. Accessed 14 Aug 2012.
- Abrahams N, Martin LJ, Jewkes R, Mathews S, Vetten L, et al. (2008) The epidemiology and the pathology of suspected rape homicide in South Africa. *Forensic Sci Int* 178: 132–138.
- Abrahams N, Jewkes R, Martin L, Mathews S, Vetten L, et al. (2009) Mortality of women from intimate partner violence in South Africa: a national epidemiological study. *Violence Vict* 24: 546–556.
- Norman R, Schneider M, Bradshaw D, Jewkes R, Abrahams N, et al. (2010) Interpersonal violence: an important risk factor for disease and injury in South Africa. *Popul Health Metr* 8: 32.



59. Mathews S, Abrahams N, Jewkes R, Martin LJ, Lombard C, et al. (2008) Intimate femicide-suicide in South Africa: a cross-sectional study. *Bull World Health Organ* 86: 552–558.
60. Norman R, Matzopoulos R, Groenewald P, Bradshaw D (2007) The high burden of injuries in South Africa. *Bull World Health Organ* 85: 695–702.
61. Rajaratnam JK, Marcus JR, Levin-Rector A, Chalupka AN, Wang H, et al. (2010) Worldwide mortality in men and women aged 15–59 years from 1970 to 2010: a systematic analysis. *Lancet* 375: 1704–1720.
62. Botha JL, Bradshaw D (1985) African vital statistics - A black hole? *S Afr Med J* 67: 977–981.
63. Murray CJL, Chen LC (1994) Dynamics and patterns of mortality change. In: Chen LC, Kleinman A, Ware NC, editors. *Health and social change in international perspective*. Boston: Harvard School of Public Health. 3–23.
64. Frenk J, Bobadilla JL, Sepulveda J, Cervantes ML (1989) Health transition in middle-income countries: New challenges for health care. *Health Policy Plan* 4: 29–39.
65. Middleton D, Anderson R, Billingsly T, Virgil NBM, Wimberly Y, et al. (2011) Death certification: issues and interventions. *Open Journal of Preventive Medicine* 1: 167–170.
66. Johansson LA, Westerling R, Rosenberg HM (2006) Methodology of studies evaluating death certificate accuracy were flawed. *J Clin Epidemiol* 59: 125–131.
67. Aung E, Rao C, Walker S (2010) Teaching cause-of-death certification: lessons from international experience. *Postgrad Med J* 86: 143–152.
68. Pillay-van Wyk V, Bradshaw D, Groenewald P, Laubscher R (2011) Improving the quality of medical certification of cause of death: the time is now! *S Afr Med J* 101: 626.
69. Mayosi BM, Lawn JE, van Niekerk A, Bradshaw D, Abdool Karim SS, et al. (2012) Health in South Africa: changes and challenges since 2009. *Lancet* 380: 2029–2043.
70. Groenewald P, Msemburi W, Neethling I, Day C, Tuoane-Nkhazi M, et al. (2012) Burden of Disease. In: Day C, Barron P, Massyn N, Padarath A, English R, editors. *District Health Barometer 2010/11*. Durban, South Africa: Health Systems Trust. 131–150.
71. United Nations, African Union Commission, African Development Bank (2011) *Africa programme on accelerated improvement of civil registration and vital statistics (APAI-CRVS)*. New York: Economic Commission for Africa, United Nations.
72. Department of Health (South Africa) (2012) *Health Data Advisory and Co-ordination Committee (HDACC) report*. Pretoria: Department of Health.
73. Bradshaw D, Dorrington R, Laubscher R (2012) *Rapid mortality surveillance report 2011*. Cape Town: South African Medical Research Council.
74. Joubert J, Rao C, Bradshaw D, Dorrington RE, Vos T, et al. (2012) Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa. Available: <http://www.globalhealthaction.net/index.php/gha/article/view/19263>. Accessed 28 Dec 2012.

The following figures and 'Text S1' have been inserted from the supplementary material of Paper 2 on the PLOS ONE website.

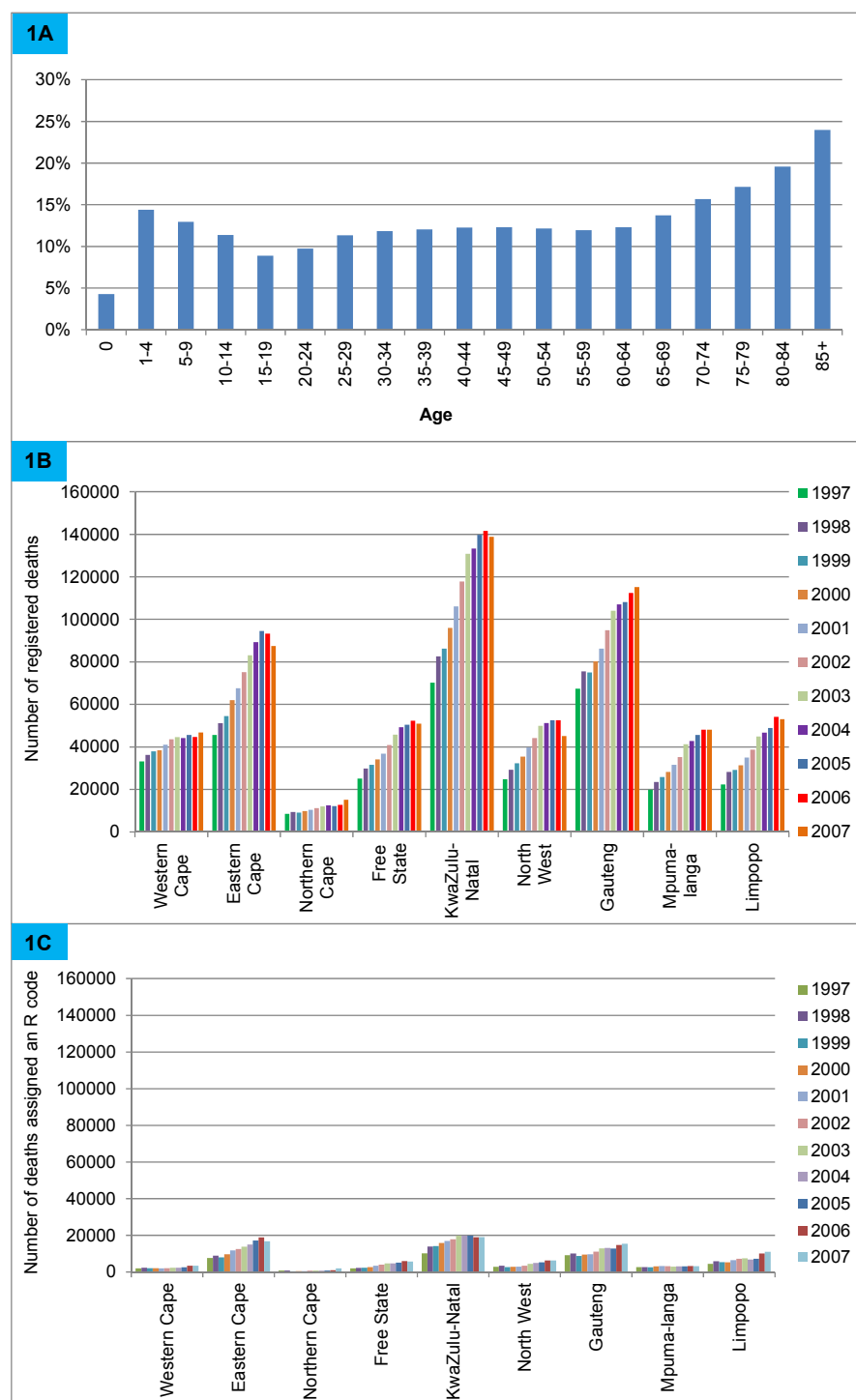


Figure S1A: Chapter R codes (R00–R99) in each age group as a percentage of total deaths in that age group, South Africa, 1997–2007. B. Trends in the number of registered deaths by province of death, South Africa, 1997–2007. C. Trends in the number of deaths assigned an R code (R00–R99) by province of death, South Africa, 1997–2007. Source: StatsSA vital registration data.<sup>154</sup>

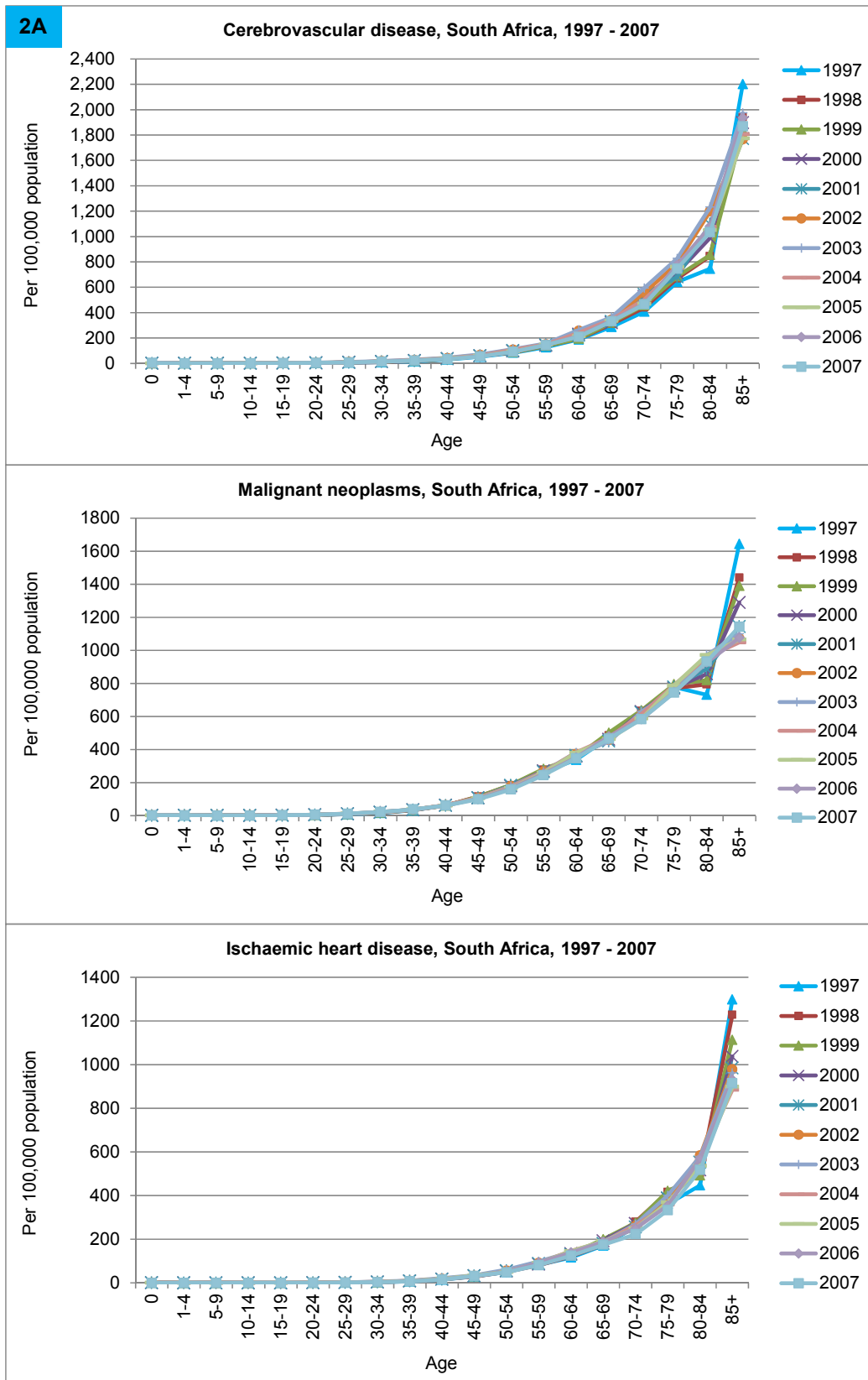


Figure S2A: Unadjusted age-specific death rates by year of death for three major cause groups, South Africa, 1997–2007. Source: Mortality data from StatsSA vital registration;<sup>154</sup> population data from ASSA2008.<sup>155</sup>

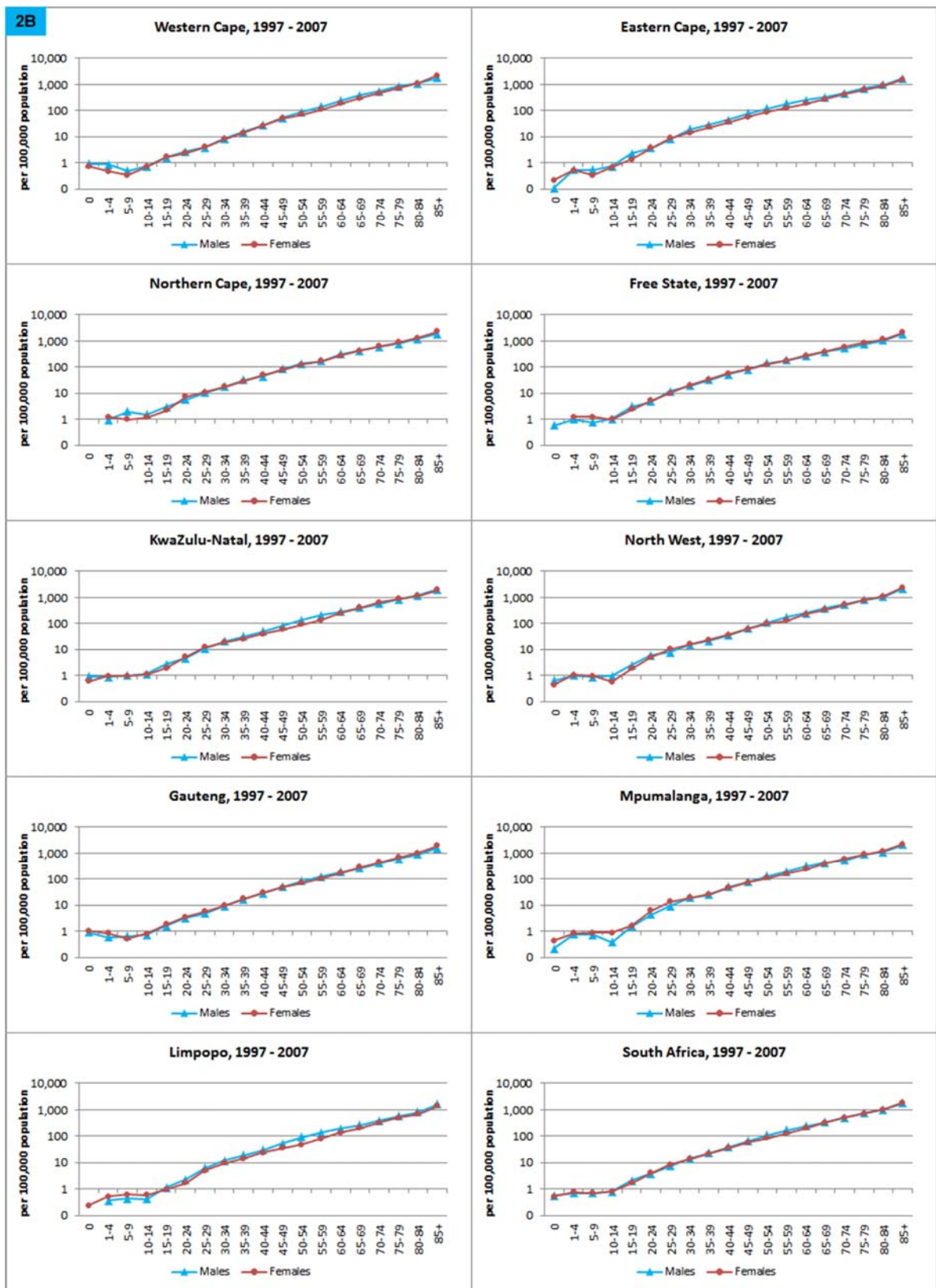


Figure S2B: Unadjusted age- and sex-specific rates for cerebrovascular deaths, by province and nationally, 1997–2007: Log scale. Source: Mortality data from StatsSA vital registration;<sup>154</sup> population data from ASSA2008.<sup>155</sup>

**Text S1:** Sex-specific and age-dependent causes/cause groups.

The aggregate dataset was examined for departures from 10 sex-specific causes: maternal haemorrhage, maternal sepsis, hypertensive disorders of pregnancy, obstructed labour, abortion, other maternal conditions, cervix uteri cancer, corpus uteri cancer, ovary cancer, and prostate cancer.

Age patterns were examined for plausibility and consistency in 27 typically age-dependent causes/cause groups: maternal conditions; low birth weight; birth asphyxia and birth trauma; other peri-natal conditions; 16 cancers (mouth and oro-pharynx; oesophagus; stomach; colon and rectum; liver; pancreas; trachea, bronchus and lung; melanoma and other skin; breast; cervix uteri; corpus uteri; ovary; prostate; bladder; lymphomas and multiple myeloma; leukaemia); rheumatic heart disease; hypertensive disease; ischaemic heart disease; stroke; inflammatory heart disease; benign prostatic hypertrophy; and self-inflicted injuries.

### **3.6 Concluding remarks**

#### **3.6.1 Evaluation studies, tools, frameworks, and criteria**

Although the need for countries to assess and document the quality of their mortality statistics was expressed 25 years ago in the WHO's World Health Statistics Quarterly,<sup>150</sup> it was not until 2001 that a country-specific evaluation of mortality data was conducted in India.<sup>152</sup> Four years later, an evaluation was conducted for China,<sup>10</sup> followed by an evaluation for Brazil<sup>126</sup> three years later. Comparative evaluations of country data submitted to international agencies have been around for longer, one of which had been conducted close to 65 years ago,<sup>15</sup> although the first multi-country comparative study to assess the availability and quality of *cause-of-death data* was only conducted in 1990.<sup>150</sup> The multi-country evaluation by Mathers *et al* has been critiqued, suggesting that the methods do not produce sufficiently reliable estimates of the differences among countries,<sup>156</sup> and that the evaluation neglected to consider the effect of the skills and knowledge of cause-of-death coders on data quality.<sup>157</sup> A constructive authors' response to these critiques provided useful insight into the limitations in country data that influence the analysis possibilities in multi-country evaluations.<sup>158</sup>

For this thesis, the country-specific evaluations, the multi-country comparative studies, and the tools and frameworks published by institutions such as the Health Metrics Network and The University of Queensland HIS Hub, were found useful as they enable the review of a wide

spectrum of evaluation components, criteria, and data items. It was possible to select a framework and criteria that best suits the primary aim of the thesis – to evaluate the quality of South Africa’s CRVS mortality data, thereby facilitating a response to research question 2: *What criteria can be used for evaluating national death and cause-of-death statistics from civil registration in South Africa?*

### **3.6.2 Quality of national mortality statistics from civil registration in South Africa**

Six satisfactory ratings out nine quality assessment criteria, and improvements from “low” to “medium-high quality” with regard to certain criteria from previous comparative evaluations, point to the opportunity to rate the quality of the assessed data “satisfactory”, and so provide an overall or general answer to research question 3: *What is the quality of national mortality statistics from civil registration for the period 1997 – 2007?*

However, it is not ideal, likely neither correct, to provide a general answer, as it is important to examine each criterion separately, to be able to seek and identify the relevant cause(s) of bias, error, or misinformation. In this way, it is possible to discern that the criteria rated ‘unsatisfactory’, i.e. *content validity* and *use of ill-defined and non-specific codes*, point to considerable shortcomings with local cause-of-death data. It was also possible to identify two criteria, commonly having a substantial effect on the reliability and representativeness of mortality data (*completeness of death registration*), and profoundly affecting the validity of cause-of-death data (*content validity*), which could benefit from further research. These criteria and their respective data elements, i.e. death registration data and cause-of-death data, are incorporated in further research presented in Chapters 4 and 5.

### **3.6.3 Suggested comprehensive framework for South Africa and countries with comparable statistical development**

During the assessment and analysis of the CRVS mortality data for South Africa, it became clear that it would be useful, at least for South Africa, to (a) consolidate relevant data attributes and criteria from different authors as found in the literature; (b) add criteria not found in the literature, though perceived as suitable and useful for quantifying known and newly-identified shortcomings, in the South African data; and (c) combine these in one comprehensive evaluation framework, as done in Table 3.2 below.

In addition, the student became aware that not all criteria in the various tools and instruments, though useful, could be taken to be understandable or comprehensible by all readers. To help

resolve this and make the framework more user-friendly, simple assessment questions, though phrased to effectively capture the essence to be measured in the criterion, have been added as a third column to Table 3.2.

Furthermore, assessing a criterion without an informed threshold value, or desired response, leaves the researcher without the ability to appraise whether performance of a criterion was satisfactory or not. Drawing on suggestions in previous evaluations, where the particular criteria were used,<sup>8-10,17,126,150,152</sup> the assessment questions and suggested thresholds/desired responses were formulated and added as part of the framework.

The framework is included for its possible utility as a road map for countries in the region that are initiating, re-designing, or improving their CRVS system. From a recent United Nations report, it is clear that there are substantial gaps in the development and performance of CRVS systems in the Southern African Development Community (SADC) countries, and that numerous remedial actions are needed.<sup>4</sup> For example, only 4 of 14 countries have been in a position by 2009 to provide an estimate of the completeness of death registration in their country, despite all these countries having centralized civil registration systems in which the legal responsibility for civil registration is assumed by specific ministries or the Office of the President. Only Mauritius and South Africa were able to provide annual vital statistics from their civil registration systems, while the remaining countries relied on censuses and sample surveys to generate basic statistics.<sup>4</sup>

It would be ideal to evaluate all suggested criteria to comprehensively assess the quality of CRVS mortality data. However, a country could decide which attributes or criteria are most appropriate or most important to measure at a specific time, depending on the country's general level of statistical development, the level of development of its CRVS system, specifically, or its particular health policy or information needs. As the CRVS system improves, the suggested thresholds could be adapted or refined to align with such improvements, or to raise the bar for further improvements.

For some of the SADC countries where limited technical capacity is reported,<sup>4</sup> and others in a similar position, however, it may be more useful to start with the *Rapid assessment of national civil registration and vital statistics systems tool*,<sup>18</sup> or the "Simple 10-step process" as described in the *Mortality statistics assessment guide and toolkit* developed by the HIS Hub and WHO.<sup>159</sup>

Table 3.2: Comprehensive evaluation framework for assessing the quality of national mortality data from civil registration

General attributes	Criteria	Assessment questions	Suggested threshold or desired response
Generalizability	1. Coverage	What % of the population is covered by the civil registration (CR) system?	100% of national or sample population
	2. Covered completeness of registration	Within the covered population, what % of deaths is registered into the CR system?	≥ 90%
	3. Total completeness of registration	Within the total population, what proportion of deaths is reported into the CR system?	Variable, depending on population covered by system
	4. Completeness of key descriptive information	Within the registered population, what % of deaths have key socio-demographic variables (age, sex, ID number, date of birth, date of death, place of death, place of usual residence) available?	≥ 99%: sex, age, date of death ≥ 95%: ID number, date of birth ≥ 90%: other
	5. Completeness of medically-certified cause of death	In what proportion of registered deaths are causes of death assigned by a medical practitioner?	≥ 90%
Reliability	6. Epidemiological consistency	To what extent are cause-of-death patterns consistent with the total level of mortality?	< 2 standard deviations from the mean of predicted levels
	7. Temporal consistency	To what extent is cause-specific mortality consistent over time?	< 2-3% annual death rate fluctuation in leading causes, unless explained by local epidemiological phenomena
Validity	8. Content validity	To what extent are certification and coding of the underlying cause accurate?	If no sensitivity/ specificity/ kappa scores from validation studies, consider criteria 9&10 and extent of errors on DNF
	9. Use of ill-defined & non-specific codes	What proportion of registered deaths is assigned an ill-defined/non-specific cause?	< 10% of deaths
	10. Use of improbable classifications	What proportion of deaths is assigned an improbable age- or sex-cause?	< 1% of deaths
Policy relevance	11. Timeliness	What is the time gap between the end of the reference period (year of death) and the time of publication of final tabulations?	≤ 2 years
	12. Regularity	How regularly are reports/tabulations published?	≥ 1 time per year
	13. Routine sub-national availability	Are death and cause-of-death data available at sub-national jurisdictions for analysis?	Yes, at least at provincial/state level, preferably district level
	14. Routine residence, age- and sex-specific classification	Are cause-of-death data available by usual residence, sex & age classifications useful for epidemiological, public health and demographic analysis?	Yes, by usual residence, sex, and age groups as advised by WHO and United Nations tabulation recommendations
Comparability across time, space, and settings	15. Timely registration	What proportions of deaths is registered and processed timely?	≥ 95% of deaths registered within 1 year of death
	16. Standard classification system	Have standard classifications been used in certifying, coding, and disseminating all causes of deaths?	Yes, preferably ICD, preferably revision 9 or 10; preferably at 3- or 4-digit code level
	17. Use of country-specific codes	What proportion of registered deaths is assigned a country-specific code?	< 1% of deaths
Accessibility	18. Availability	Are mortality data publically available for analysis by responsible data users?	Yes, all variables with non-identifiable & non-confidential data
	19. Transparency	Are adequate metadata available?	Yes, with codebook and background descriptions
	20. Suitability	Are the data available in formats suitable for a) local and b) international absorption and analyses?	Yes, a) in locally-preferred formats; and b) electronically in unit-record format
	21. Responsiveness	Is the user service a) easily contactable and b) responsive?	Yes, both easily contactable and responsive with feedback and/or action upon requests
	22. Affordability	Are the data affordable, if not for free?	Preferably available for free, or at minimum local cost



Notes to Table 3.2: The attributes and criteria in white text were used in previous evaluations of country mortality data.<sup>8-10,17,126,150,152</sup> The attributes and criteria in black were taken or adapted from Mahapatra *et al.*<sup>9</sup> The criteria in purple were added by the student. Drawing on suggestions in previous evaluations where the particular criteria were used,<sup>8-10,17,126,150,152</sup> the assessment questions and suggested thresholds/desired responses were formulated by the student and added as an integral part of the framework.

Selected assessment concepts have been arranged in table format in the work of Mathers *et al.*, 2005<sup>8</sup> (page 175, Table 4) and Mahapatra *et al.*, 2007<sup>9</sup> (page 1654, Table 1). The student is not aware of another country-specific assessment framework published in this inclusive and comprehensive format, accommodating general attributes, criteria, assessment questions, and suggested thresholds/desired responses in one graphic presentation.

## 4 Chapter 4: Assessing the feasibility, utility, and quality of linking mortality data from the civil registration system and a health and demographic surveillance site

---

### 4.1 Introduction

Chapter 3 describes a well-established civil registration and vital statistics (CRVS) system in South Africa, offering routine all-cause and cause-specific mortality data with national coverage. The system captures and codes 3-digit ICD-10 cause-of-death data, certified and coded according to international standards, and available by age, sex and sub-national jurisdiction. However, two criteria that have a profound effect on the validity and representativeness of mortality data—*the use of ill-defined and non-specific codes* and *content validity*—were rated as unsatisfactory in the evaluation of national CRVS mortality data in Chapter 3. The former criterion was rated poorly over the entire period of investigation (1997 to 2007),<sup>131</sup> and more recent analysis of cause-of-death data up to 2011, demonstrates continued unsatisfactory performance in relation to this criterion.<sup>50</sup> For *content validity*, ongoing problems were reported with the quality of cause-of-death data for a number of urban studies with less than 750 decedents each,<sup>121,123,160</sup> as well as national studies, examining cause-of-death data for hundreds of thousands of decedents.<sup>93,161-163</sup> Yet, accurate information concerning causes of death is essential to inform health service delivery and research prioritization.

Official agencies responsible for compiling mortality statistics may monitor and correct evident errors on death certificates with traditional editing techniques. Such edits are often based on the compatibility between the reported diagnosis, age and sex.<sup>164</sup> However, the majority of errors on death certificates are non-obvious,<sup>165</sup> and investigation that goes beyond traditional edits, often beyond the death certificate, are needed to identify these errors.<sup>164</sup>

In addition, for all-cause mortality, although compliance with the criterion of *completeness* of death registration was rated as satisfactory in Chapter 3, there are limitations to the indirect methods currently used to estimate completeness.<sup>166</sup> Death distribution methods, for example, depend on assumptions such as minimal misreporting of age that completeness is constant by age, and the study population is closed to migration.<sup>166-169</sup> For South Africa, Machedze found age misreporting in older-age deaths,<sup>170</sup> and Dorrington and Bradshaw,<sup>171</sup> as well as

Darikwa and Dorrington,<sup>172</sup> found differential completeness of death registration by age. There is also considerable uncertainty regarding the extent of international migration into South Africa<sup>131</sup> (see Section 3.4). These findings violate the mentioned assumptions, and highlight the shortcomings of using these indirect methods.

Record linkage, a direct method using empirical data, has been identified as a useful additional tool to assess and verify the quality of both cause-specific and all-cause mortality.<sup>173-175</sup> This methodology may circumvent some of the limiting factors in studies that seek to evaluate cause-specific and all-cause mortality, and has increasingly been used to provide empirical evidence about the quality of mortality data, particularly so in developed economies.<sup>174,176-187</sup> Section 4.2 gives an overview of this methodological approach, and Section 4.3 refers to studies in different countries in which it was possible to link CRVS data with mortality data from other sources.

After ascertaining whether, which, and how existing South African data collections identified in the data-source review in Chapter 2 can be useful in assessing the quality of local CRVS mortality data, a collaboration and sub-study had been proposed to link local CRVS data with mortality data from the Agincourt HDSS (Section 4.4). Research questions 4 and 5 below guided the development of the proposed sub-study, as well as the write-up and analysis of the first research paper from the study:

Research question 4: *Would a linkage study between South African CRVS and the Agincourt HDSS mortality data collections be possible, and, if so, would such linkage be of good quality and useful?*

Research question 5: *What can the mortality data collected in Agincourt HDSS indicate about the completeness of death-registration and factors associated with registration?*

Section 4.5 provides the first of two research papers published from the sub-study, answering these questions. The paper provides the background and rationale for the study; describes the data sources and the matching methods; assesses the quality of the matching exercise; seeks insight into the characteristics of those HDSS records for which a matching record could be found in the CRVS system; and discusses the study results. This section will have a strong focus on the methodology followed and the utility and quality of the linkage, while Chapter 5,

in answering research questions 6 and 7, will focus on comparing cause-of-death information between the data sources.

## 4.2 Record linkage

### 4.2.1 Origin and definition of record linkage

The original ideas underpinning modern record linkage are attributed to geneticist Howard Newcombe.<sup>188-190</sup> Newcombe's ideas, based on odds ratios of frequencies and decision rules for allocating matches and non-matches, have been implemented in software that has been used in many epidemiological studies.<sup>191,192</sup> The formal mathematical foundations of record linkage have been extended and formalized by Fellegi and Sunter, who developed a mathematical model to provide a theoretical framework for computer recognition of records in two files that represent the same entity (person/object/event/business).<sup>193</sup> Fellegi and Sunter presented an optimal decision procedure whereby cut-off threshold weights can be computed by making a decision about acceptable probabilities for false matches and false non-matches. Two threshold weights are established, and all composite weights higher than the high threshold weight are considered matched. Weights below the low threshold are considered non-matches. Weights between the two thresholds, which suggest possible matches, are cases for clerical or manual review.<sup>194</sup>

Modifications and extensions of this work have been suggested and implemented by scientists such as Howe and Lindsay, who suggested extensions of the Newcombe methodology to deal with the problem of partial agreement of identifying data items;<sup>176</sup> Jaro, who developed a linear sum assignment approach to matching, extending the probability model;<sup>194</sup> Winkler, who further enhanced the decision rules for situations in which compared data items have errors;<sup>195</sup> and Zhu et al., who implemented a value-based weight scaling modification into the Fellegi-Sunter maximum-likelihood linkage algorithm.<sup>196</sup>

Herzog *et al* offered a simple definition of record linkage in 2010, i.e. “the bringing together of information from two records that are believed to relate to the same entity”.<sup>197</sup> This corresponds well with the definition of Newcombe et al. who, 50 years before, referred to it as “the bringing together of two or more separately recorded pieces of information concerning a particular individual or family”.<sup>188</sup> The concept of ‘identifying information’ is added in Howe’s

definition, defining record linkage as “the process of comparing two or more records which contain identifying information to determine whether those records refer to the same entity”.<sup>191</sup>

#### 4.2.2 Record linkage approaches

Two main and commonly-used computerized record linkage approaches are deterministic and probabilistic linkage.<sup>184,198-200</sup> Deterministic linkage techniques, also called “exact”, “all-or-none” and “hard” matching,<sup>199,200</sup> were commonly used during the early phases of record linkage. These require an exact match on a single identifier (e.g. a unique personal identity number), or a group of identifiers or combinations of data fields (e.g. a name, surname, date of birth, and date of death), that uniquely identify a given person or other entity. As this approach requires an exact match of the variables used, all identifiers should agree for a link to be made. Deterministic linking is often used when unique identifiers, such as a driver’s licence or social security number, are available. However, it is possible to miss a proportion of true matches due to typographical errors in the identifier(s); missing values; reporting errors (e.g. in date of birth); variations in the format or coding of the data across datasets (e.g. dates, locations, causes of death, abbreviations); or changes to the identifier over time (e.g. maiden vs married name, changes of place names).<sup>184,198,200,201</sup>

In probabilistic linkage, a pair of records from two datasets are classified as belonging to the same individual, or not, based on the statistical probability that common identifiers in the datasets belong to the same entity, using weights or scores to gauge the agreement and disagreement between the records.<sup>184,193-195,202,203</sup> The statistical probability, and not the evidence of an exact match, as needed in deterministic approaches, is then used to decide whether a pair of records refers to the same entity. Probabilistic linkage is therefore more suitable in exercises where unique identifiers are not available, or where there is evidence of variation in reporting, coding or transcription of identifiers.

A combination of deterministic and probabilistic approaches can also be used, as has successfully been done in the work of, for example, Bohensky *et al*<sup>204</sup> and Méray *et al*.<sup>205</sup>

#### 4.2.3 Identifiers used in record linkage

A range of identifiers can be used in record linkage, provided they are common to the different data sources that are used for linkage. As linkage exercises using less informative identifiers

could cause linkage errors,<sup>206</sup> it is important to identify the optimal identifiers from the pool of potential identifiers for each study. Optimal identifiers have been identified, for example, in the work of Van den Brandt *et al* (date of birth, first four characters of family name, and gender);<sup>207</sup> Quantin *et al* (date of birth, first name, and last name);<sup>206</sup> Li *et al* (date of birth, sex, and last name);<sup>184</sup> and that of Méray *et al* (mother's date of birth, and ZIP code).<sup>205</sup>

In the case of linking the records of persons/individuals, generally-used identifiers include:

- personal identifiers, such as name(s), surname, or maiden name;
- geographic information, such as the individual's residential address, or village, city, county, district, or postal/zip code of residence;
- demographic information, such as gender;
- dates of vital events, including day, month and year of birth; day, month and year of death; and
- where available, a unique identity number (e.g. health insurance number, vehicle license number, or a civil or legally-acquired personal identity number).

The particular subject or field of study could require additional or alternative identifiers. In public health, epidemiological, or health services studies, for example, the list of potential identifiers can include name/address of the relevant hospital/medical practice/ pharmacy; duration of hospitalization; date of hospital admission/discharge/transfer; main diagnoses; date of general practitioner/specialist consultation; prescription drug name; previous/current disease; risk factor information; surgical procedure codes; diagnostic material results; and pathological indicators. Such identifiers have been used in a number of studies, including those which aimed to estimate the incidence or prevalence of disease;<sup>208-210</sup> identify new or prevalent users of prescription medications;<sup>211</sup> and assess the relationship between risk factors and health outcomes.<sup>180</sup>

#### **4.2.4 Wide and increased use of data linkage for research in developed countries**

The establishment of data linkage centres, and the expansion of linkage activities in Australia, Canada, the United States, and countries in Europe, Scandinavia, and the United Kingdom, are an indication of the proliferation of data linkage initiatives.<sup>178,204,212,213</sup> In 2010, Bohensky *et al*<sup>204</sup> reported that data linkage in research studies has increased almost 6-fold over the

preceding 20 years. For public health and epidemiological research in particular, Van den Brandt et al.,<sup>207</sup> Newman and Brown,<sup>178</sup> Kelman et al.,<sup>214</sup> Qayad and Zhang,<sup>203</sup> and Ferrante and Boyd<sup>177</sup> reported increased use of data linkage too. In the United Kingdom, for example, almost half of the 105 clinical databases surveyed by Black et al. in 2003, had been used in routine data linkage to other databases.<sup>213</sup> In Australia, Evans et al. reported in 2011 that 68% of the 28 clinical registries investigated, routinely undertook some form of data linkage to assess data quality, and to obtain specified outcome information, such as disease status or death.<sup>215</sup>

The literature reports considerable agreement, where data sources and data items make it possible, that data linkage has become a preferred option to generate more informative and more complete datasets in a time- and cost-efficient manner for public health and epidemiological studies.<sup>174,175</sup> Compare, for example, the commercial return on research infrastructure investment at the Western Australia Data Linkage System that has exceeded 1000%.<sup>175</sup> Linked information can additionally provide an opportunity to create new—either more comprehensive or tailor-made and subject-focussed datasets from existing sources.<sup>174,204</sup> Particularly in developed countries, record linkage has been applied extensively and progressively to generate high-quality datasets for public health and epidemiological studies. Several centres have been noted for ‘highly productive’ research by utilizing linked administrative data.<sup>174,175,216</sup> These include the Western Australia Data Linkage System,<sup>175</sup> the national Population Health Research Network in Australia,<sup>217</sup> Oxford Linkage Study,<sup>218</sup> MigMed2 Database in Sweden,<sup>219</sup> the Centre for Health Services and Policy Research in British Columbia, Canada,<sup>220</sup> and the Rochester Epidemiology Project.<sup>221</sup>

Moreover, data linkage has reduced methodological problems relating to recall, loss-to-follow-up, selection, and thus, response and reporting bias.<sup>175</sup> Linkage has facilitated contributions to medical, epidemiological, and population health knowledge as found in the majority of studies cited in this thesis. For example, by 2008, the Western Australia Data Linkage System has supported over 400 studies, and more than 250 journal publications and 35 research higher degrees,<sup>175</sup> and there have been identifiable advances in population health arising from the research based on linkage.<sup>175,177,191,197</sup>

#### 4.2.5 Increased linkage: reasons and enabling factors

These increases are a result of the development and refinements of the tools and opportunities for data linkage over time, advancing from manual matching exercises, to developing in-house programmes that integrate multiple databases, to using high-speed computers and commercial software tailored for contemporary data matching.<sup>203</sup> In particular, advances in the affordability and capacity of computing equipment were an enabling factor.<sup>175</sup> Linkage activities moreover increased as a result of researchers and research institutions departing from traditional methods that are prohibitively resource-consuming—in terms of financial costs, and human-resource and infrastructural investments—and likely subject to complicating factors referred to before.<sup>176,183,222-225</sup>

In contrast, linked administrative health databases comprised of data already collected, for large populations for different purposes, can be made available for research at a lower cost.<sup>174,175</sup> Jutte *et al* report that, more inclusive than disease-specific registries, and more extensive than routine vital statistics, administrative health databases are increasingly linked to population-wide institutional databases. These include information from a variety of data sources, including outpatient clinical data; nursing records; pharmaceutical data; immunisation records; emergency room records; home care information; maternal care records; disease registries; census data; and population registers,<sup>174</sup> providing a wide platform for research opportunities and the generation of new knowledge.

Holman *et al* report that research-enabling infrastructure has become an important planning focus in a number of countries, including the United Kingdom, the United States, Canada, Australia, and countries in Europe.<sup>175</sup> A number of enabling factors for developing linkage work in Australia have been identified: having a policy of access by any *bona fide* researcher with approved ethics application and a local collaborator; a commitment to the translation of research findings into policy and practice which engender participation by clinicians, consumer advocates, and a broad, countrywide constituency of researchers; and the creation of a sense of community across agency boundaries, as a result of movements of senior public-health personnel between the academic and government sectors.<sup>175</sup>



#### **4.2.6 Data linkage in Africa**

While record linkage has been applied extensively, progressively, and productively in high-income or developed countries,<sup>174,175,216</sup> record linkage in Africa is much less common,<sup>202</sup> whether it is in the context of data available from HDSSs, national surveys, population censuses, disease registries, or civil registration. Obstacles to record linkage in this setting include imprecision in the reporting of dates, variation in the transcription of names, the lack of identification systems that provide for the unique identification of individuals,<sup>202</sup> and likely also a lack of technical expertise and limited awareness of the value of data linkage.

#### **4.3 Studies linking CRVS data with mortality data from other sources**

Linkage studies aiming to match deaths captured in CRVS systems with deaths from other data sources have been conducted successfully in a number of countries. These studies were conducted with diverse purposes, including to assess the feasibility and quality of mortality data linkage among different data sources; verify the number of deaths or completeness of death reporting in a specified period and space; identify the optimal choice of personal identifiers for linkage; identify participant or population characteristics that compromise complete linkage and the sources of such bias; assess the reliability of cause-of-death data on death certificates; consider the implications of study findings for civil registration mortality statistics; as well as combinations of these reasons. Selected studies, with purposes similar to the purpose of the linkage study for this thesis (Section 4.5), are referred to below.

##### **4.3.1 Assessing the feasibility, quality and optimal identifiers of linking CRVS and other mortality data**

Studies to assess the feasibility, success, or quality of data linkage between CRVS and other data sources include that of Roos *et al*, linking data from the Canadian Mortality Data Base, Manitoba Health Services Commission (MHSC), and the Manitoba Cancer Foundation. A combination of family registration number, sex, birth year, and initial was used to match the records, with 96% of MHSC records being matched with Canadian Mortality Data Base records. The study showed that the record linkage rate improved significantly with improvements in data quality over time.<sup>198</sup> Newman and Brown linked patient data from a hospital database with death data from the California Department of Health Services (Deaths of California residents) and from the United States (USA) Social Security Administration

(USA Deaths), using both unique (social security number) and non-unique identifiers (last name, first name, date of birth, middle initial, race, county of residence and sex). Linkage between civil registration and hospital data was conducted successfully, linking 97% of all in-hospital deaths, and 99% of deaths of patients with social security numbers. The authors reported linkage challenges among infants (< 1 year) due to more cases with missing identifiers (e.g. name and social security number) compared to cases involving older persons.<sup>178</sup>

Linking California state-wide hospital discharge records with a database of death certificates for all persons dying in California and California residents who die elsewhere, Zingmond and co-authors successfully linked 95% of in-hospital death records which contained social security numbers to the death records. Subsequent calculation of linkage accuracy showed an accuracy rate of 99%.<sup>182</sup> Herrchen and colleagues successfully linked vital statistics birth data with hospital discharge data in California, USA. Using zip code, hospital code, date of birth, gender, payment source for hospital stay, neonate's ethnicity, and type of delivery, a very high percentage, 98.6%, of the births were matched with patient records in the hospital discharge data.<sup>180</sup>

For residents of the Calgary Health Region of Alberta in Canada, Li *et al* linked death records from (a) the Vital Statistics Registry and the Alberta Health Care Insurance Plan Registry, and then from (b) the Vital Statistic Registry and hospital discharge data. Testing different combinations of identifiers, the investigators found that the combination of surname, sex and date of birth achieved the best linkage rate, i.e. 88% and 93%, respectively, for (a) and (b). The identical recording, or not, of residents' unique lifetime Personal Health Number in the data sources used, enabled an assessment of the correct linkage rate in the study, achieving 97% and 99%, respectively, for (a) and (b). As in the work of Newman and Brown,<sup>178</sup> Li and colleagues found many infant records with missing or inaccurate variables, and subsequently excluded infants in this study.<sup>184</sup>

#### **4.3.2 Assessing numbers of deaths and completeness of death registration, and identifying sources of bias via record linkage**

In Bohol province in the Philippines, Carter *et al* collected death records from civil registration offices, hospitals and health centres, and parish churches for the years 2002 to 2007, and manually matched the records. With a capture-recapture analysis, a total number of deaths was estimated. Completeness of death registration in the civil registration system was estimated at

72%, and it was found that 5 – 10% of total deaths were not reported under any system.<sup>186</sup> In Tonga, Hufanga *et al* obtained and manually matched death records from the Civil Registry and the Ministry of Health, the latter receiving death records from the Health Information System and the Reproductive Health System. Using a capture-recapture analysis, the authors were able to estimate the total numbers of deaths for Tonga for the years 2001 – 2004 and 2005 – 2009, and estimated completeness at 98% and 88% for these periods, respectively.<sup>185</sup> In both these studies, completeness of death recording varied by age, with child deaths less likely than adult deaths to be recorded.<sup>185,186</sup>

In their study of Calgary residents in Canada, linking the Vital Statistics Registry and the Alberta Health Care Insurance Plan Registry, Li *et al* found that the linkage rate varied by age. Over the four-year study period the linkage rate was consistently lower for the age groups 10 – 19 and 20 – 29 years, and considerably lower for the age group 1-9 years.<sup>184</sup> The work of Zingmond and others showed that, when linking hospital discharge data containing a social security number with death certificate data in California, the infants and older adults were more likely than those 1 to 64 years old to have unlinked records; unlinked record rates were two to three times higher for older females compared to older males among those 65 years or older; and Hispanics and non-Hispanic Blacks were consistently more likely than non-Hispanic Whites or Asians to have unlinked records.<sup>182</sup>

In their review, Bohensky *et al* identified a number of participant characteristics, such as age, gender, ethnicity, socio-economic and health status, that were associated with sub-optimal linkage which influences the linkage rate.<sup>204</sup> For example, nine studies in their review indicated that people in lower socio-economic groupings and with lower levels of education were less likely to have matched records, mainly due to lower consent rates for data linkage and less complete data for people of lower socio-economic status.<sup>204</sup> Seven studies showed that people in minority ethnic groups had lower linkage rates. Reasons include a lower likelihood of having a social security number recorded, minority groups being treated at health facilities with less complete data, and lower rates of consent to data linkage.<sup>204</sup> These findings demonstrate that participant characteristics can influence the completeness of data linkage, which may result in systematic biases in reported outcomes such as the relative risk of mortality by age, ethnicity or socio-economic status. This highlights the need for researchers and decision-makers to consider such biases when interpreting and reporting linkage results.<sup>204,226,227</sup>

### 4.3.3 Assessing the reliability of cause-of-death data in mortality data sources through record linkage

Linking data from death certificates of reproductive-age women with foetal death and live birth records from the Vital Statistics Administration in Maryland, USA, and reviewing medical records for evidence of current or recent pregnancy at the time of death, Horon found that only 62% of maternal deaths were recorded on death certificates, implying a substantial under-estimation of maternal deaths when death certificates alone were used.<sup>179</sup> Amin and others linked cases of hepatitis B and hepatitis C notified to the New South Wales Health Department Notifiable Diseases Database (NDD) to deaths in the Australian National Death Index (NDI). Probabilistic linkage was performed for the years 1990 - 2002. For hepatitis B and hepatitis C deaths, cancers (38%) and external causes (28%) were among the most frequently misclassified underlying causes of death reported on death certificates, pointing to considerable under-reporting of hepatitis B and C.<sup>228</sup>

In two research papers,<sup>229,230</sup> Johansson and Westerling report the linkage of death certificate data for 1995 from the Swedish National Cause of Death Register with hospital discharge data from the Swedish Hospital Discharge Register. The authors found that the last main hospital diagnosis and underlying cause of death agreed in less than half (46%) of 69,818 cases, but indicated that cause-of-death information on hospital discharge records and death certificates supplemented each other well, and that the linkage was useful in that conditions absent from one data source may be found in the other.<sup>229,230</sup> In a similar exercise, Klijs *et al* linked death certificate cause-of-death data from the Dutch Cause of Death Registration and hospital discharge data from the National Medical Registration, using a unique personal identification number.<sup>181</sup> They aimed to assess the extent to which registered causes of death represent morbidity at the end of life, and concluded that, except for cancers, registered causes poorly represented diseases recorded during hospitalisation. Similar to the work of Johansson and Westerling,<sup>229,230</sup> Klijs *et al*<sup>181</sup> compared *underlying causes* from death certificates with *main causes* on hospital discharge records. As the definition and purpose of *underlying* and *main cause* on these two types of records do not concur, the low levels of agreement were to be anticipated.

Mühlhauser *et al* compared multiple data sources, including interviews with family physicians and relatives, and autopsy, physician and hospital records, related to deaths from a study cohort of 3,674 patients with Type 1 diabetes who were treated with insulin before age 31 and

monitored for a mean of 10 years. The authors concluded that death certificates were not reliable sources of information for Type I diabetes. Substantial misclassification occurred to conditions such as cardiac and cerebrovascular disease, and diabetes was mentioned on only 71% of death certificates.<sup>231</sup> Findings of unreliable diabetes diagnoses on civil registration death certificates were also reported by Pattaraarchachai *et al*<sup>232</sup> in Thailand, finding about six times fewer deaths from diabetes as a registered underlying cause among Thai hospital deaths, compared to the number of deaths with diabetes that were ascertained via medical record review.<sup>232</sup> For non-hospital deaths in Thailand too, Polprasert *et al* found that diabetes was under-recorded as an underlying cause of death in civil registration data compared to verbal autopsy data.<sup>233</sup>

#### **4.4 Proposing a linkage study with South African mortality data**

The many successful and valuable linkage studies elsewhere as referred to above, using a variety of source materials and data sources, reinforced the idea that a linkage study may be possible in South Africa, despite the obstacles to linkage in an African setting (Section 4.2.6). After assessing the characteristics, variables, and availability of the different mortality data sources in South Africa (Chapter 2), the Agincourt HDSS/Stats SA linkage study was proposed. Mortality data available from the Agincourt HDSS and the CRVS system were deemed suitable for the intended matching study after reviewing the methods of mortality data collection and coding, considering the variables available from each data source, and selecting the candidate variables that were common to both databases.

A collaboration was proposed among the Agincourt HDSS, Burden of Disease Research Unit of the South African Medical Research Council, School of Population Health at the University of Queensland, and Statistics South Africa, to conduct a study that aims to:

- assess the feasibility, quality and utility of record linkage between the mortality data collections from the national CRVS system and Agincourt HDSS (research question 4);
- assess the completeness of death-registration and factors associated with registration (research question 5)
- provide empirical evidence about the quality and utility of rural civil registration cause-of-death data (research question 6, in Chapter 5).

**4.5 Paper 3: “Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa”**



RESEARCH

Open Access

# Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa

Chodziwadziwa W Kabudula<sup>1\*</sup>, Jané D Joubert<sup>2,3</sup>, Maletela Tuwane-Nkhasi<sup>4</sup>, Kathleen Kahn<sup>1,5,6</sup>, Chalapati Rao<sup>3</sup>, Francisc Xavier Gómez-Olivé<sup>1</sup>, Paul Mee<sup>1,5</sup>, Stephen Tollman<sup>1,5,6</sup>, Alan D Lopez<sup>7</sup>, Theo Vos<sup>8</sup> and Debbie Bradshaw<sup>2</sup>

## Abstract

**Background:** Health and Demographic Surveillance Systems (HDSS) collect independent mortality data that could be used for assessing the quality of mortality data in national civil registration (CR) systems in low- and middle-income countries. However, the use of HDSS data for such purposes depends on the quality of record linkage between the two data sources. We describe and evaluate the quality of record linkage between HDSS and CR mortality data in South Africa with HDSS data from Agincourt HDSS.

**Methods:** We applied deterministic and probabilistic record linkage approaches to mortality records from 2006 to 2009 from the Agincourt HDSS and those in the CR system. Quality of the matches generated by the probabilistic approach was evaluated using sensitivity and positive predictive value (PPV) calculated from a subset of records that were linked using national identity number. Matched and unmatched records from the Agincourt HDSS were compared to identify characteristics associated with successful matching. In addition, the distribution of background characteristics in all deaths that occurred in 2009 and those linked to CR records was compared to assess systematic bias in the resulting record-linked dataset in the latest time period.

**Results:** Deterministic and probabilistic record linkage approaches combined linked a total of 2264 out of 3726 (60.8%) mortality records from the Agincourt HDSS to those in the CR system. Probabilistic approaches independently linked 1969 (87.0%) of the linked records. In a subset of 708 records that were linked using national identity number, the probabilistic approaches yielded sensitivity of 90.0% and PPV of 98.5%. Records belonging to more vulnerable people, including poorer persons, young children, and non-South Africans were less likely to be matched. Nevertheless, distribution of most background characteristics was similar between all Agincourt HDSS deaths and those matched to CR records in the latest time period.

**Conclusion:** This study shows that record linkage of mortality data from HDSS and CR systems is possible and can be useful in South Africa. The study identifies predictors for death registration and data items and registration system characteristics that could be improved to achieve more optimal future matching possibilities.

**Keywords:** Health and demographic surveillance system (HDSS), Agincourt HDSS, Record linkage, Civil registration system, Death registration, South Africa, Mortality

\* Correspondence: chodziwadziwa.kabudula@wits.ac.za

<sup>1</sup>MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Full list of author information is available at the end of the article

## Background

Reliable and valid statistics on the levels and causes of mortality are widely acknowledged as essential information for monitoring the impact of health interventions and developing public health policies and programs for improving population health [1-4]. An adequate and complete civil registration (CR) system is the ideal source from which to draw such information [2,5].

Well-functioning CR systems do not exist in the majority of African countries [6]. South Africa is one of the few that produce mortality statistics from a CR system [7,8], but previous assessments rated their quality as low [2,9]. In recent years, the country has adopted the Africa Programme on Accelerated Improvement of Civil Registration and Vital Statistics (APAI-CRVS) [6], building on the focused initiatives by Statistics South Africa, the Department of Health, and a group of researchers since the 1990s to improve and strengthen its CR system and cause of death information [10-12]. Therefore, there is a continuous need for assessing the quality of CR mortality data to ascertain the impact of these initiatives and identify remaining gaps and options for further improvement.

A number of criteria, organized into a framework of four quality concepts (generalizability, reliability, validity, and policy relevance), have been proposed for comprehensive assessment of the quality of mortality data in CR systems [13,14]. Although most criteria can be evaluated directly from the mortality data recorded in the CR system and administrative information on the system, data from other sources are also required [13]. Combining vital-event data sources, and cooperation among the custodians of such data sources, was encouraged at the 2012 International Network for the Demographic Evaluation of Populations and Their Health in developing countries (INDEPTH) - African Census Analysis Project (ACAP) Bellagio meeting on using longitudinal INDEPTH data, national censuses, Demographic and Health Surveys, and other national surveys for better health policy in Africa [15].

In South Africa, three INDEPTH Health and Demographic Surveillance Systems (HDSS) collect mortality data in rural populations [16-18]. Such data provide an opportunity for comparison with CR data. However, this requires record linkage between the two data sources, which has not been attempted before. Both data sources are protected by strict data-use clauses to protect the confidentiality of the identity and other information of the deceased. Once linked, comparison would also depend on the quality of the matched records.

This paper describes the practical steps we took to set up and execute record linkage of mortality data and evaluates the quality of the matched records between the CR system and the longest-running of the three HDSS centers in South Africa, the Agincourt HDSS

[19,20]. It describes how we overcame the challenges of bringing together data that are kept in secure databases and environments almost 600 kilometers apart, each governed by data-security policies that prohibit the off-site and non-staff use of unit-record data that contain personal identifiers.

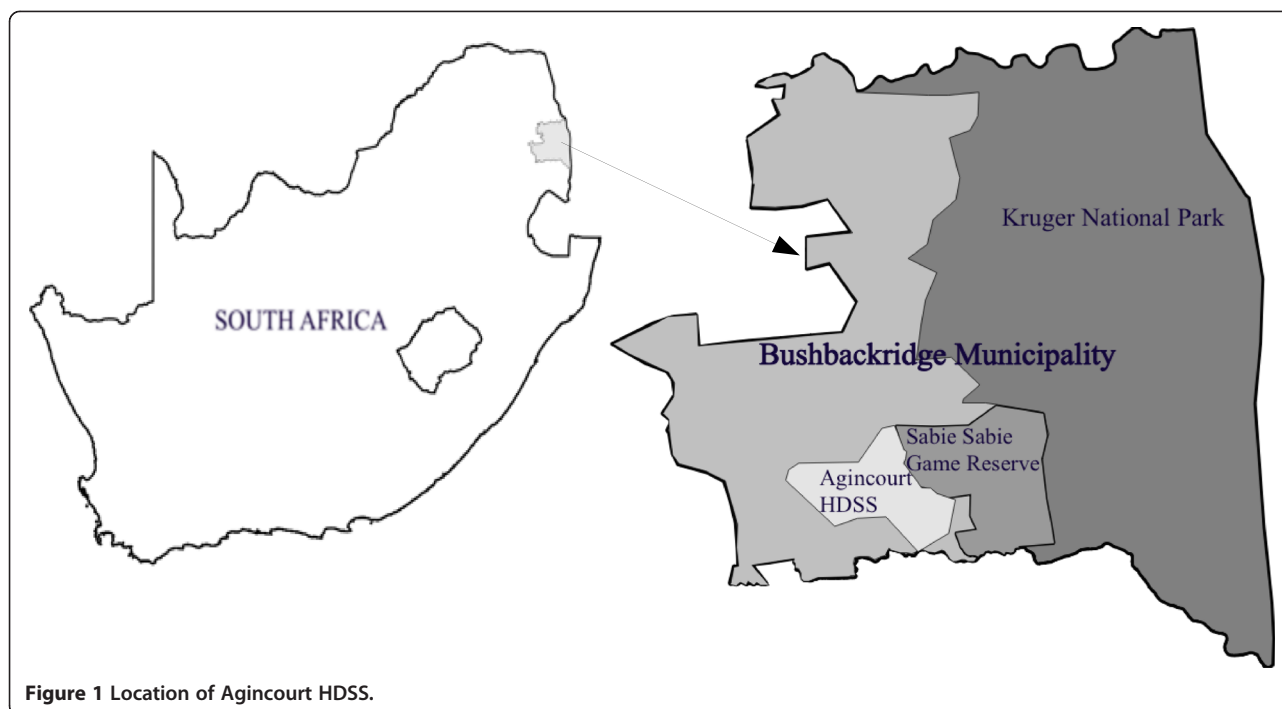
## Methods

### Data sources

Records of individuals who died from 1 January 2006 to 31 December 2009 were extracted from the Agincourt HDSS database and saved under password protection on a portable device. An Agincourt HDSS staff member who is familiar with the collection, processing, and coding of mortality data and the stringent data-use policies at Agincourt, and who had previous experience in electronic record linkage, securely brought the data files to Statistics South Africa's (Stats SA) head office in Pretoria. After confidentiality and data-security agreements were undertaken and signed by the Agincourt HDSS staff member and other members of the record linkage team, the non-Stats SA team members were given access to the secure environment in the Stats SA building where CR data for deaths that occurred within the same period were made available for linkage.

Information about deaths in the Agincourt HDSS was collected as part of annual updates of vital events in a surveillance population occupying 27 villages in Bushbuckridge municipality, Mpumalanga province, South Africa (Figure 1) [19,20]. The population is largely Tsonga-speaking, and a third are of Mozambican descent who arrived in the study area in the early to mid-1980s as refugees and/or their descendants. The population has been under surveillance since 1992. Residency status and vital events have been updated at approximately 15- to 18-month intervals between 1993 and 1999, and annually since 1999. During the annual update, an individual who was present at home for at least six months in the last 12 months is considered a permanent resident. Permanent residency status is also assigned to an individual who in-migrated or a child who was born prior to the annual update and is considered as a permanent resident by the household informant. Individuals who were present at home for less than six months in the previous 12 months due to work-related, education-related, and other reasons are assigned a temporary residency status. For each death recorded during the annual update, a verbal autopsy (VA) interview is conducted with caregivers of deceased individuals one to 11 months after death to elicit signs and symptoms of the illness or injury prior to death using a locally validated, local-language VA instrument [20,21]. Because vital events are updated every year, death events missed in one year are captured the following year since the deceased individuals still appear on pre-populated household rosters. Hence, completeness of





**Figure 1** Location of Agincourt HDSS.

recording of deaths into the Agincourt HDSS is very high despite some under-recording of neonatal deaths. In addition, the data items collected pertaining to the characteristics of the event of death and the deceased person cover most of the core topics of themes for vital statistics as recommended by the United Nations [22]. These include date of death, date of death registration when a death certificate is available, place of death, cause of death derived from verbal responses to autopsy interviews, date of birth of the deceased, place of usual residence of the deceased, and marital status of the deceased.

The CR data were captured by Stats SA from *Notification of death/still-birth* forms (Form BI-1663) that were submitted to the Department of Home Affairs offices for death registration as required by the country's Births and Deaths Registration Act No 51 of 1992 [23]. As required by the Act, different sections of the form are completed by (i) the person reporting the death, (ii) a medical practitioner (where a medical practitioner is not available, a traditional leader may complete the Death Report (Form BI-1680)), and (iii) a Home Affairs official or member of the South African Police Services if the former is not available [24,25].

#### Record linkage procedures

We applied deterministic and probabilistic record linkage approaches to link the Agincourt HDSS and CR mortality data. Variables common to both data sources that we used are: national identity number (a unique 13-digit number assigned to South African citizens), surname, sex, day of

death, month of death, year of death, day of birth, month of birth, year of birth, institution/place of death, and village name. For village name, village of the household of the deceased individual in the Agincourt HDSS was matched to place of birth, residency, and death in the CR records. Due to the recording of local tribal area names rather than the official village names for some deaths on the CR death registration forms, the place names in the CR records were mapped to their equivalent Agincourt HDSS village names prior to the record linkage exercise.

In deterministic record linkage, a pair of records from two data sources is considered to belong to the same individual if it matches on a unique identifier such as national identification number or a set of conventional personal identifiers (e.g., the combination of first name, last name, and date of birth) [26-29]. We defined 12 deterministic linkage rules based on different combinations of the common variables as presented in Table 1. Record linkage using these rules proceeded iteratively. Records matched by one rule were removed from the pool of records to be matched with subsequent rules in both datasets. The Jaro-Winkler string comparator (JW) [30], which is particularly well-suited for personal names [31], returning values between 0 (complete disagreement) and 1 (exact agreement) as a measure of similarity between two strings [30,32], was used to accommodate typographical errors on surnames. We set a cut-off for designating pairs of surnames as matches to a JW score  $\geq 0.85$ , which is higher than in previous studies [30,33].

**Table 1 Deterministic matches**

Rule number	Description	Matches in trimmed CR dataset	Matches in full CR dataset	Total matches
1	Match on National ID No	708	161	869
2	Match on Surname, Sex, Date of birth, Date of death	128	28	156
3	Match on Surname, Sex, Date of birth, Year of death, Month of death	88	30	118
4	Match on Surname, Sex, Year of birth, Month of birth, Date of death	34	6	40
5	Match on JW(Surname) > =0.85, Sex, Date of birth, Date of death	39	-	39
6	Match on JW(Surname) > =0.85, Sex, Date of birth, Year of death, Month of death	37	-	37
7	Match on JW(Surname) > =0.85, Sex, Year of birth, Month of birth, Date of death	6	-	6
8	Match on JW(Surname) > =0.85, Sex, Year of birth, Year of death, Agincourt HDSS village = CR place of birth	207	-	207
9	Match on JW(Surname) > =0.85, Sex, Year of birth, Year of death, Agincourt HDSS village = CR place of residence	67	-	67
10	Match on JW(Surname) > =0.85, Sex, Year of birth, Year of death, Agincourt HDSS village = CR place of death	23	-	23
11	Match on JW(Surname) > =0.85, Sex, Year of birth, Date of death, died at hospital	30	-	30
12	Match on JW(Surname) > =0.85, Sex, Date of birth, Year of death, died at hospital	27	-	27
Total		1,394	225	1,619

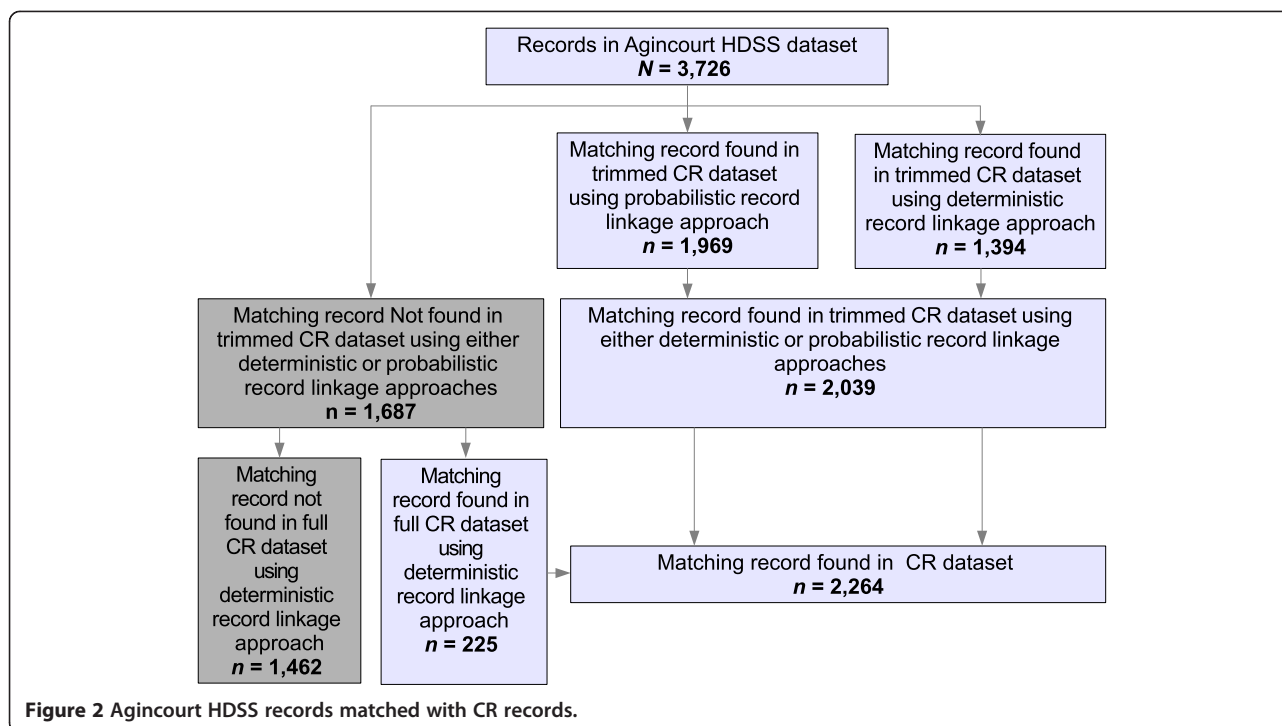
In probabilistic record linkage, a pair of records from two data sources is classified as a match based on the statistical probability that the values of common variables from the two data sources belong to the same individual [32,34-38]. Each matching variable is assigned a weight that indicates its contribution to the probability of accurately designating a pair of records as a match or non-match [29,32,36]. The weight of a matching variable,  $i$ , is calculated from the probability that records belonging to the same individual agree, denoted by  $m_i$ , and the probability that records belonging to different individuals agree, denoted by  $u_i$  [32,36,38]. Record pairs where variable  $i$  agrees receive a weight value of  $\log_2 \frac{m_i}{u_i}$ , and those where the variable disagrees get a weight value of  $\log_2 \frac{1-m_i}{1-u_i}$ . A record pair is classified as a match if the sum of the weights on all the matching variables is above a particular threshold value. We estimated  $m_i$  and  $u_i$  values for all matching variables, except national identity number, using the Expectation Maximization (EM) algorithm [32,39,40]. Only surname pairs with a JW score  $\geq 0.85$  were considered as matches. Similar to the work of Méray et al. [41] and Tromp et al. [42], the threshold value for determining which record pairs were matches was derived from an estimate of the proportion of true matches among all possible record pair combinations produced by the EM algorithm. The estimated proportion of true matches was multiplied by the total number of all possible record pair combinations to obtain the total number of

true matches. Thereafter, all possible record pair combinations were sorted in descending order of the sum of the weights on all matching variables and the top  $n$  record pairs, where  $n$  equals the calculated number of true matches, were designated as matches.

The number of possible record pair comparisons in two files to be linked is equal to the product of the numbers of records on each file, which can be enormous. Therefore, we used blocking to reduce the number of record pair comparisons [32]. We restricted the comparisons to “blocks” or “pockets” of record pairs with exact matches on one or more variables. We applied the 12 deterministic rules to link the Agincourt HDSS dataset with a trimmed CR dataset that had records for which the deceased was either born, resident, or died within the Bushbuckridge municipality. We repeated the linkage between the Agincourt HDSS dataset with the trimmed CR dataset using the probabilistic approach described above with further blocking on (i) sex and year of birth, and (ii) sex and year of death. Finally, we applied the first four strict deterministic rules in Table 1 to link the thus-far unmatched records in the Agincourt HDSS dataset with records in the full CR dataset (compare Figure 2).

#### Evaluation of record linkage results

Since we set strict deterministic matching rules with very narrow margins for error, evaluation of the record linkage results focused on matches generated by the



**Figure 2** Agincourt HDSS records matched with CR records.

probabilistic record linkage approach. Their quality was evaluated using sensitivity and PPV calculated from a subset of records that were linked by means of national identity number. This is justifiable because national identity numbers contain a check digit that prevents incorrect matching. We also compared characteristics of the deceased individuals in the Agincourt HDSS dataset whose records were matched and unmatched to records in the CR dataset in logistic regression models to identify characteristics associated with successful matching. Variables selected for analysis included sex, age, nationality, having a national identity number, residency status, level of education, wealth quintile, year of death, and place of death. Wealth quintiles were derived from data on ownership of assets such as cattle, a car, and cell phone and access to amenities including drinking water and sanitation using principal component analysis [43]. In addition, the distribution of background characteristics in all deaths that occurred in 2009 and those linked to CR records was compared using Pearson Chi squared tests to assess systematic bias in the resulting record-linked dataset in the latest time period.

#### Software

The record linkage of the data between the two data sources was done using Microsoft SQL Server 2008 which had the EM algorithm implemented in Microsoft C# programming language, integrated in it as a common language runtime (CLR) function. The JW algorithm we used is part of the SimMetrics library [44]. It was also

integrated in Microsoft SQL Server 2008 as a CLR function. Stata (version 11.2, Stata Corporation, Texas, USA) was used for data analysis.

#### Ethics

The study received ethical approvals from the University of Queensland School of Population Health Research Ethics Committee (approval no. JJ010911), the South African Medical Research Council Ethics Committee (EC008-6/2011), and the University of the Witwatersrand Human Research Ethics Committee (Medical) (M120106).

#### Results

The Agincourt HDSS and full CR datasets, respectively, contained 3,726 and 2,464,915 unit records of individuals who died from 2006 to 2009. Place of birth, residence, or death in 29,416 of the records in the CR dataset was within the Bushbuckridge municipality. Overall, 2,264 (60.8%) of the 3,726 records from Agincourt HDSS were matched to records in the CR dataset (Figure 2). The matched record pairs comprised of 2,039 record pairs that were obtained by linking Agincourt HDSS records to records in the trimmed CR dataset, and 225 record pairs obtained by linking the remaining Agincourt HDSS records with records in the full CR dataset. Deterministic and probabilistic record linkage approaches, respectively and independently, produced 1,394 and 1,969 of the record pairs that were obtained by linking Agincourt HDSS records to records in the trimmed CR dataset, and 1,324 (95.0%) of the records that were matched using the

deterministic approach were also matched using the probabilistic approach. The weights computed for probabilistic record linkage for each of the variables in the two blocking schemes are presented in Tables 2 and 3. The weights indicate that village had the highest agreement weight followed by surname, and for the block on sex and year of death, record pairs that agreed on all the variables were assigned an overall weight score of 32.91344, whereas those that disagreed on all the variables were assigned an overall weight score of -14.02270.

Most of the record pairs that were generated by linking the remaining Agincourt HDSS records with records in the full CR dataset had Hazyview, a town about 40 km away from the Agincourt HDSS, as the reported place of birth, residence, or death in the CR dataset. There were also a few cases for which the reported place of birth, residence, or death in the CR dataset is indeed within the Agincourt HDSS study site, such as Belfast and Somerset, but had not been assigned to the Bushbuckridge municipality in the CR system. For example, one of the death records from Somerset village in the Agincourt HDSS dataset was in the CR dataset assigned to Somerset West, a town in the Western Cape province. Over half (53.7%) of the combined deterministic matches were found via the deceased's identity number (Table 1).

In a subset of 708 records from the Agincourt HDSS that were deterministically linked by means of national identification number, the probabilistic approaches yielded sensitivity of 90.0% and a positive predictive value of 98.5%.

Table 4 presents a number of characteristics of the deceased in the Agincourt HDSS dataset and their association with successful matching to individuals in the CR dataset. In a univariate model, higher matching likelihood was associated with age (age >=5 years), nationality (South African), having a national identity number recorded in the VA system, residency status prior to death (permanently residing in the Agincourt HDSS study area), education (at least some primary education, but higher likelihood with secondary), wealth (from second to highest wealth quintile), year of death (more recent year), and

place of death (place of death specified). In a multivariable model, matched and non-matched cases differed significantly in terms of age, nationality, having a national identity number recorded, residency status, wealth (from middle to highest wealth quintile), and place of death (place of death specified). Having adjusted for all variables in the model, having a national identity number recorded increased the odds to be matched by almost 14 times.

Although a number of characteristics prevented successful record linkage of mortality between the Agincourt HDSS and CR system for the period considered in this study, in the latest time period (2009), except for infants and non-South Africans, there were no significant differences in the distribution of background characteristics in all Agincourt HDSS deaths compared to those matched with CR records (Table 5).

## Discussion

In South Africa, there are no comprehensive systems of pre-linked health data covering large or entire populations such as the Manitoba Population Health Information System in Canada [45] or systems that routinely or periodically link data at any level of jurisdiction. In this study, we have assessed the feasibility of setting up and executing record linkage of mortality data and evaluated the quality of the matched records between the Agincourt HDSS and the CR system. The study was motivated by the unexplored potential of HDSS as sources of independent mortality data for assessing the quality of mortality data in CR systems in low-and middle-income countries.

Using deterministic and probabilistic approaches, our study yielded a matching rate of 60.8% for mortality records from 2006 to 2009, with sensitivity of 90% and PPV of 98.5% for the probabilistic linkage. This matching rate was influenced by a number of limitations relating to the amount, accuracy, completeness, and consistency of information available for the linkage process [46]. First, we had a small number of common variables in the two datasets. Second, collection of the ideal unique-identifier variable, national identity number, was introduced gradually in the Agincourt HDSS over the period of our investigation,

**Table 2 Weights for the probabilistic linkage approach with blocking on sex and year of death**

Variable	$m_i$	$u_i$	Agreement weight	Disagreement weight
Surname	0.80987	0.01581	5.67838	-2.37191
Day of birth	0.62776	0.03498	4.16563	-1.37431
Month of birth	0.68986	0.08655	2.99473	-1.55840
Year of birth	0.80879	0.01643	5.62171	-2.36291
Month of death	0.80330	0.08366	3.26334	-2.21987
Day of death	0.53200	0.03253	4.03141	-1.04770
Village	0.60572	0.00832	6.18540	-1.33066
Institution/place of death	0.82912	0.42244	0.97283	-1.75695

**Table 3 Weights for the probabilistic linkage approach with blocking on sex and year of birth**

Variable	$m_i$	$u_i$	Agreement weight	Disagreement weight
Surname	0.81578	0.01588	5.68252	-2.41737
Day of birth	0.69855	0.03605	4.27643	-1.67704
Month of birth	0.75919	0.08734	3.11968	-1.92216
Year of death	0.98414	0.28645	1.78059	-5.49130
Month of death	0.80748	0.08367	3.27068	-2.25083
Day of death	0.52955	0.03286	4.01026	-1.03967
Village	0.60579	0.00808	6.22918	-1.33126
Institution/place of death	0.82785	0.43602	0.92498	-1.71199

starting only in 2007. However, it is worth noting that as of 2013, national identity number was available on 68% of the individuals still under surveillance in the Agincourt HDSS. Therefore, national identity number has an increased future potential as a unique matching variable. Third, we set strict deterministic matching rules with narrow margins for error, such as in the case of the spellings of surnames. Fourth, there has been a particular problem with the reporting of tribal area names instead of village names for some deaths in the death registration system. As more than one village is contained in a tribal area, it is not possible to correct this data entry. Last, the use of proxy respondents, inevitably, in both VA and CR systems, and that VA interviews are conducted one to 11 months after death, may also have reduced the accuracy of individual-level information.

While the record linkage approach employed in this study would typically allow the assessment of completeness using a standard two-source capture-recapture analysis [47,48], it is not possible in our study. This stems from difficulties in identifying CR deaths that occurred within the Agincourt HDSS borders due to the recording of local tribal area names rather than the official village names on the CR death registration forms for some deaths. The three tribal areas containing the study site additionally include areas not covered by the Agincourt HDSS. Furthermore, the places of birth, death and residence in the CR data, reported by the relative or friend of the deceased, were not verified against the StatsSA official or Agincourt HDSS colloquial place names. Valuable lessons were learned in this regard, and recommendations are offered in the Conclusion.

Even though the matching rate in this study is low and it is not possible to assess completeness of death registration using a standard two-source capture-recapture analysis due to the limitations above, the similarity in the distribution of most of the background characteristics in all Agincourt HDSS deaths compared to those matched with CR records in the latest time period (2009) suggests that the record-linked data can enhance understanding of death registration practices into the

CR system through identifying subgroups likely to be underrepresented in the CR data. For example, the finding that after adjusting for other variables, matching rates are significantly lower for records belonging to more vulnerable people, including poorer persons, children <5 years, and non-South Africans could possibly be interpreted to mean that their deaths are less likely to be registered. In addition, adding cause of death data to the record-linked data can also allow cause attribution and leading cause of death comparisons between the data sources. Such analyses, accompanied by careful interpretation, can form a useful basis from where to adjust cause of death data according to observed biases. At the individual level, misclassification patterns can be identified, which can offer insight into newly identified and re-occurring biases in cause of death attribution. Cause of death analyses using the record-linked data generated in this linkage study are presented in a forthcoming paper [49].

### Conclusion

Despite strict policies to protect the confidentiality and safety of the data reported into each system, record linkage of mortality data between a CR system and an HDSS was possible in our study. To our knowledge, our study is the first in South Africa and possibly in sub-Saharan Africa to assess the feasibility and utility of linking HDSS and CR mortality data. The resultant data are useful for assessing selected population and individual health measures as referred to above, and hold potential to improve rural data quality.

We suggest the following five crucial contributions for further fruitful linkage exercises: the routine collection of national identity number in all the South African HDSSs; collaborative efforts to address place-name inconsistencies; recording of actual village/town/suburb names on death notification forms instead of tribal area names or adequate provision to provide for both; the development of an electronic place-name database, linked to detailed maps, against which to verify place names reported into the CR system, for use by Home Affairs registration

**Table 4 Factors predictive of successful matching of death records between Agincourt HDSS and South African CR system**

Variable	n	Matched n (%)	Univariate Odds ratio (95% confidence interval)	Multivariable Odds ratio (95% confidence interval)
	3726	2264 (60.8)		
<b>Sex</b>				
Female	1771	1104 (62.34)	1.00	1.00
Male	1955	1160 (59.34)	0.88 (0.77-1.01)	0.97 (0.82-1.14)
<b>Age (years)</b>				
<5	555	213 (38.38)	1.00	1.00
5-14	106	67 (63.21)	2.76 (1.79-4.24)***	2.83 (1.64-4.90)***
15-49	1729	1126 (65.12)	3.00 (2.46-3.65)***	2.89 (2.11-3.96)***
50-64	575	368 (64)	2.86 (2.24-3.63)***	2.24 (1.65-3.05)***
65+	761	490 (64.39)	2.90 (2.31-3.64)***	1.87 (1.41-2.49)***
<b>Nationality</b>				
Other nationality	1191	569 (47.77)	1.00	1.00
South African	2531	1695 (66.97)	2.22 (1.93-2.55)***	2.05 (1.70-2.48)***
<b>National Identity number recorded in VA system</b>				
Not available	2722	1324 (48.64)	1.00	1.00
Available	1004	940 (93.63)	15.51 (11.91-20.2)***	13.65 (10.12-18.43)***
<b>Residency status</b>				
Temporary and other	1211	642 (53.01)	1.00	1.00
Permanent	2515	1622 (64.49)	1.61 (1.40-1.85)***	1.28 (1.06-1.54)*
<b>Education</b>				
None	1720	936 (54.42)	1.00	1.00
Primary	1070	694 (64.86)	1.55 (1.32-1.81)***	0.92 (0.72-1.17)
Post primary	733	512 (69.85)	1.94 (1.61-2.33)***	1.02 (0.75-1.39)
<b>Wealth quintile</b>				
Lowest	605	308 (50.91)	1.00	1.00
Second	631	375 (59.43)	1.41 (1.13-1.77)**	1.21 (0.93-1.58)
Middle	647	406 (62.75)	1.62 (1.3-2.04)***	1.44 (1.1-1.88)*
Fourth	752	494 (65.69)	1.85 (1.48-2.3)***	1.59 (1.22-2.07)**
Highest	745	501 (67.25)	1.98 (1.59-2.47)***	1.73 (1.32-2.27)***
<b>Year of death</b>				
2006	885	453 (51.19)	1.00	1.00
2007	901	518 (57.49)	1.29 (1.07-1.55)**	1.16 (0.94-1.44)
2008	1024	655 (63.96)	1.69 (1.41-2.03)***	0.99 (0.79-1.24)
2009	916	638 (69.65)	2.19 (1.8-2.65)***	0.83 (0.64-1.06)
<b>Place of death</b>				
Hospital	1759	1117 (63.5)	1.00	1.00
Health center	42	25 (59.52)	0.85 (0.45-1.58)	1.07 (0.51-2.23)
Clinic	26	19 (73.08)	1.56 (0.65-3.73)	1.23 (0.41-3.68)

**Table 4 Factors predictive of successful matching of death records between Agincourt HDSS and South African CR system (Continued)**

Home	1525	937 (61.44)	0.92 (0.79-1.06)	1.12 (0.93-1.34)
Vehicle accident site	102	59 (57.84)	0.79 (0.53-1.18)	0.79 (0.48-1.29)
Other	255	98 (38.43)	0.36 (0.27-0.47)***	0.43 (0.31-0.6)***
Unknown	17	9 (52.94)	0.65 (0.25-1.68)	1.13 (0.31-4.06)

Statistical significance: \*\*\*P < 0.001; \*\*P < 0.01; \*P < 0.05.

**Table 5 Background characteristics of all 2009 Agincourt HDSS deaths compared to those matched with CR records**

Variable	All deaths in Agincourt HDSS (n = 846)		Deaths matched with CR records (n = 618)	
	n (%)		n (%)	p-value
<b>Sex</b>				
Female	411 (48.58)		307 (49.68)	
Male	435 (51.42)		311 (50.32)	0.679
<b>Age (years)</b>				
1-4	48 (5.67)		22 (3.56)	
5-14	25 (2.96)		20 (3.24)	
15-49	405 (47.87)		299 (48.38)	
50-64	138 (16.31)		109 (17.64)	
65+	230 (27.19)		168 (27.18)	0.431
<b>Nationality</b>				
Other	246 (29.08)		145 (23.46)	
South African	599 (70.80)		473 (76.54)	0.038
<b>Residence status</b>				
Permanent	623 (73.64)		467 (75.57)	
Temporary and other	223 (26.36)		151 (24.43)	0.404
<b>Highest level of education</b>				
None	350 (41.37)		238 (38.51)	
Primary	236 (27.90)		190 (30.74)	
Post primary	199 (23.52)		147 (23.63)	0.622
<b>Place of death</b>				
Hospital	439 (51.89)		327 (52.91)	
Health center	8 (0.95)		5 (0.81)	
Clinic	4 (0.47)		4 (0.65)	
Home	329 (38.89)		244 (39.48)	
Vehicle accident site	19 (2.25)		13 (2.10)	
Other	46 (5.44)		24 (3.88)	
Unknown	1 (0.12)		1 (0.16)	0.894
<b>Wealth quintile</b>				
Lowest	127 (15.01)		72 (11.65)	
Second	134 (15.84)		92 (14.89)	
Middle	150 (17.73)		111 (17.96)	
Fourth	184 (21.75)		142 (22.98)	
Highest	180 (21.28)		148 (23.95)	0.469

offices; and aligning study site borders with established official borders when setting up or extending HDSS sites.

Given our success in matching with surnames in this study and other studies' successes in using names [50,51], we additionally recommend that in addition to the surname, the deceased's full names (which are already captured on notice of death/stillbirth forms) be included in StatsSA datasets. Finally, concerted action among the governmental departments involved, health researchers, and relevant health data advisory committees is suggested to revitalize/modify the data fields on the notification form such that it is possible to identify the place of death, death registration, most recent employment prior to death, and residence of the deceased.

From a broader perspective, the methods and findings from this study are also of interest given the potential for application in other HDSS sites. Currently there are more than 45 HDSS sites across Africa, Asia and Oceania [16,18]. Conducting similar studies could serve to evaluate CR data where available, help identify gaps in national or sample CR systems, and where feasible, guide improved mortality and cause of death estimates. Of special interest would be the conduct of a similar study using data from an urban HDSS, such as DodaLab in Vietnam [52], to obtain empirical evidence for or against the general assumption that death registration is more complete in urban compared to rural areas and to help identify under-registered groups in urban areas. Such an empirical approach has potential to strengthen the evidence base for population health assessment and policy in developing countries where CR systems are weak.

Finally, our study provides scarce empirical evidence about factors affecting death registration, which has implications for strategies to accelerate death registration in countries with deficient CR systems.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

JJ wrote the study proposal and ethics applications and coordinated the collaboration, data preparation, and matching exercise. DB conceptualized the paper with inputs from JJ and CWK. CWK extracted the Agincourt HDSS data, and with inputs from DB and JJ, did the electronic matching, created the base analytic dataset, analyzed and interpreted the data, created the tables and figures, and wrote the first draft of the Introduction, Methods, and Results. JJ wrote the first draft of the Discussion and critically appraised the other sections for structure and intellectual content. CWK and JJ integrated the comments from co-authors and external referees. ADL, CR, DB, and TV consistently supplied critical inputs during all study phases. ADL, DB, KK, MT-N, ST, and TV made substantial contributions to acquisition of the data. MT-N led the vital registration data extraction and data security arrangements. All authors contributed to interpreting the data, critically reviewed the drafts, and approved the final manuscript.

#### Acknowledgments

We thank Statistics South Africa and the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) for making available data used in this study. We are also grateful to Statistics South Africa for housing the matching exercise at the head office in Pretoria; Ms Ramadimetja Matji, Ms

Aletia Barkley, and Ms Kerotse Mmatli for their participation in meetings and contributions to securing the data prior to the matching exercise; Ms Marlanie Moodley for preparing maps of the Agincourt and Bushbuckridge areas; and Ms Ria Laubscher for her assistance during the matching exercise.

#### Funding/Sponsorship

The study was supported by the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), the South African Medical Research Council (MRC), and Statistics South Africa. The study was conducted while the second author held a University of Queensland Research Scholarship and the Endeavour International Postgraduate Research Scholarship at the University of Queensland, Brisbane, Australia. The Agincourt HDSS is funded by the Medical Research Council and University of the Witwatersrand, South Africa, Wellcome Trust, UK (grant no. 058893/Z/99/A, 069683/Z/02/Z, 085477/Z/08/Z), and National Institute on Aging of the NIH (grants 1R24AG032112-01 and 5R24AG032112-03). This paper was first presented at the INDEPTH Scientific Conference, October 2013, and was supported by an INDEPTH travel award. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

#### Author details

<sup>1</sup>MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. <sup>2</sup>Burden of Disease Research Unit, South African Medical Research Council, Parow Valley, Western Cape, South Africa. <sup>3</sup>School of Population Health, The University of Queensland, Herston, Brisbane, Queensland, Australia. <sup>4</sup>Health and Vital Statistics, Statistics South Africa, Pretoria, South Africa. <sup>5</sup>Umeå Centre for Global Health Research, Division of Epidemiology and Global Health, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. <sup>6</sup>INDEPTH Network, Accra, Ghana. <sup>7</sup>Melbourne School of Population and Global Health, The University of Melbourne, Carlton, Victoria, Australia. <sup>8</sup>Institute of Health Metrics and Evaluation, University of Washington, Seattle, USA.

Received: 17 April 2014 Accepted: 11 August 2014

Published: 30 August 2014

#### References

- Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, AlMazroo MA, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Abdulhak AB, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C: **Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010.** *Lancet* 2012, **380**:2095–2128.
- Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, Rao C, Szreter S: **Civil registration systems and vital statistics: successes and missed opportunities.** *Lancet* 2007, **370**:1653–1663.
- Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, Levin-Rector A, Levitz CE, Lopez AD, Murray CJ: **Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010.** *Lancet* 2012, **380**:2071–2094.
- Carter KL, Rao C, Lopez AD, Taylor R: **Mortality and cause-of-death reporting and analysis systems in seven pacific island countries.** *BMC Public Health* 2012, **12**:436.
- Rao C, Osterberger B, Anh TD, MacDonald M, Chuc NTK, Hill PS: **Compiling mortality statistics from civil registration systems in Viet Nam: the long road ahead.** *Bull World Health Organ* 2009, **87**:58–65.
- United Nations, African Union Commission, African Development Bank: *Africa Programme on Accelerated Improvement of Civil Registration and Vital Statistics (APAI-CRVS)*. New York: Economic Commission for Africa, United Nations; 2011.
- Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S, AbouZahr C: **A scandal of invisibility: making everyone count by counting everyone.** *Lancet* 2007, **370**:1569–1577.
- Bradshaw D, Groenewald P, Laubscher R, Nannan N, Nojilana B, Norman R, Pieterse D, Schneider M, Bourne DE, Timaeus I, Dorrington R, Johnson L: **Initial burden of disease estimates for South Africa, 2000.** *SAMJ* 2003, **93**:682–688.



9. Mathers CD, Ma Fat D, Inoue M, Rao C, Lopez AD: **Counting the dead and what they died from: an assessment of the global status of cause of death data.** *Bull World Health Organ* 2005, **83**:171–177c.
10. Bradshaw D, Kielkowsk D, Sitas F: **New birth and death registration forms - a foundation for the future, a challenge for health workers?** *SAMJ* 1998, **88**:971–974.
11. Rao C, Bradshaw D, Mathers CD: **Improving death registration and statistics in developing countries: lessons from sub-Saharan Africa.** *South Afr J Demogr* 2004, **9**:81–99.
12. Bah S: **Multiple forces working in unison: the case of rapid improvement of vital statistics in South Africa post-1996.** *World Health Popul* 2009, **11**:50–59.
13. Rao C, Lopez AD, Yang G, Begg S, Ma J: **Evaluating national cause-of-death statistics: principles and application to the case of China.** *Bull World Health Organ* 2005, **83**:618–625.
14. Joubert J, Rao C, Bradshaw D, Vos T, Lopez AD: **Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007.** *PLoS ONE* 2013, **8**:e64592.
15. INDEPTH Network: *Using Longitudinal INDEPTH Data, National Censuses, DHS, and Other National Surveys for Better Health Policy in Africa. Report of Meeting nr 1352.* Bellagio, Italy: INDEPTH Network; 2012.
16. Sankoh O: **Global health estimates: stronger collaboration needed with low-and middle-income countries.** *PLoS Med* 2010, **7**:e1001005.
17. Ye Y, Wamukoya M, Ezech A, Emina J, Sankoh O: **Health and demographic surveillance systems: a step towards full civil registration and vital statistics system in sub-Sahara Africa?** *BMC Public Health* 2012, **12**:741.
18. Sankoh O, Byass P: **The INDEPTH Network: filling vital gaps in global epidemiology.** *Int J Epidemiol* 2012, **41**:579–588.
19. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, Afolabi SA, Clark BD, Kabudula CW, Khosa A, Khoza S, Shabangu MG, Silaule B, Tibane JB, Wagner RG, Garenne ML, Clark SJ, Tollman SM: **Profile: Agincourt Health and Socio-demographic Surveillance System.** *Int J Epidemiol* 2012, **41**:988–1001.
20. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, Shabangu M, Gomez-Olive FX, Mokoena O, Garenne ML: **Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt Health and Demographic Surveillance System.** *Scand J Public Health* 2007, **35**:8–20.
21. Kahn K, Tollman SM, Garenne M, Gear JSS: **Validation and application of verbal autopsies in a rural area of South Africa.** *Trop Med Int Health* 2000, **5**:824–831.
22. Division UNS: *Principles and Recommendations for a Vital Statistics System. Revision 3. Final Draft.* New York: United Nations; 2013.
23. Republic of South Africa: *Births and Deaths Registration Act, 1992 (No. 51 of 1992).* In: *Government Gazette No. 13953.* Cape Town: Government Printer; 1992.
24. Statistics South Africa: *Mortality and Causes of Death in South Africa, 2008: Findings from Death Notification. Statistical Release P0309.3.* Pretoria: Statistics South Africa; 2010.
25. **Deaths Certificates.** [http://www.home-affairs.gov.za/index.php/death-certificates1]
26. Li B, Quan H, Fong A, Lu M: **Assessing record linkage between health care and Vital Statistics databases using deterministic methods.** *BMC Health Serv Res* 2006, **6**:48.
27. Machado CJ: **A literature review of record linkage procedures focusing on infant health outcomes.** *Cadernos de Saúde Pública* 2004, **20**:362–371.
28. Maso LD, Braga C, Franceschi S: **Methodology used for software for automated linkage in Italy (SALI).** *Comp Biomed Research* 2001, **34**:395.
29. Victor TW, Mera RM: **Record linkage of health care insurance claims.** *J Am Med Inform Assoc* 2001, **8**:281–288.
30. Winkler WE: **String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage.** In *Proceedings of the Section on Survey Research Methods.* Alexandria: American Statistical Association; 1990:354–359.
31. Durham E, Xue Y, Kantarcioglu M, Malin B: **Quantifying the correctness, computational complexity, and security of privacy-preserving string comparators for record linkage.** *Inform Fusion* 2012, **13**:245–259.
32. Jaro MA: **Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida.** *J Am Stat Assoc* 1989, **84**:414–420.
33. Sariyar M, Borg A, Pommerening K: **Evaluation of record linkage methods for iterative insertions.** *Methods Inf Med* 2009, **48**:429–437.
34. Howe GR: **Use of computerized record linkage in cohort studies.** *Epidemiol Rev* 1998, **20**:112–121.
35. Beauchamp A, Tonkin AM, Kelsall H, Sundararajan V, English DR, Sundaresan L, Wolfe R, Turrell G, Giles GG, Peeters A: **Validation of de-identified record linkage to ascertain hospital admissions in a cohort study.** *BMC Med Res Methodol* 2011, **11**:42.
36. Cook L, Olson L, Dean J: **Probabilistic record linkage: relationships between file sizes, identifiers, and match weights.** *Methods Inf Med* 2001, **40**:196–203.
37. Jaro MA: **Probabilistic linkage of large public health data files.** *Stat Med* 1995, **14**:491–498.
38. Nitsch D, Morton S, DeStavola BL, Clark H, Leon DA: **How good is probabilistic record linkage to reconstruct reproductive histories? Results from the Aberdeen children of the 1950 s study.** *BMC Med Res Methodol* 2006, **6**:15.
39. Grannis SJ, Overhage JM, Hui S, McDonald CJ: **Analysis of a probabilistic record linkage technique without human review.** *AMIA Annu Symp Proc* 2003, **2003**:259–263.
40. Herzog TN, Scheuren F, Winkler WE: *Data Quality and Record Linkage Techniques.* Heidelberg: Springer; 2007.
41. Méray N, Reitsma JB, Ravelli AC, Bonsel GJ: **Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number.** *J Clin Epidemiol* 2007, **60**:883–e881.
42. Tromp M, Ravelli AC, Bonsel GJ, Hasman A, Reitsma JB: **Results from simulated data sets: probabilistic record linkage outperforms deterministic record linkage.** *J Clin Epidemiol* 2011, **64**:565–572.
43. Filmer D, Pritchett LH: **Estimating wealth effects without expenditure data or tears: an application to educational enrollments in states of India.** *Demography* 2001, **38**:115–132.
44. *SimMetrics*; [http://sourceforge.net/projects/simmetrics]
45. Roos NP, Black CD, Frohlich N, Decoster C, Cohen MM, Tataray DJ, Mustard CA, Toll F, Carriere KC, Burchill CA, MacWilliam L, Bogdanovic B: **A population-based health information system.** *Med Care* 1995, **33**:DS13–DS20.
46. Karmel R, Rosman D: **Linkage of health and aged care service events: comparing linkage and event selection methods.** *BMC Health Serv Res* 2008, **8**:149.
47. Chandrasekar C, Deming WE: **On a method of estimating birth and death rates and the extent of registration.** *J Am Stat Assoc* 1949, **44**:101–115.
48. Hook EB, Regal RR: **Capture-recapture methods in epidemiology: methods and limitations.** *Epidemiol Rev* 1995, **17**:243–264.
49. Joubert J, Bradshaw D, Kabudula C, Rao C, Kahn K, Mee P, Tollman S, Lopez AD, Vos T: **Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006–09.** *Int J Epidemiol* In press.
50. Kabudula CW, Clark BD, Gómez-Olivé FX, Tollman S, Menken J, Reniers G: **The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa.** *BMC Med Res Methodol* 2014, **14**:71.
51. Quantin C, Binquet C, Bourquard K, Pattisina R, Gouyon-Cornet B, Ferdynus C, Gouyon J-B, Allaert F-A: **Which are the best identifiers for record linkage?** *Inf Health and Social Care* 2004, **29**:221–227.
52. Tran TK, Eriksson B, Nguyen CT, Horby P, Bondjers G, Petzold M: **DodaLab: an urban health and demographic surveillance site, the first three years in Hanoi, Vietnam.** *Scand J Public Health* 2012, **40**:765–772.

doi:10.1186/s12963-014-0023-z

**Cite this article as:** Kabudula et al.: Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa. *Population Health Metrics* 2014 **12**:23.

## 4.6 Concluding remarks

After identifying and confirming quality problems with selected aspects of South African mortality data in the evaluation study in Chapter 3, this chapter documented background information about record linkage as an additional methodology to assess these aspects of data quality, and present the methods and results from a linkage study that was successfully set up and executed to match death records from two independently administrated surveillance systems.

South Africa has no system that routinely or periodically link data at any level of jurisdiction. For South Africa, and possibly sub-Saharan Africa, the chapter contributes to the literature the first study to successfully link HDSS and official routine CVSR mortality data, despite operational and methodological challenges.

It would have been very useful if it was possible to measure the level of completeness of CRVS death registration in the studied rural area, in particular because completeness in rural areas is unknown in the country, and because matching studies are the only direct method for assessing completeness once the related assumptions apply. Almost all the underlying assumptions for a capture-recapture study design were met in the study, i.e. homogeneity of capture probability (i.e. that for each data source each individual has the same probability of being included in the source); source independence (i.e. that a person's presence in one source does not influence the probability of its presence in the other source); that the population under study is closed (i.e. no in- and outward migration, and where entry into or exit from the population occur, then such movement is documented and taken account of); that the matching of cases has been done appropriately; and that all identified cases should be true cases (i.e. that all individuals are deceased).<sup>234</sup> The only assumption not met, relates to the expectation that all the cases should be within the time-space unit that are studied, which was not possible to assess due to the recording of tribal area names, rather than village names, on selected CRVS death notification forms while the three tribal areas each include additional areas not covered by the Agincourt field site. Such lack of coterminous or common boundaries needs to be addressed, in particular if co-ordinates are not publically available at very small area sizes. This leads to 'missing' decedents in the Agincourt field site boundaries, and in this way would produce an under-count of the real number of decedents. The real number is critically required in the frequencies that are needed to reliably calculate capture-recapture completeness.

Despite being unable to calculate completeness, the linkage and the resultant linked dataset were found useful in a number of ways:

- A number of data items and registration system characteristics were identified that could be improved in order to achieve more optimal matching possibilities in future;
- Documented methods and the findings point to the potential for application in other HDSS sites where conduct of similar studies may serve to help identify gaps in national or sample CR systems, and, where feasible, guide improved mortality and cause-of-death estimates;
- Predictive factors for successful matching of death records across databases were identified, which has potential to enhance the understanding of death registration practices into the CR system through identifying subgroups that are likely to be underrepresented in the CR data, and thereby providing scarce empirical evidence about factors affecting death registration, as well as pointers for corrective and advocacy action. This has implications for strategies to accelerate death registration in countries with deficient CR systems.
- Finally, the linkage provided the opportunity to compare cause-of-death data generated by the two data systems, providing an essential step in this research project that allowed the core analysis of comparing causes of death between the CRVS and HDSS systems that will be presented in the next chapter.

Chapter 5 holds this core analysis and its findings, aiming to compare causes of death as attributed in each data system, and to quantify the level of agreement of cause attribution between death certificate and verbal autopsy data in order to provide empirical evidence about the quality and utility of the official CRVS cause-of-death data in this rural area.



## 5 Chapter 5: Record-linkage as a tool for comparing cause-of-death data from civil registration and other data sources

---

### 5.1 Introduction

As outlined in Chapter 1, it is critical to know the cause-of-death profile of the population for effective public health decision-making. It is therefore essential that the individual death certificate hold an accurate and consistent description of the causes of death.

Although mortality data sources in South Africa are many and varied for measuring the level of mortality (Chapter 2), much fewer data sources are available for cause-of-death analysis, and sources for detailed, representative cause-of-death data are rare. The CRVS system is the only nationally-representative data source that offers detailed cause data, certified and coded according to international standards. However, the findings in Chapter 3 signified considerable shortcomings with these data.

Different source materials and approaches have been used to evaluate and validate cause-of-death information on death certificates in a number of countries as briefly overviewed in Section 5.2. Section 5.3 highlights that, although many of these source materials and data sources are available in South Africa, only medical records were used in validation studies before.

Section 5.4 includes the second research paper from the Agincourt HDSS/Stats SA linkage study. It focuses on the cause-of-death data as ascertained via verbal autopsy in the Agincourt HDSS and death certification in the CRVS system. In response to research question 6 below, the level of agreement of cause attribution between the data sources, misclassification patterns, and cause-specific mortality fractions are reported and discussed. To answer research question 7, the linkage study results are applied in an illustration to the cause-of-death data of the district that surrounds the Agincourt HDSS.

Research question 6: *What can the verbal autopsy data collected in the Agincourt HDSS indicate about the quality and utility of cause-of-death attribution in the civil registration system?*

Research question 7: *Are the linkage study results useful for developing adjustment factors for cause-of-death attribution at the district level?*

## **5.2 Methods for evaluating and validating cause-of-death information**

### **5.2.1 Electronic tools**

Electronic tools like CoDEdit (Cause of death editing) are useful for performing routine checks on cause-of-death data at the compilation stage to flag gross errors and minimize basic and obvious errors in the data.<sup>235</sup> ANACoD (Analysing mortality levels and causes of death) offers a more comprehensive but quick analysis of data on mortality levels and causes of death, and is useful for an automatic review for errors and assessing the plausibility of patterns emerging from the data.<sup>235</sup>

### **5.2.2 Statistical and mathematical methods**

In the absence of empirical data sources, or where empirical sources are known to hold unreliable data, statistical and mathematical methods and modelling may also assist in determining or modifying the likely cause-of-death profile in a population. Examples include the burden of disease work done in South Africa by Bradshaw *et al.*,<sup>93</sup> and the estimation of HIV/AIDS deaths by Dorrington *et al.*,<sup>236-239</sup> Groenewald *et al.*<sup>161</sup> and Birnbaum *et al.*<sup>162</sup> in their respective efforts to enhance cause-of-death information for South Africa. Modelling, however, commonly needs expert-intensive collaboration over different health focus-areas, and rely on a range of assumptions and statistical conjectures. These requirements highlight the importance of avoiding at all costs inaccurate diagnosis, certification, and coding of the cause of death.

### **5.2.3 Validation studies against a reference standard**

In contrast to electronic tools and mathematical approaches to check and model cause-of-death data, validation studies against a reference standard offer the opportunity to evaluate and improve the quality and enhance the utility of cause-of-death data. For such studies, different source materials and approaches have been used to identify mis-attribution, errors, or inconsistencies on individual death certificates, and to form an opinion of the accuracy of cause-of-death information at the population level. Source materials for comparing, evaluating or validating certified causes of death include medical records and clinical notes assessed by

clinical review committees or review panels comprising disease specialists or experts;<sup>231,240-243,244-247</sup> autopsy reports and pathology results,<sup>248-250</sup> police, coroner and medical examiner reports,<sup>251,252</sup> interviews with the certifying physician<sup>253,254</sup> or the relatives of the deceased;<sup>242,255,256</sup> health insurance claims;<sup>184,203,257</sup> verbal autopsy questionnaires;<sup>258</sup> entries into disease-specific registers;<sup>259,260</sup> participant records in follow-up studies,<sup>183,223</sup> or combinations of these.<sup>183,203,242,244,256,257,261-263</sup>

### 5.2.3.1 Pathological autopsy review

Post-mortem autopsy is generally considered the ‘gold standard’ against which the accuracy of causes of death can be ascertained. However, due to financial constraints and negative attitudes to autopsy procedures among relatives and clinicians, autopsy rates generally are low<sup>164,243,264</sup> and most countries experienced a declining autopsy rate.<sup>164</sup> Kircher *et al*,<sup>265</sup> Sington and Cottrell,<sup>248</sup> Ravakhah,<sup>266</sup> and Engel *et al*<sup>250</sup> found in their respective studies a substantial discrepancy between the diagnosis provided on death certificates compared with that during autopsy. In their meta-analysis of discrepancies between clinical and autopsy diagnoses, Roulson *et al* found little improvement in the overall rate of such discrepancies over a 40-year period since the 1960s, and argue that autopsies still are the most accurate method of verifying clinical diagnoses and death certification, and determining the cause of death.<sup>264</sup> Johansson<sup>164</sup> and Shojania<sup>267</sup> found in their reviews a large number of studies indicating that the autopsy remain an important method to improve the quality of mortality data, and have called for more autopsies<sup>164</sup> or at least the ongoing use of the autopsy.<sup>267</sup>

### 5.2.3.2 Medical record review

Given decreasing autopsy rates, and the factors limiting the use of autopsy in large samples of deaths, independent reviews of medical records, adhering to certain standard practices and protocols, can be used as a reasonable alternative for validating causes of death.<sup>47,268,269</sup> Underlying causes of death have been compared in many countries with information on medical records to assess the validity of routinely reported causes of death, mostly in developed countries.<sup>234,270</sup> In their review of the methods used in studies which assessed cause-of-death accuracy of deaths reported to official mortality statistics agencies, Johansson *et al* found that most studies, 89%, made use of medical records. Two of these studies assessed the accuracy of asthma deaths in Denmark,<sup>244,271</sup> both finding that the accuracy of death certification in asthma deaths was poor. In official statistics for Warsaw in Poland, May found that mortality from

asthma and COPD has been underestimated when compared to hospital and clinic records.<sup>272</sup> Reid and other concluded from hospital and general practice records that asthma death certification in Northern England provided a markedly inaccurate picture of asthma mortality, and it was agreed that 43% of the cases had probably never suffered from asthma.<sup>273</sup> Similarly, Smyth *et al* found in Northern Ireland that the death certificate cause of death was a relatively poor indicator of the "true" cause of death for both asthma (29% sensitivity for all deaths under 75 years in Northern Ireland in 1987) and COPD (69% sensitivity) as validated from necropsy reports, general practitioners' case notes, hospital records, and general practitioner and relative questionnaires.<sup>274</sup>

Mühlhauser and colleagues studied hospital, out-patient, and necropsy records, among other sources, of cohort-study participants with confirmed insulin-dependent Type 1 diabetes mellitus in Düsseldorf, Germany.<sup>231</sup> They found that death certificate information did not allow the identification of about one-third of confirmed diabetic cases as patients with diabetes. About 50% of the cases who were considered from the studied sources to have died of hypoglycaemia or ketoacidosis, were differently classified on the death certificate, indicating vast underreporting of Type I diabetes. Findings of unreliable diabetes diagnoses on death certificates were also reported by Rampatige *et al* in Colombo in Sri Lanka;<sup>234</sup> Andresen *et al* in King County, Washington, USA;<sup>275</sup> and Pattaraarchachai *et al* in Thailand.<sup>232</sup> Citing 15 different studies in various countries where researchers examined death certificates of patients known to have had diabetes, Andresen *et al* reported that from as few as 32% to 92% of the death certificates investigated in these studies had diabetes recorded on them.<sup>275</sup>

While Johansson and colleagues' review of different methods used in validation studies, found that 89% of studies made use of medical records, a recent literature review by Rampatige *et al* found that no standard framework exists for medical record reviews, apart from studies following similar basic steps.<sup>270</sup> In response, Rampatige *et al* developed a methodological framework that outlines a process to follow, explains the diagnostic criteria used for diagnosing the cause of death, and includes quality assessments of the medical records and hospital environment.<sup>270</sup>



### 5.2.3.3 Verbal autopsy review

In populations with inadequate CRVS systems, or where health services are not readily available to medically certify deaths, verbal autopsies have become of established value, and are critically important in helping to classify broad or leading mortality patterns.<sup>276-280</sup> This view of the importance of verbal autopsies remain despite acknowledging the limitations of deriving causes of death from a verbal autopsy approach,<sup>281,282</sup> and the overview conclusion by the Population Health Metrics Research Consortium Gold Standard Verbal Autopsy Validation Study<sup>276</sup> that physician coding of verbal autopsy cause-of-death assignment may not be as robust as previously thought.<sup>282</sup>

Causes of childhood deaths have been the focus of a number of successful verbal autopsy validation studies in Africa and Asia, for example in Namibia,<sup>283</sup> Kenya,<sup>284</sup> and Tanzania,<sup>285,286</sup> Pakistan<sup>287</sup> and India.<sup>288</sup> Similarly have verbal autopsies been used in validation studies of maternal causes, for example in rural Bangladesh,<sup>289</sup> and rural areas in each of Tanzania, Ethiopia, and Ghana.<sup>290</sup> Chandramohan *et al*, for example, aiming in their study in the latter three African countries to assess the validity of verbal autopsies, reported that verbal autopsy estimates of cause-specific maternal mortality were comparable to expected values for most maternal causes. The authors conclude that certain direct causes of hospital-based maternal mortality can be determined by means of verbal autopsies with a reasonable level of confidence.<sup>290</sup> Setel *et al*, including both children and adults in their study which aimed to validate verbal autopsy procedures against medical records in Tanzania, found that verbal autopsies produced accurate cause-specific mortality fractions for nine causes in different age groups, i.e. for pneumonia; malaria; HIV/AIDS; tuberculosis; cerebrovascular diseases; injuries; birth asphyxia; intrauterine complications; and direct maternal causes. Study results for 20 other causes approached the threshold for good performance, and the authors concluded that verbal autopsy, in all age groups, reliably estimated cause-specific mortality fractions for diseases of public health importance.<sup>286</sup>

In the South Indian city Chennai, Gajalakshmi *et al* compared the certified cause of death from the Vital Statistics Department with the probable cause of death as ascertained via a large verbal autopsy study with 48 357 adult deaths that occurred during 1995 to 1997.<sup>278</sup> Of the vital statistics-certified deaths, 37% had been attributed to unspecified natural and unknown causes, compared to 7% as ascertained by verbal autopsy. It was possible to verify cancer deaths with data from the Chennai Cancer Registry, and the validity of the verbal autopsy diagnosis was

hence assessed for cancer deaths after linking vital statistics deaths with the Registry records to confirm the deaths attributable to cancer. No false positive cancer deaths had been reported by verbal autopsy, and the sensitivity of the verbal autopsy to identify cancer deaths was 94%.<sup>278</sup> In contrast, in an earlier study, the sensitivity of the death certificate to identify cancer deaths in the same city was 57% for 1992 and 1993 deaths.<sup>291</sup>

The value of a verbal autopsy approach is shown in both hospital and out-of-hospital deaths in a cause-of-death study in Thailand<sup>292</sup> where verbal autopsy diagnoses first were validated against reference diagnoses from a medical record review for in-hospital deaths,<sup>232</sup> and the validated study findings then were used to adjust the causes of death which were derived from verbal autopsies for out-of-hospital deaths.<sup>233</sup> The authors concluded that verbal autopsy methods adapted to local use in Thailand enabled a substantial reduction of ill-defined diagnoses in registered deaths; a plausible assessment of cause-specific mortality patterns; and that routine incorporation of verbal autopsy methods could significantly improve the quality of cause-of-death data in Thailand.<sup>233,293</sup>

While cause-of-death validation ideally involves the comparison of the underlying cause of death from a particular source with a reference or 'true' underlying cause for the same death as derived from pathological autopsy or medical records, the verbal autopsy cause of death have been used as reference diagnosis where pathological autopsy or medical records were unavailable, unreliable, or done in small proportions of a population only. Using standard verbal autopsy procedures and verbal autopsy diagnoses as the reference diagnosis to evaluate the quality of registered cause-of-death statistics from the Disease Surveillance Points system in rural China, Wang *et al* found the registered cause statistics to be reasonably reliable for several leading causes of death, but substantially incorrect for other major conditions such as chronic obstructive pulmonary disease, tuberculosis, and ischaemic heart disease. The former two conditions were found significantly under-reported, and the latter significantly over-reported in the routine statistics.<sup>258</sup>

#### **5.2.3.4 Comparison with cause-of-death information from long-term follow-up studies, administrative and health databases, and disease registers**

Long-term follow-up studies, which include patient or disease monitoring till death, and cause-of-death ascertainment in clinical trials where mortality is an outcome of interest, may provide useful records and cause-of-death data to compare certified causes with.<sup>174,247</sup> In addition,

evaluations of the quality and validity of death certificate diagnoses can be undertaken by comparison to disease databases and registers. Examples of such databases and registers are the National Hospital Discharge Register in Sweden;<sup>229,230</sup> General Practitioner/General Practice Research Databases in a number of countries including Italy,<sup>294</sup> Norway,<sup>295</sup> and the United Kingdom;<sup>296</sup> and disease-specific registers, such as the Multiple Sclerosis register in France,<sup>209</sup> National Registry of Childhood Tumours in the United Kingdom,<sup>213</sup> Scottish Motor Neurone Disease Register,<sup>213</sup> All Wales Injury Surveillance System,<sup>213</sup> Cancer Registries in several regions and countries such as Europe, the United Kingdom and Ireland,<sup>297</sup> Australia and Canada,<sup>298</sup> and in Africa, Asia, the Caribbean and Central America;<sup>299</sup> the Irish Register for ALS/MND (amyotrophic lateral sclerosis/motor neuron disease);<sup>259</sup> and the cause-of-death register at the Southern Urals Biophysics Institute in Ozyorsk, Russia, documenting deaths in the Mayak workers cohort employed at the Mayak Nuclear Facility between 1948 and 1982.<sup>260</sup>

While large-scale longitudinal studies traditionally have been valuable tools to provide insight to the web of factors that affect health outcomes, including insight into the relationships among underlying, immediate, preceding and contributory causes of death, they usually are expensive, time-consuming, difficult to coordinate, and complicated by a number of factors. These factors include under-representation of minority groups, loss-to-follow-up, patient consent, and exclusion criteria.<sup>174</sup> Particularly when conducted over extended periods and multiple sites, such resource-intensive factors and complications may exclude the approach as an option, or are at least a disadvantage.<sup>174,176</sup> These factors had been observed among other in the work of Bangdiwala *et al* in the Lipid Research Clinics Program Mortality Follow-up Study;<sup>223</sup> Coady *et al* in the Atherosclerosis Risk in Communities (ARIC) Study;<sup>183</sup> Ederer *et al* in the Colorectal Cancer Death Minnesota Study;<sup>225</sup> and Moy<sup>247</sup> and Hawkins<sup>222</sup> and their colleagues in the Collaborative Ocular Melanoma Study (COMS).

#### **5.2.3.5 Record linkage**

In contrast, the advantages and increased use of record linkage in comparing and validating information from a particular source by drawing on a variety of other sources,<sup>174,175,204</sup> has been referred to in Section 4.2.4. Linkage has been used to assess the quality of, or validate, CRVS cause-of-death data by drawing on source materials, databases and registers as mentioned above. Examples and summary results of such studies are reported in Section 4.3.3.

### 5.3 Potential sources for evaluating causes of death in South Africa

In South Africa, despite urban-rural, inter-provincial, and inter-racial differences in health service delivery,<sup>71</sup> most of the source materials listed in Section 5.2.3 are routinely prepared during health service delivery to a patient, and could potentially be used in validation studies to compare, evaluate and validate certified causes of death in the CRVS system. In deaths where a pathological autopsy is required, autopsy reports are prepared and filed in the mortuaries where the autopsy was conducted. Radiology and pathology reports are prepared and kept at such facilities, and copies are sent for inclusion in patient files in hospitals and clinical practices where the patient is being attended to. For hospital patients, clinical notes by family or attending clinicians and hospital records are kept at hospitals. Upon visiting a clinician at a general or specialist practice, clinical notes are commonly made by clinicians and kept in practice files. In deaths where police investigation is involved, police reports and declarations are kept at the police station of the attending officer/s. Private health insurance companies, serving between 15% and 17% of the population between 1985 and 2010,<sup>300</sup> require detailed patient, health provider, and patient treatment information which is kept at both the service provider (for example, hospital, medical practice, or pharmacy) and the health insurance company.

Of these source materials, there is evidence in the literature of only medical records having been used before in a number of small-scale urban studies to evaluate the quality of death certificate diagnoses<sup>120,121,123,160,301</sup> (compare Chapter 3, Paper 2, page 5). Medical records have also been used in the rural study of Kahn *et al* to validate causes of death from verbal autopsies conducted in the Agincourt HDSS.<sup>302</sup>

The mortality data source review in Chapter 2 informs about a variety of local mortality data sources that contains cause-of-death information. These include censuses and surveys, which do not represent traditional ways of collecting cause-of-death data. As such, causes are only broadly indicated as natural or unnatural, pregnancy or delivery-related, and accident or violence-related causes. However, more useful cause-of-death data are available from databases, registers, and surveillance systems, similar to those mentioned in Section 5.2.3.4. These include facility-based mortality audit databases, such as the Child Healthcare Problem Identification Programme (Child PIP)<sup>303</sup>; and registers such as the National Tuberculosis Registry (NTR) and Electronic Tuberculosis Register (ETR.Net);<sup>304,305</sup> and the National Cancer Registry (NCR)<sup>306</sup> and Eastern Cape Province Cancer Registry.<sup>307</sup> Mortality

surveillance initiatives include the National Injury Mortality Surveillance System<sup>308</sup> and three small-area HDSSs co-ordinated by INDEPTH,<sup>309</sup> of which the Agincourt HDSS is the longest-running, recording vital events since 1992 in deep-rural north-east South Africa (Section 1.5). No evidence was found that any of these databases, registers and surveillance systems was used before to evaluate the quality of registered causes of death.

#### **5.4 Drawing on the Agincourt HDSS/Statistics South Africa linkage study**

The literature shows considerable agreement that data linkage has become a preferred option, where data sources and data items make it possible, to generate useful datasets in a time- and cost-efficient manner for public health and epidemiological studies.<sup>174,175</sup> Section 4.5 describes the Agincourt HDSS/Stats SA linkage study, linking death records from the Agincourt HDSS and the national CRVS system, to create a new database for use in this thesis. Both data sources collect and capture cause-of-death information for deceased individuals, including the underlying cause of death. The resultant linked data set was drawn upon for detailed cause-of-death analysis and cause-of-death comparison between the data sources.

#### **5.5 Paper 4: “Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006-09”**



Original article

## Record-linkage comparison of verbal autopsy and routine civil registration death certification in rural north-east South Africa: 2006–09

Jané Joubert,<sup>1,2\*</sup> Debbie Bradshaw,<sup>1</sup> Chodziwadziwa Kabudula,<sup>3</sup> Chalapati Rao,<sup>2</sup> Kathleen Kahn,<sup>3,4,5</sup> Paul Mee,<sup>3,4</sup> Stephen Tollman,<sup>3,4,5</sup> Alan D Lopez<sup>6</sup> and Theo Vos<sup>7</sup>

<sup>1</sup>Burden of Disease Research Unit, South African Medical Research Council, Parow Vallei, Western Cape, South Africa, <sup>2</sup>School of Population Health, The University of Queensland, Brisbane, QLD, Australia, <sup>3</sup>MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, University of the Witwatersrand, Johannesburg, South Africa, <sup>4</sup>Umeå Centre for Global Health Research, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden, <sup>5</sup>INDEPTH Network, Accra, Ghana, <sup>6</sup>Melbourne School of Population and Global Health, The University of Melbourne, Carlton, VIC, Australia and <sup>7</sup>Institute of Health Metrics and Evaluation, University of Washington, Seattle, USA

\*Corresponding author. Jané Joubert, Burden of Disease Research Unit, South African Medical Research Council, PO Box 19070, Tygerberg, 7505, South Africa. E-mail: jane.joubert@mrc.ac.za

Accepted 14 July 2014

### Abstract

**Background:** South African civil registration (CR) provides a key data source for local health decision making, and informs the levels and causes of mortality in data-lacking sub-Saharan African countries. We linked mortality data from CR and the Agincourt Health and Socio-demographic Surveillance System (Agincourt HDSS) to examine the quality of rural CR data.

**Methods:** Deterministic and probabilistic techniques were used to link death data from 2006 to 2009. Causes of death were aggregated into the WHO Mortality Tabulation List 1 and a locally relevant short list of 15 causes. The matching rate was compared with informant-reported death registration. Using the VA diagnoses as reference, misclassification patterns, sensitivity, positive predictive values and cause-specific mortality fractions (CSMFs) were calculated for the short list.

**Results:** A matching rate of 61% [95% confidence interval (CI): 59.2 to 62.3] was attained, lower than the informant-reported registration rate of 85% (CI: 83.4 to 85.8). For the 2264 matched cases, cause agreement was 15% (kappa 0.1083, CI: 0.0995 to 0.1171) for the WHO list, and 23% (kappa 0.1631, CI: 0.1511 to 0.1751) for the short list. CSMFs were significantly different for all but four (tuberculosis, cerebrovascular disease, other heart disease, and ill-defined natural) of the 15 causes evaluated.

**Conclusion:** Despite data limitations, it is feasible to link official CR and HDSS verbal autopsy data. Data linkage proved a promising method to provide empirical evidence about

the quality and utility of rural CR mortality data. Agreement of individual causes of death was low but, at the population level, careful interpretation of the CR data can assist health prioritization and planning.

**Key words:** Mortality, data quality, causes of death, vital statistics, verbal autopsy, data linkage, Agincourt Health and Demographic Surveillance System, Statistics South Africa, rural South Africa

### Key Messages

- Civil registration (CR) is a well-established official national system in South Africa, and mortality information from CR is a key data source that can guide health prioritization. However, studies continue to indicate quality problems with cause-of-death information.
- This first study, linking mortality data from CR and from the Agincourt Health and Demographic Surveillance System, demonstrates that data linkage between these sources is possible.
- The study offers the first empirical evidence of the extent and diversity of misattribution of HIV deaths in CR data from a rural setting in South Africa.
- Urban and national findings of systematic biases in CR cause-of-death data are confirmed by this rural study, pointing to the countrywide urgency to improve CR cause-of-death data.
- In the interim, however, it is encouraging that the confirmed biases can facilitate adjustment of cause profiles after careful interpretation to better inform rural health prioritization.

## Introduction

Reliable and valid mortality data are key inputs for appropriately aligning a population's health care delivery with its health care needs. However, there is a lack of such information in many low- and middle-income countries, with particular limitations in sub-Saharan Africa.<sup>1-3</sup> Data for South Africa, an upper-middle income country, were categorized by the World Health Organization (WHO) in the group of countries with unsatisfactory levels of completeness of death registration and low-quality cause-of-death information.<sup>3,4</sup> Since these assessments, a number of national initiatives have focused on improving completeness levels and quality of cause-of-death statistics from civil registration (CR).<sup>5,6</sup> South African CR has national, all-inclusive geographical and population coverage as mandated by the Births and Deaths Registration Act of 1992 and subsequent amendments. Completeness of adult death registration was estimated to be 90% in 2000,<sup>7</sup> and 89% and 78%, respectively, for infants and children under 5 years of age by 2006.<sup>8</sup> Moreover, cause-of-death information from routine death registration is produced satisfactorily by Statistics South Africa (Stats SA) in terms of timeliness and sub-national availability.<sup>9</sup> Whereas South African CR data are a key source for national health priorities,<sup>10</sup> and are used to estimate levels and causes of mortality in many sub-Saharan African countries<sup>11</sup> where there

is little mortality information elsewhere,<sup>1</sup> death registration remains incomplete,<sup>8,12</sup> and studies continue to indicate quality problems with cause-of-death information.<sup>13-16</sup> Cause quality is particularly compromised by the under-recording of HIV/AIDS as an underlying cause as found by reviewing medical records<sup>13,14,17</sup> and as indicated by examining the plausibility of national age distributions for HIV-recipient or -indicator conditions.<sup>15,16,18</sup>

In addition to the annual cause-of-death reports produced from CR by Stats SA,<sup>19</sup> cause-of-death data are collected at three South African study sites of the International Network for the Demographic Evaluation of Populations and Their Health (INDEPTH).<sup>20-22</sup> Deaths have been monitored since 1992 at the Agincourt Health and Socio-demographic Surveillance Site (Agincourt HDSS), the oldest of these sites, using a verbal autopsy (VA) tool to establish the probable cause of death.<sup>20</sup> However, no study to date has attempted to link mortality data from CR and an HDSS to assess completeness of rural death registration in the CR system, or the quality of CR cause-of-death data in a rural area in South Africa.

Linking mortality data from lay reports in VA with mortality data from medical certification in routine civil registration may pose challenges due to data confidentiality and the particular methods used to ascertain information surrounding the cause of death. In addition, in the

absence of 'gold standard' sources such as post-mortem pathological autopsies, it may not be straightforward to decide on the source of data that will best serve as a reference standard. Quality problems with CR cause-of-death information continue to be reported, as referenced above. Simultaneously, the limitations of deriving causes of death from a VA approach are acknowledged.<sup>23</sup> However, a number of studies point to plausible, valid and reliable cause-of-death results from VAs in the Agincourt HDSS. These include a local validation study comparing Agincourt HDSS VA diagnoses from 1992 to 1995 against clinician-derived diagnoses from hospital records, which found that the VA diagnoses closely approximated to that of the hospital records, with high sensitivity, specificity and positive predictive values for injuries and infectious and parasitic diseases, and reasonable accuracy for non-communicable diseases.<sup>24</sup> Other studies detected plausible time trends in HIV-related mortality from physician-certified VA diagnoses in Agincourt,<sup>25–27</sup> which correspond closely to the patterns estimated in the National Burden of Disease Studies.<sup>18,28</sup> Furthermore, for 6153 Agincourt deaths from 1992 to 2005, using two very different approaches to cause-of-death analysis and attribution (i.e. physician-assessment vs probabilistic modelling with the InterVA model) closely comparable results were found for the major causes of death in the area: over the 14-year period, both approaches present closely similar results from VAs with increasing domination by HIV-related mortality, combined with large numbers of injury deaths, and relatively low non-HIV-related infectious disease mortality. The 10 leading causes accounted for 83% and 88% of all deaths, according to physician and probabilistic interpretation, respectively, with 8 of the 10 leading causes common to both methodologies.<sup>29</sup> Another study focused on physician vs InterVA assessment for HIV-related deaths in the period 1992–2005,<sup>25</sup> and a remarkably similar development of HIV-mortality over time was reported, estimating the overall HIV-mortality rate at 18.4% and 18.6% under physician and modelled interpretations, respectively.<sup>25</sup> Moreover, considerable agreement over time in the five leading causes by age was found for the period 1992–2005.<sup>26</sup>

Against this background, we aimed to link and compare mortality data from the national CR system and the Agincourt HDSS, for the same individuals who died during the period 2006–09. Our objectives were: to quantify the level of completeness of death registration in the CR system; to compare leading causes of death from each data source; and to quantify the level of agreement of cause attribution between CR and VA data; in order to provide empirical evidence about the quality and utility of rural CR cause-of-death data.

## Methods

### Study setting and data sources

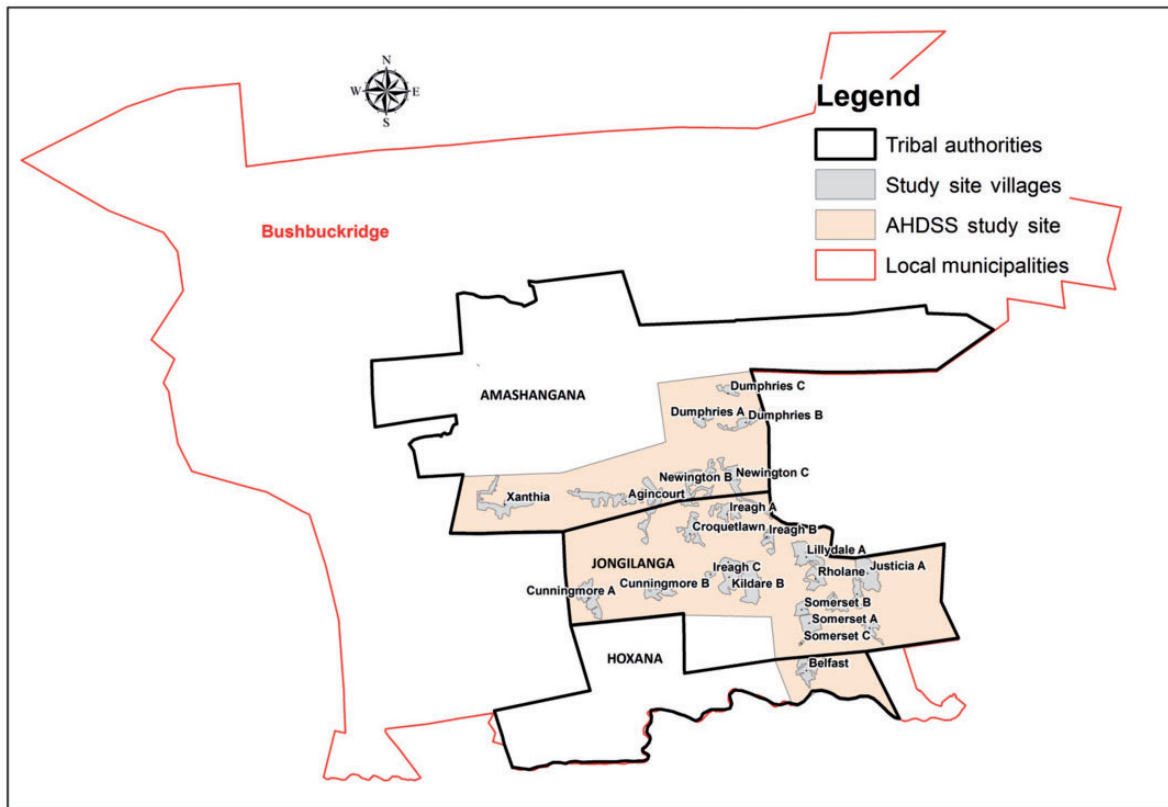
The Agincourt HDSS, located in the Bushbuckridge district of Mpumalanga province in rural north-east South Africa (Figure 1), had a population of 87 000 people in 2009. The study site covers 420 km<sup>2</sup>, comprising 27 villages with limited development infrastructure, serviced by two health centres and six primary health care clinics within the site, and three district hospitals located 25 to 60 km away.<sup>20,26</sup>

A baseline population enumeration in 1992 has been followed by annual updates of resident status and vital events. For all deaths, trained fieldworkers interview the closest carer of the deceased to elicit signs and symptoms of the illness or injury preceding death, using a locally-validated, local-language VA instrument. Two medical doctors independently review the VA information and assign probable immediate, contributory and underlying causes using ICD-10 conventions. When a consensus cause cannot be reached, a third clinician, blind to earlier findings, assesses the details. The cause is coded 'undetermined' if an agreement cannot be reached.<sup>20,24</sup>

### National civil registration system and death registration

The Births and Deaths Registration Act requires a clinician to complete the death notification form (Form BI-1663) that includes the ICD recommended format for reporting the immediate, antecedent, underlying and contributory causes of death.<sup>30</sup> For deaths in health facilities, attending or on-duty clinicians complete the form. For natural deaths at home, the deceased is taken to a morgue by undertakers who arrange for a clinician to examine the deceased and complete the form. In such instances, insufficiently available medical information about the deceased is commonly supplemented by information from relatives.<sup>31,32</sup> When a clinician is not available, as may happen in some remote rural areas, a Death Report (Form BI-1680) must be completed by an authorized traditional leader to certify the death and describe the circumstances around it.<sup>31,32</sup> Approximately 10% of deaths are certified in this way.<sup>33</sup> Unnatural deaths are subject to medico-legal investigation pursuant to the terms of the Inquests Act of 1959, and the deceased is taken to a government morgue where an autopsy is conducted. For death registration, the notification is submitted to a regional office of the Department of Home Affairs. All forms are subsequently compiled at national level, and then delivered to Stats SA where trained nosologists code causes of death to ICD-10 three-digit codes.<sup>30</sup> They then determine the underlying cause of a





**Figure 1.** Agincourt HDSS villages within the study site borders, stretching across three tribal areas: Amashangana, Jongilanga and Hoxana, within the Bushbuckridge Municipality. Source: GIS section, Agincourt HDSS, created by Paul Mee.

death using the Automated Classification of Medical Entities software (ACME 2000.05).<sup>34</sup>

### Data extraction and linkage

Following ethical undertakings regarding the confidentiality and security of data, relevant mortality and birth data for the period 1 January 2006 to 31 December 2009 were extracted from the Agincourt HDSS database and linked with CR data at the premises of Stats SA. Eleven common variables were used for matching these death records: national identity number (a unique 13-digit number assigned to South African citizens), surname, sex, day of birth, month of birth, year of birth, day of death, month of death, year of death, village name and institution/venue where the death took place. Information for individuals who were either born, resident or died in the Bushbuckridge Municipality was extracted from Stats SA's CR database. From this pool, records were extracted if the deceased was either born, resided or died in one of the tribal areas in which the Agincourt HDSS is located, i.e. Amashangana, Jongilanga and Hoxana (Figure 1).

Deterministic and probabilistic record linkage approaches were applied to find matches using routines implemented in T-SQL, the proprietary implementation of

the SQL standard in the SQL\*Server™ software package.<sup>35</sup> An anonymized, de-identified data set was created, including an indicator for matching status. The linkage methodology has been detailed elsewhere.<sup>36</sup>

### Data analysis

Stata 12<sup>37</sup> and Microsoft Excel 2010 were used. The matching rate was calculated using the proportion of total Agincourt HDSS records that could be matched to CR records. The rate was compared with the proportion of HDSS records in which it was reported that the death had been registered in the CR system.

Causes from both sources were aggregated into the 103 causes of the WHO Mortality Tabulation List 1 (WHO list),<sup>38</sup> used in similar studies elsewhere.<sup>39,40</sup> Causes not coded according to standard ICD conventions, were recoded: one CR case assigned U51 (extensively drug-resistant tuberculosis) was recoded to A16 (respiratory tuberculosis); 10 VA cases assigned S and T codes (certain consequences of external causes) to Y34 (undetermined injury); and four VA cases assigned Z codes (factors influencing health status and contact with health services) to R99 (other ill-defined and unspecified natural causes).

**Table 1.** WHO Mortality Tabulation List 1 and ICD-10 codes for the short list of causes of death

Short list conditions	WHO Mortality Tabulation List 1 codes	ICD-10 codes
1 Diarrhoea	1-003	A09
2 Tuberculosis	1-005, 1-006	A15-A16
3 HIV disease	1-020	B20-B24
4 Remaining infectious & parasitic disease	1-002, 1-004, 1-007 to 1-010, 1-012 to 1-025	All remaining Ch A & B codes
5 Neoplasms	1-026	C00-D48
6 Diabetes	1-052	E10-E14
7 Meningitis	1-011, 1-059	A39, G00, G03
8 Hypertensive disease	1-066	I10-I14
9 Remaining heart disease	1-065, 1-067, 1-068	I00-I09, I20-I25, I26-I51
10 Cerebrovascular disease	1-069	I60-I69
11 Acute lower respiratory infections	1-073 to 1-075	J10-J11, J12-J18, J20-J22
12 Other respiratory disease	1-076 to 1-077	J40-J47, J30-J39, J60-J98
13 Symptoms & ill-defined conditions	1-094	R00-R99
14 External causes	1-095	V01-Y89
15 Remaining natural causes	All remaining codes	All remaining codes

To make the analyses more relevant to the local mortality burden, causes were further aggregated into a short list with 15 causes/cause groups (Table 1). Considering the use of pseudonyms for HIV deaths as reported in previous studies,<sup>13–15</sup> investigating the age pattern of these deaths and examining the immediate, antecedent and contributory causes where available, 130 CR cases from ICD-10 codes B33 (other viral diseases), B45 (cryptococcosis), B59 (pneumocystosis), C46 (Kaposi sarcoma) and D84 (other immunodeficiencies) were recoded to B24 (HIV-disease). Additionally, based on previous findings,<sup>15,16</sup> clinical advice, careful examination of age patterns and taking into account other causes attributed to these deaths, 21 CR cases from E86 (volume depletion) and E87 (other disorders of fluid, electrolyte and acid-base balance) were recoded to R99; 32 CR cases from K52 (other non-infective gastroenteritis and colitis) to A09 (other gastroenteritis and colitis of infectious and unspecified origin); and one CR and two VA cases from A39 (meningococcal infection) were recoded to G03 (meningitis).

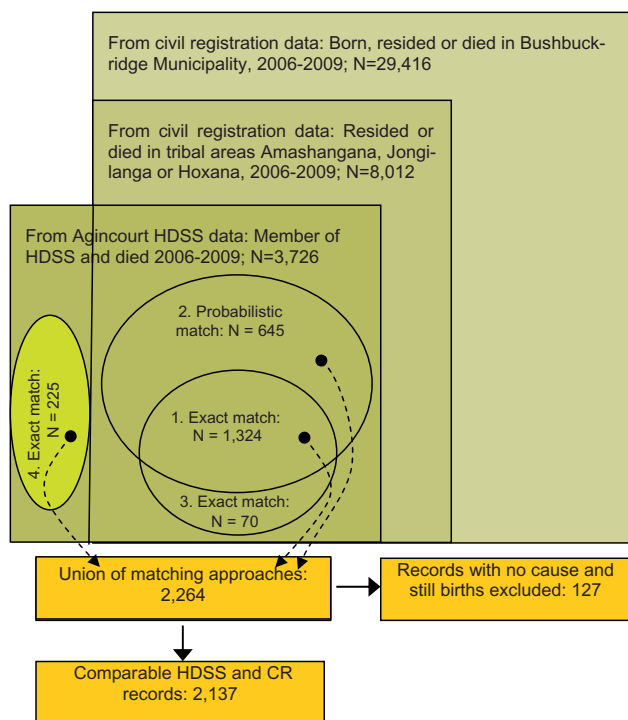
We could not make use of ‘gold standard’ sources such as post-mortem pathological autopsy or expert review of hospital records against which to validate causes of death, as only 2% of CR deaths were autopsy ascertained, and less than half occurred in hospital. Instead, as done in a study in rural China,<sup>41</sup> we used physician-certified VA causes of death as the reference diagnoses against which to examine the plausibility of the CR diagnoses. In spite of acknowledging the limitations of deriving causes of death from a VA approach,<sup>23,42</sup> our decision was informed by the studies referred to in the Introduction,<sup>18,24–29</sup> which lend support to the quality of Agincourt VA cause-of-death

data and strengthen our confidence in VA causes of death as the reference diagnoses. As a number of studies continue to report quality problems in CR cause data, with a substantial problem of under-reporting HIV/AIDS,<sup>13–17,43,44</sup> the VA diagnoses here were hence used as reference values in our study to assess the quality of the CR data.

Agreement of cause attribution was assessed with the kappa statistic and 95% confidence intervals (CIs) using the WHO List and short list. Misclassification patterns were identified by cross-tabulating the data using the short list. The sensitivity of the CR diagnoses and their positive predictive values (PPV) were calculated with 95% CIs. CSMFs were calculated as the proportion of total deaths attributable to specific conditions in each data set. The differences between CR- and VA-based fractions were expressed as percentages of the CR-based fractions, and 95% CIs were calculated using Nam and Blackwelder’s method.<sup>45</sup>

## Ethics

Ethics clearance for research involving human participants was obtained from the University of Queensland’s School of Population Health Research Ethics Committee (approval no. JJ010911), the South African Medical Research Council Ethics Committee (EC008-6/2011) and the Human Research Ethics Committee (Medical) at the University of the Witwatersrand (M120106). Ethical clearance for the collection of Agincourt HDSS and VA data was given by the University of the Witwatersrand Human Research Ethics Committee (Medical), clearance certificates M960720 and M110138.



**Figure 2.** Establishing the study population: data sources, number of deaths and number of comparable records, Agincourt HDSS, 2006–09.

## Results

### Data linkage

The complete CR data file with 4 years' data contained 2 464 915 death records nationally. Of these, 29 416 records were found for individuals whose place of birth or residence or death was recorded as within the Bushbuckridge Municipality. These included 8012 records that had place of birth or residence or death recorded as within one of the three tribal areas in which the study site is located (Figure 2). The Agincourt HDSS data file contained 3726 death records of individuals who were residents of a household within the study site and died within the reference period. Deterministic and probabilistic record linkage approaches independently identified 1394 and 1969 matches, respectively, among records in the Agincourt HDSS and Stats SA's CR data files where place of birth or residence or death in the CR data file was recorded as within the Bushbuckridge borders. Of the records that were matched with the deterministic approach, 1324 (95.0%) were also matched using the probabilistic approach. An additional 225 records were matched deterministically following careful examination of their variables and subsequent corrections to place names in the CR data, resulting in a total of 2264 matched cases (Figure 2). Over half of the deterministic matches (54%) were found via the deceased's identity number. After removing

**Table 2.** Matching rate and informant-reported level of death registration in the study population

Age	VA-CR matching rate			VA informant-reported registration		
	%	95% CI		%	95% CI	
All ages	60.8	59.2	62.3	84.6	83.4	85.8
<5 years	38.4	34.3	42.4	33.8	29.7	37.9
5+ years	64.7	63.0	66.3	93.3	92.4	94.2

22 records of stillbirths and 105 with no VA cause of death recorded, 2137 records were available for assessing cause-of-death agreement.

Of all deaths recorded in the Agincourt HDSS database, 61% were matched to a death registered in the CR system. In contrast, for 85% of the Agincourt HDSS deaths, the household informant reported that the death had been registered at the Department of Home Affairs, the entity that administers the CR system. In both scenarios, the rate for deaths under the age of 5 years was substantially lower than that for persons aged over 5 years (Table 2).

### Characteristics of the study populations

The characteristics of the matched cases were fairly similar to those that did not match (Table 3), except that the unmatched compared with matched Agincourt HDSS cases had a higher proportion of child deaths under 5 years (23% vs 9%), and a higher proportion of cases with unspecified ascertainment of the cause of death (8% vs 5%). The age groups 25–34 and 35–44 years accounted for the largest proportions of deaths in both data sets.

### Leading underlying causes

Table 4 presents the 10 leading causes of death, according to the WHO List, prior to recoding and aggregating the causes into the short list. The VA system identified HIV disease as the leading cause of death (31%), in contrast to its 21st rank in the CR system (1.2%) (not presented). Diarrhoea and pneumonia are among the leading causes in both sources, but account for considerably higher proportions in the CR list. Injuries appear in both lists.

### Cause agreement and misclassification

At the WHO list level, agreement of cause attribution between the VA and CR data was 15.1% (322/2137), yielding a kappa score of 0.1083 (95% CI: 0.0995 to 0.1171). At the much more aggregated short list level, agreement

**Table 3.** Characteristics of the study populations, Agincourt HDSS and CR data, 2006–09

Characteristics	Agincourt HDSS matched		Agincourt HDSS unmatched		CR matched		CR unmatched	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Age at death, years:								
<1	134	5.9	228	15.6	126	5.6	318	5.5
1–4	79	3.5	114	7.8	74	3.3	204	3.6
5–14	67	3.0	39	2.7	62	2.7	153	2.7
15–24	120	5.3	67	4.6	120	5.3	344	6.0
25–34	443	19.6	228	15.6	468	20.7	985	17.1
35–44	418	18.5	221	15.1	417	18.4	1045	18.2
45–54	284	12.5	169	11.6	271	12.0	760	13.2
55–64	229	10.1	125	8.6	227	10.0	551	9.6
65–74	203	9.0	95	6.5	207	9.1	526	9.2
75–84	171	7.6	109	7.5	183	8.1	544	9.5
85+	116	5.1	67	4.6	109	4.8	310	5.4
Unspecified	0	0.0	0	0.0	0	0.0	8	0.1
Total	2264	100.0	1462	100.0	2264	100.0	5748	100.0
Sex:								
Male	1160	51.2	795	54.4	1160	51.2	2774	48.3
Female	1104	48.8	667	45.6	1104	48.8	2964	51.6
Unspecified	0	0.0	0	0.0	0	0.0	10	0.2
Total	2264	100.0	1462	100.0	2264	100.0	5748	100.0
Venue of death:								
Hospital	1117	49.3	642	43.9	1058	46.7	2094	36.4
Health care centre	25	1.1	17	1.1	–	–	–	–
Clinic	19	0.8	7	0.4	–	–	–	–
Emergency room/outpatient unit	–	–	–	–	37	1.6	95	1.7
Dead on arrival	–	–	–	–	9	0.4	52	0.9
Nursing home	–	–	–	–	6	0.3	30	0.5
Home	937	41.4	588	40.2	970	42.8	2784	48.4
Vehicle accident site	59	2.6	43	2.9	–	–	–	–
Other	98	4.3	157	10.7	50	2.2	206	3.6
Unknown/unassigned/unspecified	9	0.4	8	0.5	134	5.9	487	8.5
Total	2264	100.0	1462	100.0	2264	100.0	5748	100.0
Ascertainment of cause via:								
Verbal autopsy	2157	95.3	1346	92.1	–	–	–	–
Forensic autopsy	–	–	–	–	55	2.4	116	2.0
Opinion of attending doctor	–	–	–	–	717	31.7	1547	26.9
Opinion of attending doctor on duty	–	–	–	–	454	20.1	752	13.1
Opinion of registered nurse	–	–	–	–	5	0.2	4	0.1
Interview with family member	–	–	–	–	617	27.3	1814	31.6
Other	–	–	–	–	5	0.2	19	0.3
Unknown	–	–	–	–	28	1.2	55	1.0
Unspecified	107	4.7	116	7.9	383	16.9	1441	25.1
Total	2264	100.0	1462	100.0	2264	100.0	5748	100.0

only increased to 23.2% (496/2137), kappa 0.1631 (CI: 0.1511 to 0.1751), or less than one in four cases.

Table 5 shows the misclassification patterns using the short list causes. Using the VA cause as reference diagnosis, the sensitivity of the CR system to identify external causes was relatively high (67%, 95% CI: 58.8 to 74.8), but considerably lower for natural causes. A relatively high PPV

(78%, 95% CI: 69.7 to 85.0) was calculated for external causes, but noticeably lower values for natural causes.

The CR data show considerable misclassification of HIV disease. Of 672 VA deaths attributed to HIV disease (B20–B24), only 11% were assigned B20–B24 in the CR data (Table 5), and the remainder to 73 other ICD-10 codes. The most frequent single recipient CR causes were

**Table 4.** Ten leading causes of deaths from VA and CR, according to the WHO List: Agincourt HDSS, 2006–09 (N = 2137)

Agincourt HDSS Verbal Autopsy (VA), according to WHO List			Civil Registration (CR), according to WHO List		
Rank	Ten leading causes	% of total	Rank	Ten leading causes	% of total
1	HIV disease (B20-B24)	31.4	1	Respiratory tuberculosis (A15-A16)	16.0
2	Respiratory tuberculosis (A15-A16)	14.8	2	Diarrhoea/gastro, infectious (A09)	15.6
3	Pneumonia (J12-J18)	5.2	3	Pneumonia (J12-J18)	10.7
4	Cerebrovascular diseases (I60-I69)	4.6	4	Cerebrovascular diseases (I60-I69)	5.2
5	Septicaemia (A40-A41) <sup>a</sup>	4.5	5	Other blood & immune disorders (D65-D89)	4.5
6	Meningitis (G00,G03)	3.8	6	All other external causes <sup>b</sup>	3.8
7	Diarrhoea/gastro, infectious (A09)	3.2	7	Other heart diseases (I26-I51)	3.6
8	Other heart diseases (I26-I51)	3.0	8	Hypertensive diseases (I10-I14)	3.0
9	Symptoms & ill-defined (R00-R99)	3.0	9	Other acute lower resp. infections (J20-J22)	3.0
10	Transport accidents (V01-V99)	2.4	10	Symptoms & ill-defined (R00-R99)	2.7
Top 10 causes as % of total deaths		76.1	Top 10 causes as % of total deaths		68.0

<sup>a</sup>Septicaemia is an intermediate cause of death, not an underlying cause of death. Although it is a clearly defined clinical entity, septicaemia has an underlying cause that would have precipitated the chain of events which led to death.<sup>46</sup>

<sup>b</sup>W20-W64, W75-W99, X10-X39, X50-X59, Y10-Y89.

**Table 5.** Misclassification patterns for selected causes/cause groups in the Agincourt HDSS study site, 2006–09

	Civil Registration diagnoses															Verbal Autopsy diagnoses		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	CR total	Positive predictive value %	95% Confidence Interval
1 Diarrhoea	20	47	163	23	16	3	10	3	7	10	21	3	10	8	21	365	5.5	3.4 to 8.3
2 Tuberculosis	4	124	139	21	9	2	9	1	7	4	15	0	11	11	11	368	33.7	28.9 to 38.8
3 HIV disease	2	36	73	6	4	0	9	2	1	3	8	0	1	2	9	156	46.8	38.8 to 54.9
4 Remaining infect. & parasitic disease	4	6	17	13	3	0	3	1	1	1	3	0	2	5	6	65	20.0	11.1 to 31.8
5 Neoplasms	1	4	10	10	28	0	0	4	3	1	0	3	1	7	7	72	38.9	27.6 to 51.1
6 Diabetes	3	4	4	2	1	15	0	1	2	7	1	1	1	0	3	45	33.3	20.0 to 49.0
7 Meningitis	2	4	21	4	0	2	15	1	2	1	0	1	1	0	2	56	26.8	15.8 to 40.3
8 Hypertensive disease	7	4	7	6	6	2	4	4	9	4	7	1	1	0	2	64	6.3	1.7 to 15.2
9 Remaining heart disease	4	10	18	7	4	3	0	4	10	8	6	0	4	1	5	84	11.9	5.9 to 20.8
10 Cerebrovascular disease	2	3	10	14	5	3	5	10	4	34	10	1	7	0	3	111	30.6	22.2 to 40.1
11 Acute lower respiratory infections	10	44	118	20	9	2	14	7	7	8	21	1	10	7	15	293	7.2	4.5 to 10.7
12 Other respiratory disease	0	6	31	6	1	1	2	4	1	4	3	4	1	2	6	72	5.6	1.5 to 13.6
13 Symptoms & ill-defined conditions	4	13	15	11	4	1	5	0	3	4	8	0	4	3	4	79	5.1	1.4 to 12.5
14 External causes	1	2	2	1	1	0	4	0	2	2	0	1	7	96	4	123	78.0	69.7 to 85.0
15 Remaining natural causes	5	22	44	15	7	2	4	5	15	6	10	2	5	7	35	184	19.0	13.6 to 25.4
VA total	69	329	672	159	98	36	84	43	75	99	114	15	68	143	133	2137		
Sensitivity %	29.0	37.7	10.9	8.2	28.6	41.7	17.9	9.3	13.3	34.3	18.4	26.7	5.9	67.1	26.3			
95% CI lower level	18.7	32.4	8.6	4.4	19.9	25.5	10.4	2.6	6.6	25.1	11.8	7.8	1.6	58.8	19.1			
95% CI upper level	41.2	43.2	13.5	13.6	38.6	59.2	27.7	22.1	23.2	44.6	26.8	55.1	14.4	74.8	34.7			

diarrhoea (24%), tuberculosis (20%) and pneumonia (13%), but heart failure (1.5%) and stroke (1.5%) were among the 10 leading recipient causes too. Of all recipient conditions, 21% were unexpected, including chronic respiratory diseases, peptic ulcers, diabetes, cardiac arrest and essential hypertension. Further, Table 5 suggests

higher reporting of diabetes coupled with lower reporting of cardiovascular disease (CVD) deaths in the CR compared with VA diagnoses.

Table 6 shows the agreement characteristics and percent change in CSMFs if the underlying cause from the VA diagnosis replaced that from the CR diagnosis. The largest

**Table 6.** Agreement characteristics of civil registration and verbal autopsy diagnoses for the short list causes/cause groups: Agincourt HDSS, 2006–09

Conditions	Total occurrences of the cause in CR system	On VA instrument			Total occurrences of the cause in VA system	Cause-specific mortality fraction in CR data	Cause-specific mortality fraction in VA data	Per cent difference in cause-specific mortality fraction	95% CI
		Cause confirmed on VA instrument	Cause assigned to other causes in VA system	Received from other causes in VA system					
1 Diarrhoea	365	20	345	49	69	17.1	3.2	-81	-85.2 to -75.9 <sup>a</sup>
2 Tuberculosis	368	124	244	205	329	17.2	15.4	-11	-20.7 to 0.7
3 HIV disease	156	73	83	599	672	7.3	31.4	331	268.2 to 405.0 <sup>a</sup>
4 Remaining infectious & parasitic disease	65	13	52	146	159	3.0	7.4	145	86.8 to 221.0 <sup>a</sup>
5 Neoplasms	72	28	44	70	98	3.4	4.6	36	6.2 to 75.0 <sup>a</sup>
6 Diabetes	45	15	30	21	36	2.1	1.7	-20	-43.8 to -13.3 <sup>a</sup>
7 Meningitis	56	15	41	69	84	2.6	3.9	50	11.3 to 102.7 <sup>a</sup>
8 Hypertensive disease	64	4	60	39	43	3.0	2.0	-33	-53.6 to -2.8 <sup>a</sup>
9 Remaining heart disease	84	10	74	65	75	3.9	3.5	-11	-33.2 to 19.4
10 Cerebrovascular disease	111	34	77	65	99	5.2	4.6	-11	-28.7 to 14.2
11 Acute lower respiratory infections	293	21	272	93	114	13.7	5.3	-61	-68.3 to -52.3 <sup>a</sup>
12 Other respiratory disease	72	4	68	11	15	3.4	0.7	-79	-87.7 to -65.0 <sup>a</sup>
13 Symptoms & ill-defined conditions	79	4	75	64	68	3.7	3.2	-14	-37.1 to 17.8
14 External causes	123	96	27	47	143	5.8	6.7	16	2.4 to 32.6 <sup>a</sup>
15 Remaining natural causes	184	35	149	98	133	8.6	6.2	-28	-40.7 to -12.0 <sup>a</sup>
Total	2137	496	1641	1641	2137	100.0	100.0		

<sup>a</sup>The confidence interval indicates significant changes in the cause-specific mortality fractions ( $P < 0.05$ ).

**Table 7.** External cause from VA for under-specified CR injury deaths, Agincourt HDSS 2006–09

Assigned in the Agincourt VA system		ICD-10 causes/cause groups assigned in CR system				VA total
ICD-10 cause category aggregated from single codes assigned	ICD-10 category code	Event of undetermined intent: Y10–Y34	Exposure to unspecified factor: X59	Motor- or non-motor vehicle accident, vehicle unspecified: V89	Other accidental hanging and strangulation: W76	
Pedestrian injured in transport accident	V01 – V09	2	3	5	–	10
Car occupant injured in transport accident	V40 – V49	4	7	14	–	25
Occupant of pick-up truck or van injured in transport accident	V50 – V59	–	–	1	–	1
Other land transport accidents	V80 – V89	1	–	1	–	2
Falls	W00 – W19	–	–	1	–	1
Exposure to inanimate mechanical forces	W20 – W49	–	1	–	–	1
Exposure to smoke, fire and flames	X00 – X09	1	–	–	–	1
Exposure to forces of nature	X30 – X39	–	1	–	–	1
Intentional self-harm	X60 – X84	2	2	–	9	13
Assault	X85 – Y09	6	8	–	1	15
CR total		16	22	22	10	70

difference is observed for HIV, showing an expected 331% change. Significant differences are observed for most conditions, with only four conditions (tuberculosis, remaining heart disease, cerebrovascular disease and symptoms and ill-defined conditions) not significantly different.

## Injuries

Despite the relatively high sensitivity and PPV for external causes as a broad group, detailed analysis revealed very low sensitivity of the CR to ascertain suicide (0%; 0/23) and homicide (17%; 6/35). Ten deaths assigned unintentional injury codes in the CR system were attributed to intentional codes for suicide in the VA system. In the CR data, 76% of injuries (94/123) were assigned to 'other', 'unspecified' and undetermined-intent injury codes (V89, W76, X59, Y10–Y34). Most deaths in these groupings were assigned to more specific ICD-10 codes in the VA system (Table 7).

## Discussion

Cause-of-death agreement between the data sources was low, whether aggregated by the WHO (15.1%) or the short list (23.2%). Our study, and findings from urban<sup>13,14,17</sup> and national<sup>15,16,47</sup> studies, highlight systematic biases in CR cause-of-death data in South Africa. Similar to other research that indicates high proportions (73–92%) of HIV deaths being misattributed in the CR

data,<sup>14,16,28</sup> our study indicates misattribution in 89% of HIV deaths in the CR system. Yudkin *et al.*,<sup>14</sup> for example, found tuberculosis, diarrhoea, lower respiratory and other respiratory infections, parasitic diseases, intestinal infectious diseases, meningitis, other infectious conditions, digestive disorders and ill-defined conditions as common HIV-recipient causes in their Cape Town study, comparing causes of death from CR and hospital records. Examining cause-of-death data from 1996–2001 CR records, Groenewald *et al.*<sup>15</sup> found that the following conditions increased in the same distinct age pattern as HIV/AIDS and concluded that they could be considered misattributed HIV/AIDS deaths: tuberculosis, pneumonia, diarrhoea, meningitis, other respiratory disease, non-infective gastroenteritis, other infectious and parasitic diseases, deficiency anaemia and protein energy malnutrition. These findings are resonated in a post-mortem autopsy study in a tertiary hospital in rural Eastern Cape province of South Africa, showing that the leading three causes of death among HIV-positive people were tuberculosis, pneumonia and meningitis.<sup>48</sup> Similarly, acute respiratory infections and tuberculosis were among the most frequent causes of death assigned to deaths among HIV-infected persons in a study pooling data from six sites of the Alpha Network in Africa.<sup>49</sup> Such misattribution, and the resultant underreporting of deaths from HIV/AIDS, have been attributed to cause-of-death coding practices, legal issues regarding life and health insurance, concerns regarding the confidentiality of death certificates, the fear of HIV-positive

stigmatization, and clinicians avoiding HIV/AIDS on the death certificate because of the potential for harm.<sup>14–16,50–52</sup> Whereas misattribution of HIV to infectious conditions such as diarrhoea, tuberculosis and pneumonia has been acknowledged in other studies<sup>14–16</sup> and account for the majority HIV-recipient conditions in our study, our findings highlight that HIV deaths were also misattributed to selected circulatory diseases, neoplasms and digestive, endocrine and metabolic disorders. Previously-observed misclassification patterns between diabetes and CVD are also confirmed.<sup>13,47</sup> These misclassifications and biases indicate that the CR data cannot be taken at face value and caution should be exercised in its use for research and health decision-making.

Compared with natural causes of death, better agreement was observed for external causes as a group. However, detailed cause analysis indicates that the CR diagnoses missed the majority of homicide and all suicide cases, and that the external causes for three-quarters of injury deaths remain undetermined. This is a consequence of the death notification form (DNF) not including a field for intent of injury deaths (homicide/suicide/accident/unknown). The manner of death is consequently often missing and results in the statistics having limited capacity to guide injury prevention or safety promotion interventions.

Surprisingly, the proportion of causes coded to non-specific and ill-defined causes for this area was low (3–4%, compared with about 14% nationally for the same period).<sup>31</sup> R-codes in the relevant province, Mpumalanga, were also low (9%) in the period 1997–2007, compared with neighbouring Gauteng (12%), KwaZulu-Natal (15%) and Limpopo (18%).<sup>9</sup> One possible explanation is that headmen do not certify deaths in the Agincourt study site, but further research into local certification practices is required to explain the phenomenon.

Considering the cause-of-death profile from both data sources, and acknowledging misattribution and underreporting of HIV/AIDS, it is clear that HIV/TB is a major concern in the area. In addition, the data suggest the emergence of cardiovascular disease as an epidemiological concern. Though not a vote of confidence in the CR cause-of-death certification system, this broad similarity suggests that, despite low agreement of cause attribution at the individual level, there is scope to carefully interpret and adjust CR data to identify plausible broad epidemiological patterns at the population level in rural areas, to prioritize health care needs and inform public health policy. It must be noted that the CR data, however, could be misleading with respect to specific causes.

It was not possible to apply the capture-recapture method to calculate CR completeness of death registration because the study site boundaries differed from the official

boundaries used in CR data. This resulted in challenges to identify whether a CR death occurred within the HDSS borders. Further challenges arose from the fact that a tribal area name instead of the actual village name was sometimes recorded on the DNF. Each tribal area spans an area larger than the village areas and, combined, the tribal areas span an area larger than the study site (Figure 1). Additionally, the study identified some differences in the place names of villages used by the Agincourt HDSS and StatsSA, indicating an inability to validate the colloquial name against the official CR name. Valuable lessons were learned about the need for accurate place name reporting during death reporting, place name consistency across data sources, and the alignment of study-site and official boundaries.

Although we could not calculate completeness of death registration using a standard capture-recapture approach, the matching rate of 61% could be taken as a minimum indication of completeness. However, the matching rate was adversely affected by various limiting factors, including few common variables to work with and proxy reporting of vital-event information.<sup>36,53</sup> Given these limiting factors, it is likely that completeness was higher than the matching rate. This likelihood is supported by the VA-informant responses indicating that 85% of deaths were registered into the CR system. As socially desirable answers and recall limitations both are acknowledged sources of bias that may affect informant responses, further research is needed to better inform our understanding of completeness in the area.

Accurate cause-of-death attribution, including that for HIV/AIDS, is needed in order for CR to meet its potential to directly inform decision-making in South Africa and contribute to regional estimates. There is an urgent need to improve the quality of CR cause-of-death information, with renewed and innovative efforts. It is encouraging that Stats SA, together with the Department of Health, Department of Home Affairs and Medical Research Council have set up a training initiative to improve the quality of cause-of-death certification countrywide.<sup>51</sup> In addition, an independent, systematic, scientific effort, such as a national burden of disease study utilizing all mortality data sources to identify data problems and adjust for biases, will provide valuable estimates. However, there is a need to consider further actions to address the biases. A nationally representative validation study of CR cause data against 'gold standard' instruments, such as post-mortem autopsy reports or high-quality laboratory, hospital and other medical records, would likely boost confidence in the country's CR data. Improving the quality of medical records, however, is likely a prerequisite.<sup>24,44</sup> For substantial improvements to injury data to be made, the Department of Home Affairs is urged to include a field for intent, or apparent manner of death, in a revised DNF, to ensure that



details about external causes, needed for prevention efforts, can be reported. Finally, to address the gap arising from poor cause specification for home deaths, we suggest that registration of these deaths in the CR system be augmented by the systematic collection of VAs to support the current practices for physician certification at morgues. Further, systematic VAs could also be used in those deaths certified by village headmen. The human capacity for such an initiative is potentially available within the community health worker programme being established as part of primary health care re-engineering. This, and harnessing the many years of experience and strengths from the country's three HDSSs, would provide a critical mass of human resources and training capacity to strengthen cause-of-death ascertainment through the use of VAs for out-of-hospital deaths. The WHO short-form VA questionnaire<sup>54</sup> and promising automated methods<sup>49,55–58</sup> will greatly facilitate the application of VA in routine CR systems for diagnosing out-of-hospital deaths, including in South Africa.

This first study linking national CR and HDSS data demonstrates that data linkage between these sources is possible. This offers a promising method to provide empirical evidence about the quality and utility of rural CR mortality data, and show how matching can be used to better understand, complement and improve the quality of CR mortality data. In addition, the study offers the first empirical evidence of the extent and diversity of misattribution of HIV deaths in a rural setting. Previously, such inferences were based on indirect evidence such as the implausibility of age distributions for HIV-recipient/indicator conditions. The Agincourt HDSS data hold considerable value in providing much-needed detail on external causes of injury deaths to better inform programmes aimed at reducing fatal injuries. Urban and national findings of systematic biases are confirmed by this rural study, pointing to the countrywide urgency to improve CR cause-of-death data. In the interim, however, it is encouraging that the confirmed biases can facilitate adjustment of cause profiles after careful interpretation, to better inform rural health prioritization and planning.

## Funding

This work was supported by: the Medical Research Council/Wits University/Rural Public Health and Health Transitions Research Unit (Agincourt); the South African Medical Research Council; and Statistics South Africa. The research was carried out while the first author held a University of Queensland Research Scholarship and the Endeavour International Postgraduate Research Scholarship at the University of Queensland, Brisbane, Australia. The Agincourt HDSS was funded by: the Wellcome Trust, UK (grants no. 058893/Z/99/A, 069683/Z/02/Z, 085477/Z/08/Z); the National Institute on Ageing of the NIH (grants 1R24AG032112-01 and 5R24AG032112-03); the William and Flora Hewlett Foundation

(grant 2008-1840); the Andrew W Mellon Foundation, USA; and the University of the Witwatersrand and Medical Research Council, South Africa. The funders had no role in study design, data collection, analysis, decision to publish or preparation of the manuscript.

## Author contributions

A.D.L., C.R., D.B., T.V. and J.J. conceptualized the study. J.J. wrote the study proposal and ethics applications; coordinated the collaboration, data preparation and matching exercise; conceptualized the paper; analysed and interpreted the data; created the tables and graphs; wrote the first draft of all sections of the paper; integrated inputs from co-authors; submitted the paper; and led the response to reviewers' comments. K.K. was instrumental in the ethics application process at the University of the Witwatersrand. A.L., D.B., K.K., S.T. and T.V. made substantial contributions to negotiating access to and procuring the data. C.K. extracted the Agincourt HDSS data with inputs from K.K., P.M. and S.T. C.K. did the electronic matching and created the base analytic data set with inputs from D.B., J.J. and T.V. C.R. revised the manuscript critically for structure and word economy, critically appraised the data visualizations and reviewed the text for important intellectual content. A.L., C.R., D.B. and T.V. consistently supplied detailed comments during all phases of the study and critically appraised decisions regarding the methods, findings and interpretation of the results. All authors made contributions to the design of the study, interpretation of the data and critical review of the final manuscript, and approved the final version to be published.

## Acknowledgements

We gratefully acknowledge Statistics South Africa and the Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt) for making available mortality data, and Statistics South Africa for hosting the matching exercise. We acknowledge the important contribution of the Agincourt HDSS fieldworkers and other staff members responsible for the collection and custodianship of the demographic and mortality data used in this study. We thank: Maletela Tuoane-Nkhasi for leading the data procurement, extraction and security arrangements at Statistics South Africa and overseeing the matching exercise; Ramadimetja Matji, Aletia Barkley and Kerotse Mmatli for their participation in meetings and contributions to securing the data prior to the matching exercise; Marlanie Moodley for preparing maps of the Agincourt and Bushbuckridge areas; Ria Laubscher for technical assistance during the matching exercise; Carl Lombard for statistical assistance; and Gail Williams and Peter Baker for statistical advice.

**Conflict of interest:** None declared.

## References

1. Wang H, Dwyer-Lindgren L, Lofgren KT *et al.* Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic

- analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**:2071–94.
2. Lozano R, Naghavi M, Foreman K *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;**380**:2095–128.
  3. Mahapatra P, Shibuya K, Lopez AD *et al.* Civil registration systems and vital statistics: successes and missed opportunities. *Lancet* 2007;**370**:1653–63.
  4. Mathers CD, Ma Fat D, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull World Health Organ* 2005;**83**:171–77C.
  5. Bradshaw D, Kielkowski D, Sitas F. New birth and death registration forms – A foundation for the future, a challenge for health workers? *S Afr Med J* 1998;**88**:971–74.
  6. Bah S. Multiple forces working in unison: the case of rapid improvement of vital statistics in South Africa post-1996. *World Health Popul* 2009;**11**:50–59.
  7. Dorrington RE, Bourne D, Bradshaw D, Laubscher R, Timaeus IM. *The Impact of HIV/AIDS on Adult Mortality in South Africa*. Tygerberg, South Africa: South African Medical Research Council, 2001.
  8. Darikwa TB, Dorrington RE. The level and trends of child mortality in South Africa, 1996–2006. *J Afr Popul Stud* 2011;**25**:158–72.
  9. Joubert J, Rao C, Bradshaw D, Vos T, Lopez AD. Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007. *PLoS One* 2013;**8**:e64592.
  10. Bradshaw D, Nannan N, Groenewald P *et al.* Provincial mortality in South Africa, 2000 – priority-setting for now and a benchmark for the future. *S Afr Med J* 2005;**95**:496–503.
  11. Mathers C, Lopez AD, Murray CJL. The burden of disease and mortality by condition: data, methods, and results for 2001. In: Lopez AD, Mathers C, Ezzati M, Jamison D, Murray CJL (eds). *Global Burden of Disease and Risk Factors*. New York: Oxford University Press, 2006.
  12. Dorrington RE, Bradshaw D. Maternal mortality in South Africa: lessons from a case study in the use of deaths reported by households in censuses and surveys. *J Popul Res* 2011;**28**:49–73.
  13. Burger EH, Groenewald P, Bradshaw D, Ward AM, Yudkin PL, Volmink J. Validation study of cause of death statistics in Cape Town, South Africa, found poor agreement. *J Clin Epidemiol* 2012;**65**:309–16.
  14. Yudkin PL, Burger EH, Bradshaw D, Groenewald P, Ward AM, Volmink J. Deaths caused by HIV disease under-reported in South Africa. *AIDS* 2009;**23**:1600–02.
  15. Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D. Identifying deaths from AIDS in South Africa. *AIDS* 2005;**19**:193–201.
  16. Birnbaum JK, Murray CJL, Lozano R. Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organ* 2011;**89**:278–85.
  17. Westwood AT. Childhood deaths due to HIV – the role of the new death certificate. *S Afr Med J* 2000;**90**:877.
  18. Bradshaw D, Groenewald P, Laubscher R *et al.* Initial burden of disease estimates for South Africa, 2000. *S Afr Med J* 2003;**93**:682–88.
  19. Statistics South Africa. Mortality and causes of death in South Africa, 2007: Findings from death notification. Statistical release P0309.3. Pretoria, South Africa: Statistics South Africa, 2009.
  20. Kahn K, Collinson MA, Gómez-Olivé FX *et al.* Profile: Agincourt Health and Socio-demographic Surveillance System. *Int J Epidemiol* 2012;**41**:988–1001.
  21. Herbst A, Mafojane T, Newell M-L. Verbal autopsy-based cause-specific mortality trends in rural KwaZulu-Natal, South Africa, 2000–2009. *Popul Health Metr* 2011;**9**:47.
  22. Kanjala C, Alberts M, Byass P, Burger S. Spatial and temporal clustering of mortality in Digkale HDSS in rural northern South Africa. *Glob Health Action* 2010;**3**:5236.
  23. World Health Organization. *WHO Technical Consultation on Verbal Autopsy Tools Final Report: Review of the Literature and Currently-Used Verbal Autopsy Tools*. Geneva: World Health Organization, 2005.
  24. Kahn K, Tollman SM, Garenne M, Gear JSS. Validation and application of verbal autopsies in a rural area of South Africa. *Trop Med Int Health* 2000;**5**:824–31.
  25. Byass P, Kahn K, Fottrell E, Mee P, Collinson M, Tollman S. Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. *Popul Health Metr* 2011;**9**:46.
  26. Tollman SM, Kahn K, Sartorius B, Collinson MA, Clark SJ, Garenne ML. Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study. *Lancet* 2008;**372**:893–901.
  27. Kahn K, Tollman S, Garenne ML, Gear JSS. Who dies from what? Determining cause of death in South Africa's rural north-east. *Trop Med Int Health* 1999;**4**:433–41.
  28. Pillay-van Wyk V, Msemburi W, Laubscher R *et al.* Second National Burden of Disease Study South Africa: national and subnational mortality trends, 1997–2009. *Lancet* 2013;**381**:S113.
  29. Byass P, Kahn K, Fottrell E, Collinson MA, Tollman SM. Moving from data on deaths to public health policy in Agincourt, South Africa: Approaches to analysing and understanding verbal autopsy findings. *PLoS Med* 2010;**7**:e1000325.
  30. World Health Organization. *ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision*. 2nd edn. Geneva: World Health Organization, 2004.
  31. Statistics South Africa. *Mortality and Causes of Death in South Africa, 2008: Findings from death notification. Statistical release P0309.3*. Pretoria, South Africa: Statistics South Africa, 2010.
  32. South Africa Government Services. *Application to Give Notice of Death*. [http://www.services.gov.za/services/content/Home/ServicesForPeople/TraveloutsideSA/Applicationtogivenoticeofdeath/en\\_ZA](http://www.services.gov.za/services/content/Home/ServicesForPeople/TraveloutsideSA/Applicationtogivenoticeofdeath/en_ZA) (25 September 2011, date last accessed).
  33. Bradshaw D, Pillay-Van Wyk V, Laubscher R *et al.* Cause of Death Statistics for South Africa: Challenges and Possibilities for Improvement. Cape Town: South African Medical Research Council, 2010.
  34. National Center for Health Statistics. *Instruction Manual, Part 2c: ICD-10 ACME Decision Tables for Classifying Underlying Causes of Death*. Hyattsville, MD: US Department of Health and Human Services, Public Health Services, 2001.

35. SQL Server 2008. 2008.
36. Kabudula CW, Joubert JD, Tuoane-Nkhasi M *et al.* Evaluation of record linkage of mortality data between a health and demographic surveillance system and national civil registration system in South Africa. *Popul Health Metr* (In press).
37. StataCorp LP. *Stata Statistical Software: Release 12*. College Station, TX: StataCorp LP, 2011.
38. World Health Organization. *Special Tabulation Lists for Mortality and Morbidity: Mortality Tabulation List 1. International Statistical Classification of Diseases and Health Related Problems, Tenth Revision*. 2nd edn. Geneva: World Health Organization, 2004.
39. Rao C, Kosen S, Bisara D *et al.* Tuberculosis mortality differentials in Indonesia during 2007–2008: evidence for health policy and monitoring. *Int J Tuberc Lung Dis* 2011;15:1608–14.
40. Pattaraarchachai J, Rao C, Polprasert W *et al.* Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Popul Health Metr* 2010;8:12.
41. Wang L, Yang G, Ma J *et al.* Evaluation of the quality of cause of death statistics in rural China using verbal autopsies. *J Epidemiol Community Health* 2007;61:519–526.
42. Lozano R, Lopez AD, Atkinson C *et al.* Performance of physician-certified verbal autopsies: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr* 2011;9:32.
43. Grandin W, Westwood T, Lagerdien K, King MS. Deaths at Red Cross Children's Hospital, Cape Town 1999–200 – a study of death notification forms. *S Afr Med J* 2006;96:964–68.
44. Nojilana B, Groenewald P, Bradshaw D, Reagon G. Quality of cause of death certification at an academic hospital in Cape Town, South Africa. *S Afr Med J* 2009;99:648–52.
45. Nam J-M, Blackwelder WC. Analysis of the ratio of marginal probabilities in a matched-pair setting. *Stat Med* 2002;21: 689–99.
46. Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R. Algorithms for enhancing public health utility of national causes-of-death data. *Popul Health Metr* 2010;8:9.
47. Nojilana B, Brewer L, Bradshaw D, Groenewald P, Burger EH, Levitt NS. Certification of diabetes-related mortality: the need for an international guideline. *J Clin Epidemiol* 2013;66:236–37.
48. Garcia-Jardon M, Bhat VG, Blanco-Blanco E, Stepien A. Postmortem findings in HIV/AIDS patients in a tertiary care hospital in rural South Africa. *Trop Doct* 2010;40:81–84.
49. Byass P, Calvert C, Miiro-Nakiyingi J *et al.* InterVA-4 as a public health tool for measuring HIV/AIDS mortality: a validation study from five African countries. *Glob Health Action* 2013;6:22448.
50. Dhai A, Dada MA, Kirk GM, McQuoid-Mason DJ. Confidentiality: a dying wish? *S Afr Med J* 2001;91:123–27.
51. Burger EG, Bradshaw D, Groenewald P. Completion of death notice forms. *SAMA Insider* 2012:22.
52. Rankin WW, Brennan S, Schell E, Laviwa J, Ranking SH. The stigma of being HIV-positive in Africa. *PLoS Med* 2005;2:e247.
53. Kabudula CW, Joubert JD, Tuoane-Nkhasi M, Kahn K, Bradshaw D. *Linking Demographic Surveillance Mortality Data With National Vital Registration System Data: Lessons Learned From Agincourt, South Africa*. Johannesburg, South Africa: School of Public Health, University of Witwatersrand, 2013.
54. Leitao J, Chandramohan D, Byass P *et al.* Revising the WHO verbal autopsy instrument to facilitate routine cause-of-death monitoring. *Glob Health Action* 2013;6:21518.
55. Byass P, Chandramohan D, Clark SJ *et al.* Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Glob Health Action* 2012;5:19281.
56. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJL, the Population Health Metrics Research Consortium (PHMRC). Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr* 2011;9:29.
57. James SL, Flaxman AD, Murray CJL, the Population Health Metrics Research Consortium (PHMRC). Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies. *Popul Health Metr* 2011;9:31.
58. Murray CJL, James SL, Birnbaum JK *et al.* Simplified Symptom Pattern Method for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr* 2011;9:30.

## **5.6 Cause-of-death agreement according to selected characteristics of the decedent, venue of death, and type of ascertainment of the cause of death**

Given the analytic potential in the linked dataset, additional analysis beyond the objectives of the research paper has been conducted to ascertain whether cause-of-death agreement between the CRVS and Agincourt HDSS data sources is better for particular individual characteristics of the deceased (gender and age), and according to the venue of death and the type of ascertainment of the cause of death (rows A to M in Table 5.1).

Although the level of agreement remains low in both sexes, agreement at the 95% confidence level was significantly better ( $P = 0.0000$ ) for males (26.4%; kappa 0.1910) than females (19.9%; kappa 0.1320) (row A). A significantly higher level of agreement ( $P = 0.0000$ ) was also observed in row C for causes of death ascertained by forensic autopsy, the attending physician, or an attending physician on duty (26.9%; kappa 0.1838), compared to causes ascertained by a registered nurse (in perinatal deaths) or an interview with the family (12.2%; kappa 0.0715). For in-hospital versus out-of-hospital deaths in row B, however, the difference in agreement of causes for decedents of all ages was not significant ( $P = 0.0984$ ).

Cause-of-death agreement as between the data sources was examined by age for adjacent broad age groups in rows D to I, showing no significant difference in the level of agreement when comparing these adjacent age groups. However, cause-of-death agreement was significantly higher ( $P = 0.000$ ) in row K for deaths occurring at ages younger than 65 years (24.9%; kappa 0.1671) compared to ages 65 years or older (17.5%; kappa 0.1049).

Whereas in-hospital versus out-of-hospital deaths for all ages did not show a significant difference in agreement ( $P = 0.0984$ ) (rows 6 and 7 in Table 5.1), further sub-group analysis by age shows that, in decedents 65 years or older, cause-of-death agreement between the data sources was significantly better ( $P = 0.0000$ ) among those dying in hospital (28%; kappa 0.2094) compared to those dying out of hospital (13.5%; kappa 0.0654). For children under 5 years old, cause-of-death agreement was somewhat better, though not significantly better ( $P = 0.4517$ ) for those dying in hospital (29.5%; kappa 0.1678) compared to those dying out of hospital (22.2%; kappa 0.1246) (compare the last 7 rows of Table 5.1)

Table 5.1: Cause-of-death agreement according to selected demographic characteristics of the decedent, venue of death, and type of ascertainment of the cause of death

	Agreement: %	Expected agreement: %	Kappa	Standard Error	P-value: difference between observed and expected agreement	P-value: difference between agreement % in a) and b) in column 2
<b>A. Gender:</b>						
a) Males	26.40	9.02	0.1910	0.0900	0.0000	0.0000*
b) Females	19.89	7.70	0.1320	0.0084	0.0000	
<b>B. Venue of death:</b>						
a) In-hospital	27.16	12.28	0.1696	0.0109	0.0000	0.0984
b) Out-of-hospital	19.91	5.99	0.1481	0.0071	0.0000	
<b>C. Ascertainment of cause of death by:</b>						
a) Forensic autopsy, attending physician, or attending physician on duty	26.86	10.39	0.1838	0.0093	0.0000	0.0000*
b) Nurse or interview with family of deceased	12.17	5.41	0.0715	0.0092	0.0000	
<b>D. Age at death: Infants and children:</b>						
a) <1 yr	27.08	14.27	0.1495	0.0393	0.0001	0.8012
b) 1 to 4 yrs	23.61	11.65	0.1354	0.0399	0.0003	
<b>E. Age at death: Children:</b>						
a) <5 yrs	25.60	12.86	0.1461	0.0281	0.0000	0.1788
b) 5 to 14 yrs	28.33	9.11	0.2115	0.0397	0.0000	
<b>F. Age at death: Children and middle-aged adults:</b>						
a) 5 to 14 yrs	28.33	9.11	0.2115	0.0397	0.0000	0.1668
b) 15 to 59 yrs	24.65	10.79	0.1554	0.0084	0.0000	
<b>G. Age at death: Middle-aged and older adults:</b>						
a) 15 to 59 yrs	24.65	10.79	0.1554	0.0084	0.0000	0.1986
b) 60 to 79 yrs	19.74	7.29	0.1343	0.0141	0.0000	
<b>H. Age at death: Older and oldest-older adults:</b>						
a) 60 to 79 yrs	19.74	7.29	0.1343	0.0141	0.0000	0.0618
b) 80+ yrs	16.13	8.15	0.0869	0.0211	0.0000	
<b>I. Age at death: Children under-5 and persons 5-59 years:</b>						
a) <5 yrs	25.60	12.86	0.1461	0.0281	0.0000	0.6594
b) 5 to 59 yrs	24.80	10.58	0.1590	0.0082	0.0000	
<b>J. Age at death: Children under-5 and older adults 60+:</b>						
a) <5 yrs	25.60	12.86	0.1461	0.0281	0.0000	0.3840
b) 60+ yrs	18.55	7.49	0.1196	0.0117	0.0000	

	Agreement: %	Expected agreement: %	Kappa	Standard Error	P-value: difference between observed and expected agreement	P-value: difference between agreement % in a) and b) in column 2
<b>K. Age at death: Adults &lt;65 and older adults 65 +:</b>						
a) <65 yrs	24.86	9.79	0.1671	0.0074	0.0000	0.0000*
b) 65+ yrs	17.50	7.83	0.1049	0.0130	0.0000	
<b>L. Age 65+: dying in- vs out-of-hospital</b>						
a) 65+ dying in hospital	28.03%	8.96%	0.2094	0.0263	0.0000	0.0000*
b) 65+ dying out of hospital	13.51%	7.45%	0.0654	0.0148	0.0000	
<b>M. Age &lt;5 yrs: dying in- vs out-of-hospital</b>						
a) <5 dying in hospital	29.49%	15.27%	0.1678	0.0461	0.0001	0.4517
b) <5 dying out of hospital	22.22%	11.15%	0.1246	0.0342	0.0001	
* Significant difference at the 95% confidence level						

## 5.7 Concluding remarks regarding published Paper 4 in the context of Chapter 5

### 5.7.1 Low cause-of-death agreement

The low agreement of the underlying cause could be real as a result of low reliability of the diagnoses which are dependent on the nature and accuracy of available evidence and proxy reporting, but could also be an artefact of the dissimilar cause-reporting, cause-attribution and coding practices used in the different data sources. The known significant systematic misattribution of HIV/AIDS in the CR data<sup>123,161,162</sup>—by far the largest detailed cause of death in the population<sup>93-95,163</sup>—affects agreement substantially, not only by under-reporting the numbers of deaths due to HIV/AIDS, but also by artificially and invalidly inflating a range of other conditions. The under-reporting of HIV/AIDS on the death certificate has been observed for non-hospital deaths in Thailand too, where Polprasert *et al* reported about a three-fold under-reporting in the CRVS system compared to verbal autopsy findings.<sup>233</sup> For hospital deaths in Thailand, CRVS cause-of-death data were validated against medical records, and Pattaraarchachai *et al* also reported an approximately three-fold under-reporting of HIV disease in the CR system too.<sup>232</sup>

The additional analysis in Section 5.6 highlights the need to examine causes of death beyond aggregate data. In this dataset, it was important to consider separate analysis by gender; physician/non-physician ascertainment of cause of death; and death before/after age 65.

Among decedents dying after age 65, the analysis suggests that it would additionally be desirable to distinguish between deaths that occurred in/out of hospital. Should it be possible to extend the results to an attempt to adjust the district-level CRVS cause-of-death data, this analysis indicates the need of such adjustment to be done separately for a) males and females; b) deaths certified by a physician and not; and c) decedents who died at ages younger and older than 65 years, prior to aggregating data for the population considered.

Furthermore, better agreement between the CRVS and Agincourt HDSS data found in physician-ascertained causes of death amplifies the importance and continuation of a local initiative which developed a training programme targeting certifiers/physicians in public sector hospitals in South Africa. The initiative, driven by Stats SA, the National Department of Health, the National Department of Home Affairs and the South African Medical Research Council, included the compilation of a guideline for physicians to complete the death notification form; a two-day training programme aiming to train ‘master trainers’ at facilities in all the provinces; materials for these ‘master trainers’ to use when training their colleagues; and an easy-to-use flyer with summary guidelines on DNF completion.<sup>310</sup>

### **5.7.2 Misclassification patterns**

A key outcome of a validation study is that it allows the identification, for the causes of death included, of the principal conditions involved in its misclassification. Misclassification patterns are useful for alerting certifiers, coders and users of CRVS cause-of-death data of major or systematic misclassification in the data.

The misclassification patterns in the Agincourt HDSS study area suggests higher reporting of diabetes coupled with lower reporting of cardiovascular disease deaths in the CRVS diagnoses. This pattern has also been reported in an urban study by Burger *et al* who validated CRVS causes against medical records in Cape Town.<sup>123</sup> Mühlhauser *et al*, however, found lower reporting of diabetes coupled with misclassification to cardiovascular conditions in Düsseldorf, Germany. In urban China, diabetes deaths were found to be over-reported in the registered data compared to findings from a medical record review. In studies of hospital<sup>232</sup> and non-hospital<sup>233</sup> deaths in Thailand, in contrast, considerable under-reporting of diabetes was observed in death certificates compared to medical records<sup>232</sup> and verbal autopsies.<sup>233</sup>

### 5.7.3 Cause-specific mortality fractions

Cause-specific mortality fractions for the registered and verbal autopsy data differ significantly for 11 of the 15 causes/cause groups considered in the short list. The largest difference is observed for HIV disease (331%), followed by Remaining infectious and parasitic disease (145%), and Diarrhoea (-81%), pointing to significant problems in the CRVS data to accurately attribute these pre-transitional causes/cause groups in the area. Although within a smaller margin of their true value, the difference in cause-specific mortality fractions for the remaining causes generally reflects poor quality, and hence limited utility, of CRVS cause-of-death data.

Should the verbal autopsy diagnoses reliably reflect the true cause-of-death pattern in the Agincourt HDSS, this conclusion provides an unfavourable answer to research question 6. In turn, this raises the question whether cause-specific mortality fractions can be used to adjust cause-of-death CRVS data. This question will be explored in the next section, guided by some of the steps in the methodology followed in the Thai death certification validation study.<sup>292,293</sup>

### 5.8 Application of cause-specific mortality fractions as adjustment factors

CRVS mortality data have become available by health district since 2008. In response to health-district managers' vast need for health and health-indicator information, the *District Health Barometer*, a report published since 2005, has started providing district-level mortality profiles based on vital statistics information. For the first time then, the 2011/12 District Health Barometer included the CRVS cause-of-death and premature-death profile for each health district, presented according to 140 locally-relevant causes specified for the SA NBD 2 Study. These district profiles were created after (i) redistributing deaths of unknown ages proportionally by age and sex across each known cause of death; (ii) combining causes of death used as pseudonyms for HIV/AIDS (e.g. immune suppression or retroviral disease) with HIV/AIDS deaths; (iii) proportional redistribution of ill-defined signs and symptoms (ICD-10 Chapter R) and 'garbage' codes as defined by Naghavi *et al*<sup>153</sup> to specified causes within age and sex categories; and (iv) adjusting injury deaths as described in Groenewald *et al*.<sup>13</sup>

However, the systematic misclassification of causes of death, as found in a number of studies, has not been taken care of, even though such misclassification has been found to significantly impact on population-<sup>93,123,161,162</sup> and facility-based<sup>120,160,301</sup> cause profiles. As the scale of



misclassification has vast implications for planning health service provision and preventive programmes, the report cautions that the information needs to be interpreted with care.

This user warning, and the critical need of health-district managers to have valid facts of the cause-of-death profile of their districts, prompted a final research question, i.e. “*Are the linkage study results useful for developing adjustment factors for cause-of-death attribution at the district level?*”

### **5.8.1 Exploration with adjustment factors**

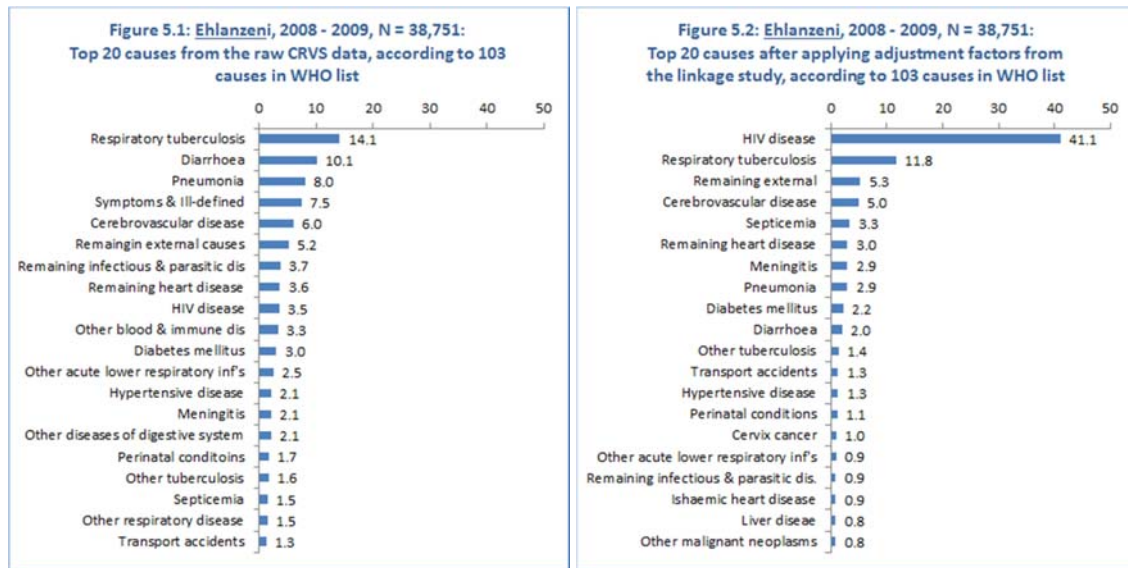
Exploratory work has been done on using the results from this validation study to adjust the cause-of-death data for the Ehlanzeni health district in which the Agincourt HDSS falls. Ehlanzeni covers an area of 27,895 km<sup>2</sup>, and has a population of 1,434,201 people, the largest of the three health districts in Mpumalanga province. Ehlanzeni is geographically made up of five local municipalities, one of which is Bushbuckridge, which is the municipal area in which the Agincourt HDSS site is located<sup>311</sup> as shown in Figure 5.8

For the period 2008-2009 a total of 38,751 deaths were registered for the Ehlanzeni health district. The percentage difference in the CSMF observed in the Agincourt HDSS/Stats SA linkage study (cf Table 6, Section 5.5), has been applied to the cause-of-death data from CRVS to derive an adjusted profile. As done in the Agincourt HDSS/Stats SA cause comparison study (Section 5.5), the underlying causes of death for Ehlanzeni were first aggregated into the 103 causes of the WHO ICD Mortality Tabulation List 1 (WHO list).<sup>312</sup> The percentage difference in cause-specific mortality fractions between the CRVS and Agincourt HDSS for each of the causes in the short list (Table 6, Section 5.5), were taken as adjustment factors to be applied to all the cause-specific mortality fractions in that group of causes in the Ehlanzeni data:

$$\text{adj CSMF} = \text{CSMF} * (100 + \text{Agincourt \% difference in CSMF})/100$$

A clear change in the cause profile is evident after applying the adjustment factors (compare Figure 5.1 and 5.2 below), of which the substantial increase of HIV disease from 3.5% to 41.1% is most noticeable. After adjustment, Diarrhoea and Pneumonia, respectively, reflect a 5- and 2.6-fold reduction, and Remaining infectious and parasitic disease and Other acute lower respiratory infections are substantially reduced in their proportional contributions to death in the adjusted profile.

Plotting the adjusted against the raw cause-specific mortality fractions for Ehlanzeni district, demonstrates that most of the 103 causes tend to cluster around the straight line. This suggests that the cause attribution in the raw data is reasonably similar to that of the adjusted profile ( $R^2 = 0.7591$ ), except for the outlier conditions labelled in the graph. These outliers strongly resemble the AIDS-indicator conditions as found by Groenewald *et al*,<sup>161</sup> Birnbaum *et al*,<sup>162</sup> Yudkin *et al*<sup>121</sup> (Figure 5.3).



Figures 5.1 & 5.2: Top 20 causes of death for Ehlanzeni district, based on raw (Figure 5.1) and adjusted CRVS data (Figure 5.2).

Source: Statistics South Africa.

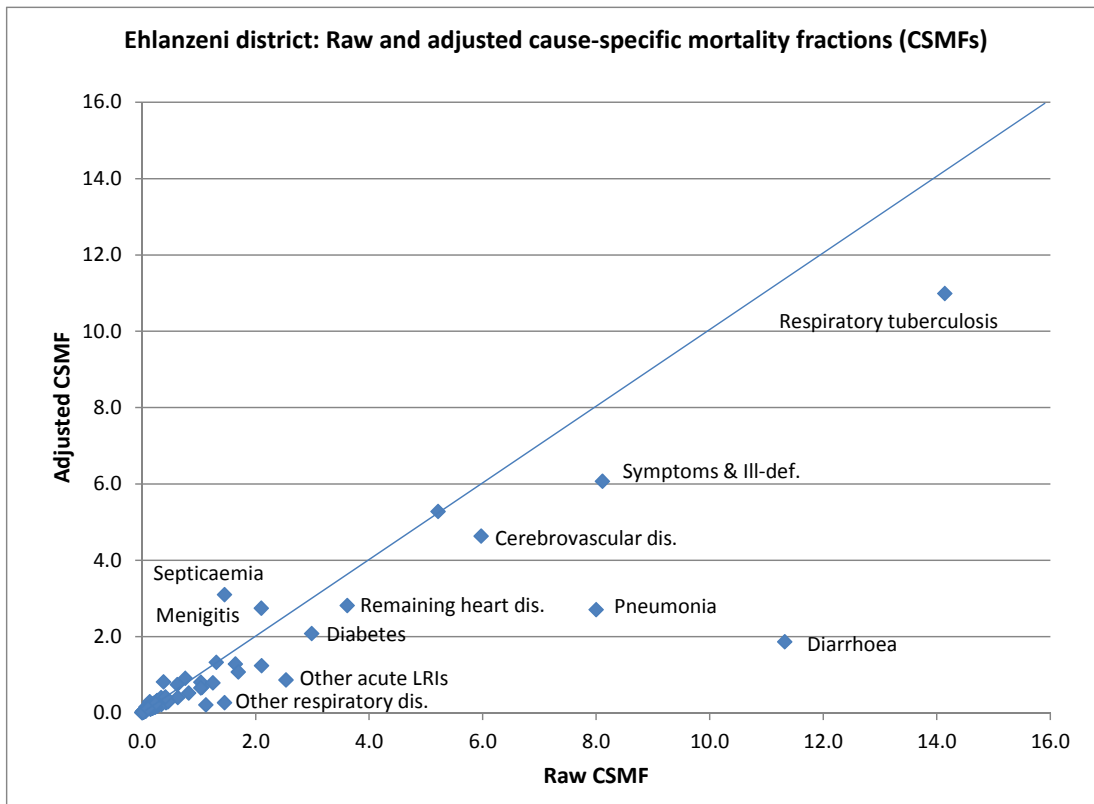
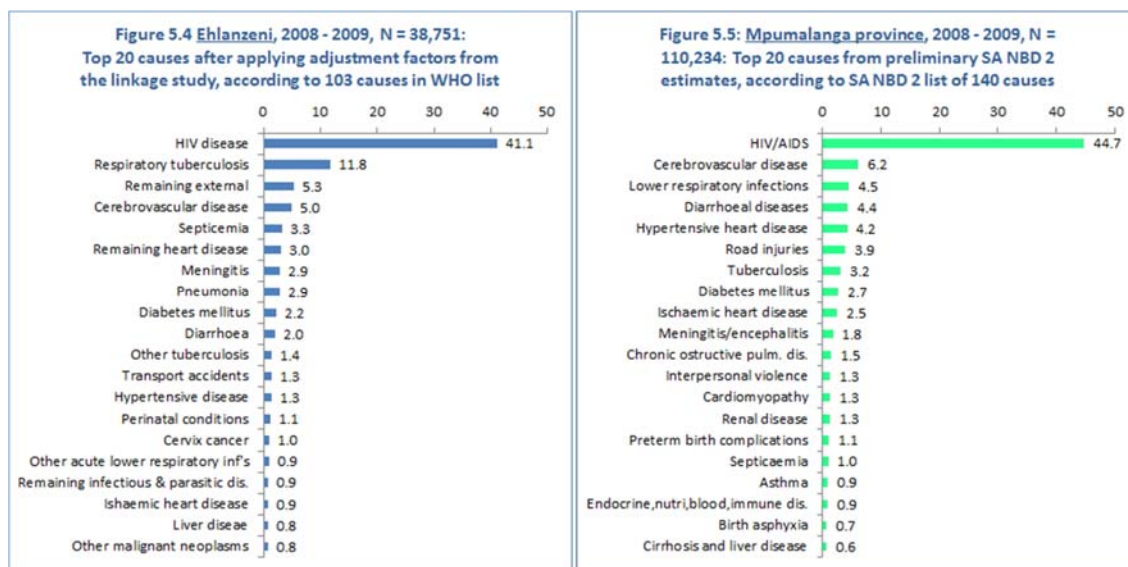


Figure 5.3: Raw and adjusted cause-specific mortality fractions for Ehlanzeni district.  
 Source: Statistics South Africa. Note: HIV disease excluded to avoid skewing the graph.

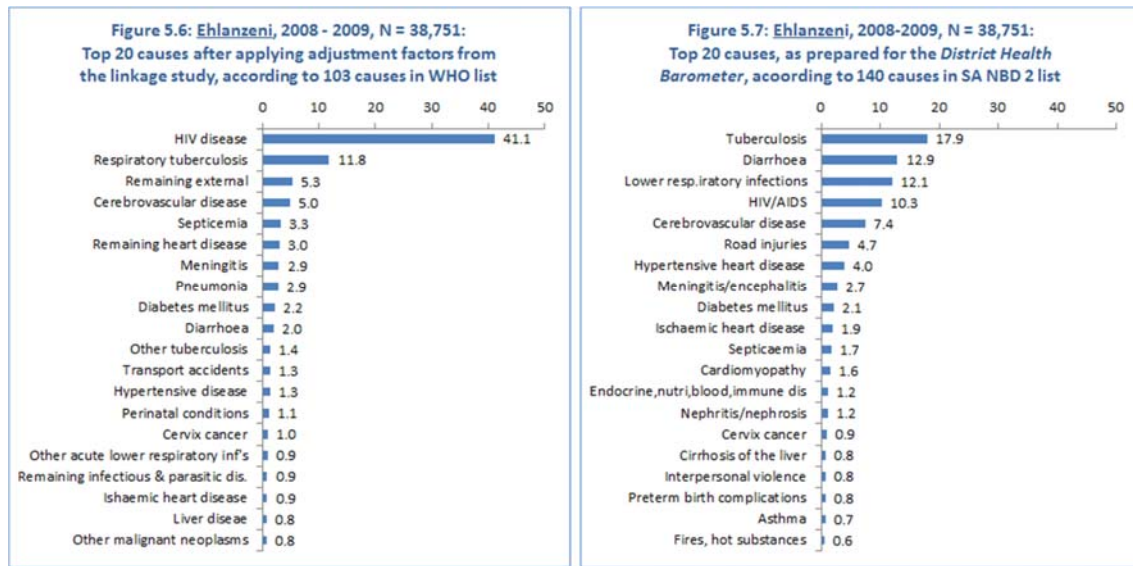
The resultant cause profile (Figure 5.4 below) has in addition been compared to preliminary estimates of the profile for Mpumalanga province from the second National Burden of Disease Study that are based on burden of disease methodologies to adjust for the challenges in causes of death data (Figure 5.5) (data extracted and supplied by Mr Ian Neethling, SA MRC). While the populations are not identical, and the cause of death list differs, the appraisal does serve as an external comparison of the adjustment factors. Figures 5.4 and 5.5 indicate excellent agreement about HIV/AIDS as a major leading cause of death, and reasonable agreement that cardiovascular disease, respiratory infections, and external causes are major groupings of disease and injury causing death in these areas.



Figures 5.4 & 5.5: Top 20 causes of death for Ehlanzeni district based on adjusted CRVS data (Figure 5.4), and top 20 causes of death for Mpumalanga province (Figure 5.5).

Source: Statistics South Africa (Ehlanzeni data) and Burden of Disease Research Unit, SA MRC (Mpumalanga data).

The adjusted CRVS estimates (Figure 5.6 below) are also compared with the profile reported by the *District Health Barometer* report (Figure 5.7). For this report, the estimates have been adjusted for (i) missing age, (ii) HIV/AIDS pseudonyms, (iii) ill-defined and ‘garbage’ codes, and to some extent, (iv) anomalous injury codes. The cause of death lists are different, but the substantial difference in the proportion of death due to HIV disease and the different rank order in the graphs below, suggest that the adjustments made for the *District Health Barometer* have been insufficient.



Figures 5.6 & 5.7: Top 20 causes of death for Ehlanzeni district based on adjusted CRVS data (Figure 5.6), and top 20 causes of death for Ehlanzeni district based on the profile presented by the *District Health Barometer* report<sup>311</sup> (Figure 5.7).

Source: Statistics South Africa.

## 5.9 Conclusion

Chapter 5 reports a promising, low-cost validation method that provided empirical evidence about the quality and utility of CRVS cause-of-death data in this rural area. It has confirmed other studies that have observed concerns about the quality of cause-of-death information.<sup>121,123,160-162,301</sup> While the evidence advances knowledge about cause-of-death data and cause-of-death research methods in South Africa, the low cause-of-death agreement should be of considerable concern to cause-of-death certifiers and coders, and researchers and health planners using CRVS cause-of-death data for research recommendations and health decision-making, respectively. Additional analysis in Section 5.6 highlights the importance of sub-group analyses of cause-of-death data, and the potential for such sub-group analysis to inform attempts to adjust district-level CRVS cause-of-death data.

The results point to a clear need to improve cause attribution at the individual level with a view to generating more reliable population-level statistics from the CRVS system. The exploratory example found reasonable agreement between the external comparison profile of Mpumalanga province and the adjusted Ehlanzeni profile, suggesting (a) that the Agincourt HDSS/Stats SA linkage study results have been useful for developing adjustment factors for cause-of-death

attribution at the district level; and (b) that it would be worthwhile exploring a similar analyses and application using cause-of-death data from the two remaining HDDS sites and their respective surrounding districts.

The Agincourt HDSS cause-of-death adjustment factors, however, cannot be used for the whole country. Adjustment factors considerably more representative than these would be needed, and, to ensure a longer list of conditions, a study of a much larger scale would be required. Given the successful study in Thailand where cause-of-death adjustment factors have been generated and applied nationally to estimate cause-specific mortality patterns for the country,<sup>293</sup> this approach needs further exploration in South Africa.

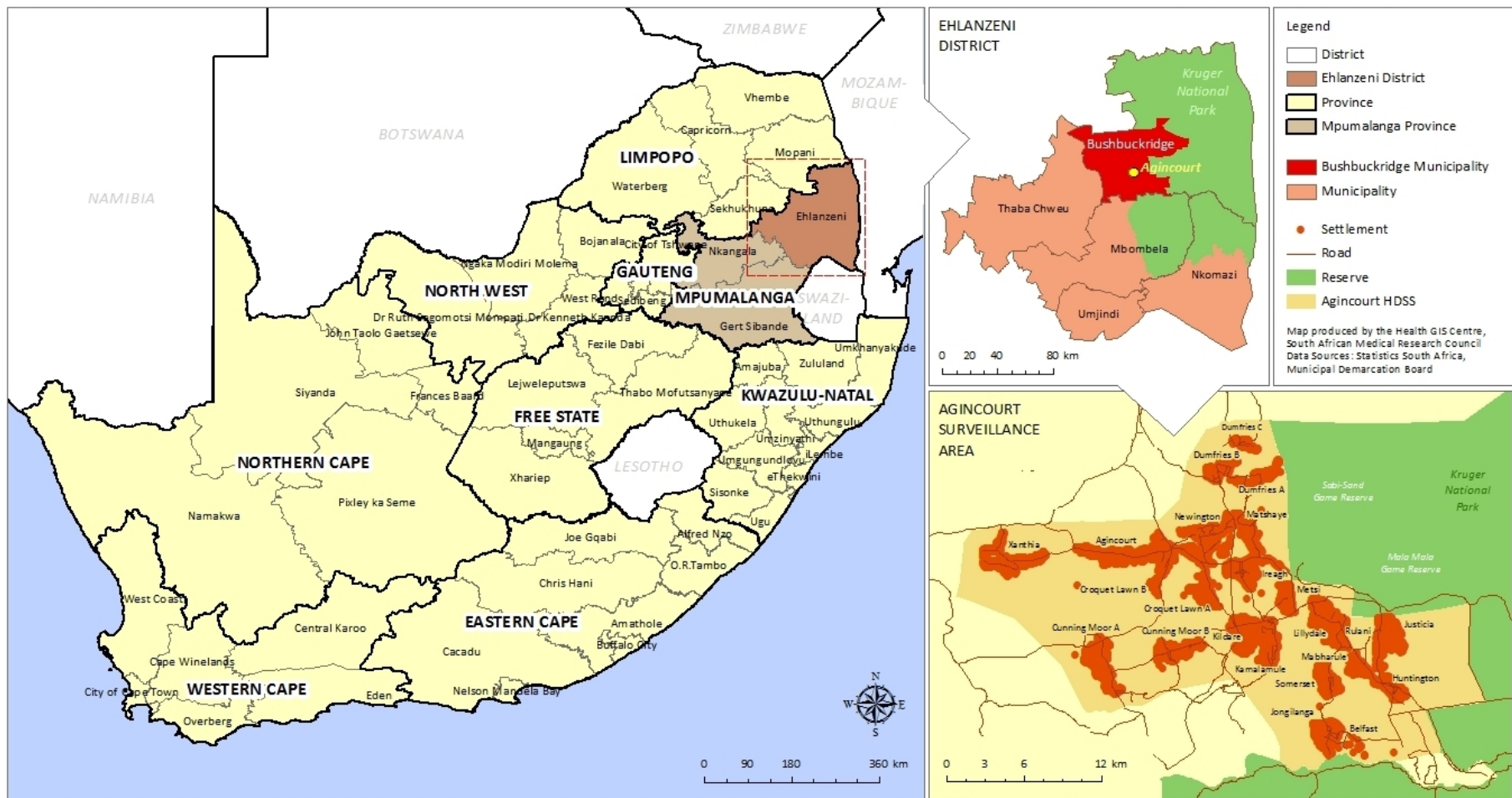


Figure 5.8: Provincial and district administrative delineations of South Africa with insets of Ehlazeni district and its five municipalities, and Agincourt HDSS with its villages.

Source: Map produced by the Health GIS Centre, South African Medical Research Council. Administrative boundaries sourced from the Municipal Demarcation Board, 2011; Settlement level data sourced from Statistics South Africa, 2011.





## 6 Chapter 6: Synopsis of findings, conclusions, and recommendations

---

The primary aim of this body of research was to assess the quality of mortality data in the national CRVS system of South Africa. Acknowledging the potential value of alternative mortality data sources to complement or evaluate the quality of civil registration data, a secondary aim was to identify and review alternative mortality data sources for South Africa, and assess the availability, strengths, and limitations of their data. In view of the primary aim, and the findings from the data source review and quality assessment, a further aim was to conduct a field study to link and compare CRVS and Agincourt HDSS mortality data, for the same individuals, with the purpose of quantifying the level of completeness of death registration into the CRVS system, and assess the level of agreement between CRVS and verbal autopsy cause-of-death data.

Seven research questions have been addressed towards achieving these aims. A synopsis of the key findings, conclusions, and recommendations related to each research question are provided in Section 6.1. Many of the recommendations have been mentioned in previous chapters. They are listed here to provide a consolidated overview, and facilitate their proximity to key findings and conclusions.

### 6.1 Synopsis of key findings, conclusions and recommendations

#### 6.1.1 Review of South Africa's mortality data sources

Research question 1: *What mortality data sources are available in democratic South Africa, and what are the limitations, strengths, uses, and public availability of data from these sources?*

In 2007, the *Who Counts* Lancet series reported that, globally, little progress has been made since the 1970s in producing data sources with reliable mortality data.<sup>5,313</sup> For the developing world, in particular, with the exception of child mortality, mortality statistics have been described as being in a “deplorable state”.<sup>44</sup> For the African continent, considerable challenges in statistical development persisted,<sup>8,9,314</sup> including slow progress in the development and maintenance of CRVS systems,<sup>4,314,315</sup> and vast under-registration of deaths, a scarcity of

representative alternative mortality data sources, and poor or no availability of cause-of-death data.<sup>3,5,8,9,43</sup>

For South Africa, however, the mortality data source review (Paper 1) documents a variety of data sources, suggesting a data-rich situation. CRVS data are available in various formats and useful for investigating levels of all-cause and cause-specific mortality. Standard mortality data items for child, adult, maternal and all-cause mortality are available from alternative national data collections, disaggregated by different jurisdictions. However, data usage of the alternative collections has been hampered by a number of shortcomings, including quality-control problems in the field, and limited technical capacity for assessing data-quality and applying analytical methods for mortality estimation.

To improve mortality measurement for South Africa, it is necessary to further invest in strengthening the country's CRVS system towards universal notification, registration, and certification of all vital events. In the absence of this possibility, a coordinated research agenda for mortality data collection, triangulation, alongside an operational agenda for quality assurance and analytical capacity development, are suggested through specific recommendations below.

#### **6.1.1.1 Recommendations**

##### **Improving data quality**

- The utility and quality of mortality data items in censuses and sample surveys should be enhanced by improving fieldworker training and implementing first-rate quality control measures in the field.
- For surveys, independent expert evaluation of sample realisation, data quality checks, and examination of data edits performed by the data proprietor/s, are recommended.
- For censuses, independent expert evaluation of content and coverage errors and editing rules are recommended.
- From a statistical perspective, censuses, surveys, and sample surveillance are not adequate substitutes for a well-functioning CRVS system.<sup>316</sup> It is therefore strongly recommended that, as far as resources can be mobilized, initiatives are focussed on strengthening the CRVS system by ensuring continued multi-sector collaboration and cooperation; and exploring the initiatives that will further improve completeness of

death registration, and the coverage and accuracy of medical certification of causes of death.

### **Promoting triangulation with the prospect of enhanced confidence in findings**

- Triangulation in research refers to the use of more than one approach in examining the same phenomenon or research question, with the prospect of enhanced confidence in the ensuing findings.<sup>317,318</sup> Denzin extended the concept beyond the conventional association with research design,<sup>319</sup> and distinguished among data, investigator, theoretical, and methodological triangulation. An agenda directed at triangulation, drawing on different data sources, methods, and investigators should be promoted. A data-linkage design holds particular promise for direct estimation of death-registration completeness, to be compared to indirect estimates in methodological triangulation. Data triangulation provides opportunity for comparing cause-of-death data in CRVS data with cause-of-death data from medical, forensic, or verbal autopsy records, and the identification of misclassification patterns of causes of death once a reference diagnosis has been established. In turn, such patterns must be used for alerting cause-of-death certifiers, coders, and data users to systematic biases in the data.

### **Strengthening analytic capacity**

- Data-quality assessment and triangulation typically require competent analytical capacity; yet such capacity is scarce in South Africa. The strengthening of analytical capacity through expert recruitment, and local training and internships in relevant fields of study, is therefore a key overarching recommendation.

#### **6.1.2 Evaluation of South Africa's mortality statistics from CRVS**

Two WHO comparative assessments rated the quality of South Africa's 1996 mortality data as low.<sup>8,9</sup> With the transition to democracy, however, civil registration underwent major transformation<sup>106</sup> (Section 3.2). Key events—including the demise of separate development policies, and passing of the Births and Deaths Registration Act of 1992 that mandates all geographic areas and all individuals to register deaths in a centralised CRVS system—facilitated improved coverage of civil registration and vital statistics.

A range of actions were initiated to improve death-registration practices and the quality of vital data (Section 3.2.2). Given the investment in time and resources, and the possibility of sharing best practices with other countries, it was important to ascertain whether or not these initiatives had an impact on the availability and quality of post-1996 CRVS data. Hence, suitable criteria were sought for single-country evaluation of the quality of mortality data from civil registration (Sections 3.3 and 3.4) in order to answer research questions 2 and 3:

Research question 2: *What criteria can be used for evaluating national death and cause-of-death statistics from civil registration in South Africa?*

Research question 3: *What is the quality of national mortality statistics from civil registration for the period 1997 – 2007?*

A broad range of criteria are available for assessing data quality in country health information systems.<sup>145,146,320</sup> For CRVS, as a component of health information systems, there are fewer frameworks and criteria for quality assessment. Substantial agreement was found in the selection of criteria for country evaluations in Brazil,<sup>126</sup> China,<sup>10</sup> and India.<sup>152</sup> These criteria were deemed suitable for, and applied successfully, to CRVS mortality data for South Africa (Sections 3.3 and 3.4).

The evaluation findings revealed considerable progress and improvement from a pre-democracy dysfunctional system, to a well-functioning system that offers (a) all-cause mortality data that can be adjusted for biases of completeness; and, (b) timely and disaggregated detailed-cause data, certified and coded according to international standards.

Satisfactory ratings were given for six out of nine criteria: coverage, completeness, temporal consistency, use of age- and sex-improbable classifications, timeliness, and availability of sub-national data, demonstrating substantial improvement in data quality (Section 3.5) compared to the 1996 data evaluated in previous evaluations.<sup>8,9</sup>

Having found satisfactory ratings for six out of nine criteria, and realizing the state of mortality data collection and dissemination 10 - 15 years earlier, it was tempting to label the quality of South Africa's CRVS data as 'satisfactory', and so provide a favourable answer to research question 3. However, the unfavourable findings for the criteria content validity and use of ill-defined and non-specific codes strongly resembled unsatisfactory findings from previous studies reporting considerable shortcomings in cause-of-death data.<sup>93,161,162,321</sup> These criteria

have substantial potential to influence the validity and reliability of cause-of-death data in the absence of expert adjustment for biases. This impedes the utility of the data for population-health monitoring, public-health planning, and needs-based resource distribution in a resource-limited country with a considerable<sup>71,93,94</sup> burden of disease.

Recommendations were made for priority actions for authorities in civil registration, health, social development, and education and training, and for further research.

### **6.1.2.1 Recommendations**

#### **System and data quality improvement in CRVS**

- Efforts to improve completeness of death registration must continue, with a particular focus on improving the reporting of deaths in infants and children aged 1-4 years.
- The Departments of Health and Education are urged to communicate to medical schools the need for mandatory and sufficient training of medical students to accurately complete a death notification form and medical certificate of death; and ensuring students' knowledge of the legal requirement and public health importance of complete and accurate reporting on the death notification form.
- For in-service medical certifiers, accessible in-person or online death-certification training, along with training in keeping good-quality medical records, are recommended.
- To facilitate sufficient detailing of external causes of deaths to suitably inform injury prevention and safety and peace promotion, the incorporation of data fields to reveal the manner and external cause of injury deaths on the death-notification form is strongly recommended.

#### **Further research**

- Reviewing best practices in other settings for improving civil registration systems, enhancing completeness of vital-event registration, and the quality of vital data is recommended for useful knowledge and practice transfer.
- Linking CRVS mortality data with other relevant data collections can facilitate the assessment, and improvement, of (a) the completeness of death reporting into the civil registration system, and (b) the validity of certified causes of death from civil

registration. These critical aspects of a successful CRVS system are further elaborated upon in Section 7.6.1.

### 6.1.3 Data linkage for data quality assessment

The evaluation in Chapter 3 indicates a well-established national CRVS system in South Africa, offering all-cause and cause-specific mortality data by age, sex and sub-national jurisdictions; available for dissemination and research annually and timeously; with full mandatory geographic coverage and high overall estimates of completeness of death registration. While these attributes signify remarkable improvements over a short period of time, two critical criteria - *content validity of the causes of death* and *the use of ill-defined and non-specific codes* - were not performing well in the assessment, with the potential to profoundly affect the validity and representativeness of the country's cause-of-death data.

Chapter 4 identified record linkage as a useful, increasingly used approach to verify the quality of both cause-specific and all-cause mortality from CRVS.<sup>174,176-187</sup> After reviewing the characteristics, variables and availability of the various South African mortality data sources (Chapter 2), and examining the methods of data collection, cause attribution, and coding in each data source (Table 2.2), mortality data from the Agincourt HDSS were deemed suitable for the record-linkage study described in Section 4.5. In line with the aims of the thesis and linkage study, the following three research questions were formulated and answered:

Research question 4: *Would a linkage study between South African CRVS and the Agincourt HDSS mortality data collections be possible and, if so, would such linkage be of good quality and useful?*

Research question 5: *What can the mortality data collected in Agincourt HDSS indicate about the completeness of death-registration and factors associated with registration?*

Research question 6: *What can the verbal-autopsy data collected in Agincourt HDSS indicate about the quality and utility of cause-of-death attribution in the civil registration system?*

In response to research question 4, Chapter 4 concludes that, despite strict data-security policies, and a small number of common variables, the linkage study was possible. Due to

limitations in the information available and few common variables, a moderate matching rate (61%) was realised. However, the quality of the matching was high, yielding high sensitivity (90%) and positive predictive values (98.5%) for the probabilistic approach. The particular value of the national 13-digit unique identification number, issued to all South Africans by birth, was highlighted.

It was not possible to calculate completeness (Section 4.6), and therefore not possible to answer the first part of research question 5. In response to the latter part, the analysis identified subgroups that are likely to be under-represented in the CRVS data, i.e. young children, persons with temporary residency status, people with no education, people in the lowest wealth quintile, and non-citizens. This analysis provides scarce and useful empirical evidence about factors affecting death registration in the area. These factors have the potential to inform strategies to improve completeness of death registration locally, and accelerate improvement in death registration in countries with emerging CRVS systems.

Agreement of cause attribution between the data sources was low, sensitivity and positive predictive values were generally low, and cause-specific mortality fractions differ significantly for most causes/cause groups. Assuming accuracy in the reference verbal-autopsy diagnoses, these unsatisfactory epidemiological findings point to poor quality and limited utility of CRVS cause-of-death data in the area, and therefore, an unfavourable answer to research question 6.

Some recommendations from the linkage study may seem to focus insignificant or general-knowledge detail. However, they are likely to have a substantial impact on future analytical possibilities in the context of the government's commitment to universal health coverage and international targets to be finalized for the Sustainable Development Goals.

#### **6.1.3.1 Recommendations**

##### **System and mortality data improvement in the CRVS system**

- Given the strong discriminatory power of the 13-digit national identity number and its particular utility in the matching study, the responsible and accurate routine collection of this unique identifier in all the local HDSSs, and other routine health information systems, is recommended.

- In view of the demonstrated usefulness of place names in matching studies in health,<sup>174,178,184</sup> there is a need for collaborative activities between Stats SA and HDSSs to address place-name inconsistencies;
- Recording of actual village/town/suburb names on death notification forms instead of tribal area names, or adequate provision for both, is recommended. This would be critical for completeness estimation in all HDSSs that include tribal areas and names.
- Aligning study site borders with established official borders when setting up or extending HDSS sites would greatly enhance the prospect of directly estimating completeness of CRVS death registration in these sites.
- The development of an electronic place-name database linked to detailed maps, for use by Home Affairs registration offices, the police, undertakers, and hospital personnel would be useful for verifying the places of death occurrence, registration, and usual residence that need to be reported on the CRVS death notification form. Augmenting civil registration certification of home deaths with the systematic collection of standard verbal autopsies is recommended to support certification by headman in remote areas, as well as physician certification at emergency units and morgues when limited information is available about the circumstances and cause of death. With an estimated one in ten deaths being reported to traditional leaders in South Africa,<sup>99</sup> and wide coverage of mobile device signals, the feasibility of death registration via mobile technology needs exploration. Lessons can be learnt from such a system used by traditional leaders in Senegal to register births via a Java applet, and in Uganda to register births via USSD codes on mobile phones and a web-based application for computers.<sup>322</sup>

### **Further research**

- New developments in the use of standardised verbal autopsy instruments,<sup>323-326</sup> and refinements of automated methods for the interpretation of verbal autopsy data,<sup>118,327-330</sup> offer easy-to-implement and reliable analytical methods that are less expensive and quicker than physician certification in traditional verbal autopsy methods.<sup>325,326,331</sup> Implementation and evaluation research related to these methods are necessary to ensure local applicability and community acceptance.



- Validation studies, using clinical information from medical records and forensic reports, are recommended to establish the veracity of cause-of-death attribution on CRVS death certificates and verbal autopsy questionnaires.

#### 6.1.4 Exploring an application of adjustment factors

District-level CRVS cause-of-death data are currently accompanied by explicit warnings of unreliable data,<sup>311</sup> while health-district managers are in critical need of accurate cause-of-death data to prioritise health responses in their jurisdictions. Following the cause-of-death comparison in the linkage study (Paper 4), district managers' need prompted a final research question:

Research question 7: *“Are the linkage study results useful for developing adjustment factors for cause-of-death attribution at the district level?”*

Applying the adjustment factors from the Agincourt HDSS/CRVS linkage study to the cause-of-death data of the surrounding Ehlanzeni district, brought about substantial change in the district cause profile. The reasonable agreement between the adjusted Ehlanzeni profile and the external comparison profile of the surrounding Mpumalanga province (see map in Figure 5.8) suggests improved cause-of-death data at the district level, and that the linkage study results have been useful for developing adjustment factors.

The substantially-changed cause-of-death profile of Ehlanzeni, however, also suggests poor cause-of-death certification at the district level in the first place. Inadequate performance of doctors in the medical certification of causes of death has been found a principal reason for inaccurate data.<sup>332-334</sup> Previous local research studies found (a) numerous errors in the completion of death notification forms,<sup>108,120,122</sup> and (b) improved capacity of medical residents and doctors in cause-of-death certification, following targeted training on correctly completing death certificates.<sup>310,335</sup>

These findings re-emphasize the need for interventions to improve the quality of cause-of-death information on the death certificate. A range of interventions have been implemented in different countries, including periodic peer-auditing of death certificates, implementation of cause-of-death query systems, provision of self-learning educational materials, and training programmes for medical students and physicians.<sup>334</sup>

#### 6.1.4.1 Recommendations

##### **System and mortality data improvement in the CRVS system**

- Interventions elsewhere should be assessed for their applicability and affordability in local settings. Changes to the Health Sciences and Medicine curricula to generate sufficient capacity in medical students for more reliable cause-of-death data, as well as continuous training of physicians, seem to be among the more viable options for South Africa, with a very large and relevant recipient audience. An assessment of the feasibility of mandatory changes to the curriculum and continuous in-service training is therefore recommended, with suggested cooperation among the Departments of Health and Education, and Stats SA.
- Considering local research findings that suggest weak in-hospital medical record keeping,<sup>120,123</sup> physicians also need training in maintaining coherent, reliable and comprehensive medical records, emphasising their utility in diagnosing the underlying cause of death.

##### **Further research**

- The application of the Agincourt HDSS adjustment factors to the district level clearly shows potential to facilitate cause-of-death profiles that provide better planning and prioritizing information to health-district managers. Against this background, and considering the successful work done in Thailand,<sup>232,292</sup> a nationally representative validation study of CRVS cause-of-death data from hospital deaths against reference diagnoses derived from post-mortem physical autopsy reports or good-quality medical records, is recommended.
- The performance of standardised verbal autopsy data-collection and cause-attribution instruments, adapted to the South African setting, should be tested against these reference diagnoses of hospital deaths in a verbal autopsy validation study.
- Subsequently, verbal autopsy diagnoses must be obtained in a nationally representative sample of non-hospital deaths registered into the CRVS system to estimate cause-specific mortality patterns in non-hospital deaths. Any biases observed from the aforementioned verbal autopsy validation study, should be adjusted for in the observed cause-specific mortality patterns found in non-hospital deaths.

Chapter 6 summarised key findings, conclusions and recommendations related to the research questions in the thesis. In view of the primary aim of this body of research, i.e. to assess the quality of mortality data in the CRVS system of South Africa, Chapter 7 presents a consolidation of the thesis findings with an emphasis on CRVS systems. The consolidation of findings is contextualized against recent key messages, opinions, directions and conclusions in relevant international literature.



### 7.1 CRVS systems: Global progress slow, but moving forward with renewed interest

Global progress in CRVS systems has been described as “disappointing and excruciatingly slow”.<sup>336</sup> However, change has begun to occur; the momentum is growing; and planning, implementing, outputs and experience in a number of new and renewed CRVS initiatives are accumulating.<sup>26,36,128,129,337</sup>

These are evidenced by the dramatically-increased trend in peer-reviewed publications addressing civil registration and/or vital statistics over the period 1966 – 2013,<sup>316</sup> and a number of comprehensive, subject-related series of papers in reputable journals.<sup>128,313,338</sup> The long list in the 2015 Lancet Series with over 50 major milestones for CRVS development during the past decade, and the list getting longer for each year as time progressed, further provides evidence of a multitude initiatives and growing momentum. These include the rapidly-increased numbers of expert and high-level meetings; the first costing study on CRVS development in 2013; the announcement of a Global Financing Facility to mobilise support for countries’ CRVS development plans in 2014; and the Bloomberg Philanthropies’ \$100 million *Data for Health Program in Developing Countries* in 2015.<sup>337</sup>

### 7.2 CRVS systems: Globally a preferred source of mortality data, but...

The fundamental principles and attributes of CRVS systems, and the functions, uses, and advantages of well-functioning systems, illustrate why a CRVS system is the recommended and preferred source of mortality data compared to alternative and substitute data sources.

It has been highlighted in the thesis, though, that many low- and middle-income countries do not have functional or well-functioning CRVS systems to cover vital event registration in the total population, neither the infrastructure, resources, and capacity to consolidate such data into vital statistics.<sup>5,337</sup>

### **7.3 CRVS systems in Africa: A prolonged bleak picture**

For Africa, international agency reports since the 1950s,<sup>4,7,30-32,147,326,339-344</sup> big-data studies of global mortality data,<sup>1,3,8,345-347</sup> and scientific journal collections<sup>5,8,9,14,46,337,348-351</sup> have painted a prolonged, particularly bleak picture of sporadic or limited data from poor-performing, incomplete, or no CRVS systems. Persistent and vast under-registration of deaths in the Africa region has been the result of such poor- and non-performance of CRVS systems, and the limited utilisation and utility of substitute data systems.<sup>3,4,6,7,326</sup>

Recently, the 2015 *Counting births and deaths* Lancet series included a performance assessment of CRVS developments in 148 countries for the period 1980 to 2012.<sup>337,351-353</sup> Applying the Vital Statistics Performance Index (VSPI), a composite index that measures six components of mortality data,<sup>17</sup> countries' performance of CRVS systems has been classified into five categories, ranging from "very high" to "very low".<sup>351</sup>

In the Africa region, three countries rated "high" or "very high", one rated "low", and fifteen rated "very low" in the VSPI. The remainder, constituting the majority countries in the continent, though, have not been rated as their civil registration systems were either non-existent or so weak that no mortality data were compiled or reported to international mortality data collections.<sup>351</sup>

### **7.4 CRVS in Africa: Regional initiatives and high-level commitments**

A number of recent regional initiatives, however, have emerged in the continent, including the CRVS workshop organised by the United Nations Economic Commission for Africa in 2009, and the first Conference of African Ministers Responsible for Civil Registration in Ethiopia in 2010.<sup>337</sup> Over 40 ministers in charge of civil registration attended the latter event, urged for an integrated approach to reform and improve CRVS systems on the continent, and made commitments to improve CRVS in their countries.

The need to collate different CRVS initiatives in Africa into a consolidated policy and advocacy framework, led to the emergence of the Africa Programme of Accelerated Improvement of Civil Registration and Vital Statistics (APAI-CRVS). Recognizing the political and socio-economic diversities of African countries, and the need to regard countries' indigenous contexts; and acknowledging the inter-sectorial and inter-disciplinary interfaces

and cooperation of successful CRVS systems, the objective of APAI-CRVS is to provide comprehensive and holistic management and programme guidance to reforming and improving CRVS systems in the region.<sup>315</sup>

## **7.5 CRVS country initiatives: mixed stages and results, but largely moving forward**

At country-level, too, there has been new and renewed interest and prioritisation in strengthening CRVS systems. Several countries, including Botswana, Ethiopia, Laos, Cambodia, and Yemen are currently completing national CRVS assessments, or formulating or implementing improvements plans.<sup>337</sup> Many have embarked on initiatives towards improved civil registration and/or vital statistics, such as Mozambique, Sri Lanka, Bangladesh, Kazakhstan, Albania, and Fiji. Some, including South Africa, Brazil, Egypt, China, Taiwan and Malaysia, have seen remarkable improvements over very short periods of time.<sup>351</sup>

## **7.6 CRVS in South Africa**

### **7.6.1 Progress and improved ratings over time**

Despite the deficient CRVS situation in most African countries (Section 7.3), this body of research has found that mortality data sources for South Africa are many and varied. Evaluating the quality of national civil registration mortality statistics, Paper 2 reports satisfactory ratings for six out of nine quality criteria for the period 1997 – 2007. Substantial improvements over a relatively short period of time have been reported in system performance as gauged by the rapidly-increased coverage and completeness of death registration, as well as improved availability, timeliness, and sub-national disaggregation of mortality data, compared to pre-democracy data<sup>119,136,354-356</sup> and the country's 1996 data evaluated in earlier WHO evaluations.<sup>8,9</sup>

The developments in death registration and improvements in vital statistics compilation and dissemination documented in Papers 1 and 2, have advanced such that it would shift South Africa's "low quality" rating for 1996 data in Mathers et al<sup>8</sup> to "medium-high quality". Such improvements for South Africa have been confirmed in international reports.<sup>44,326,337</sup> Improvements have also been quantified in the recent VSPI assessment,<sup>351</sup> showing rapid and

extensive improvement from a score of 7.8 (very low quality) for the period 1990-94, to 46.9 for 1995-99 (low), 63.2 for 2000-04 (medium), and 70.5 for 2005-09 (high).

In addition, standard mortality data items for maternal, child, adult, and all-cause mortality are available from various alternative data collections, including three population censuses; three demographic surveillance systems; a number of national surveys—some with annual data over several years; mortality audits; and disease notification programmes.

## **7.6.2 Key persisting challenges**

The findings in the thesis, collated from a multitude data sources, literature references, and the record-linkage field study, clearly points to two key components of CRVS that remain a challenge to the optimal use of mortality data for health planning: (a) accuracy of cause-of-death data, and (b) completeness of death registration.

### **7.6.2.1 Accuracy of cause-of-death data**

The accuracy of cause-of-death data, prior to undertaking the research for the thesis, has been identified as a weakness in the country's CRVS data collective, mostly in small urban sub-population studies,<sup>120,121,123</sup> but also in national studies covering a selected year or two years of death data.<sup>93,321,357</sup> In Paper 2, the quality of national cause-of-death data for an 11-year period, 1997 - 2007, have been investigated. In Paper 4, four years' cause-of-death data have been compared in a record-linkage study, matching Agincourt HDSS and CRVS data for deaths that occurred in the period 2006 - 2009. These papers, of which the analyses collectively covered approximately 5.4 million deaths, report considerable shortcomings in cause-of-death data, and highlight their limited utility for local health decision-making and resource prioritisation.

The shortcomings include misattribution of HIV/AIDS deaths in the CRVS data, the extent of which in the Agincourt/Stats SA study (89%), are similar to that in other studies (73 – 92%).<sup>121,123,162,321</sup> These findings indicate substantial quality problems with CRVS cause-of-death data in different settings over different periods of time.

For injuries, quality challenges identified in CRVS data that have been used for the National Burden of Disease Studies for 2000 and 2010,<sup>93,358</sup> were confirmed in the linkage study findings. Paper 4 reports systematic and substantial under-recording of homicide (83% not reported) and suicide (100% not reported) deaths in the CRVS data; and the systematic



attribution of external causes of injury deaths to unspecified, undetermined and ‘other’ injury codes (76%). These findings point to the likelihood of being unable to plan for appropriate health care responses and injury prevention interventions.

#### **7.6.2.2 Completeness of death registration**

As with poor content validity of cause-of-death data, incomplete death registration, when not adjusted for under-registration, can substantially affect the representativeness and utility of mortality data. Despite a trend of ongoing improvements in registration completeness of South Africa’s deaths since the early nineties, Paper 2 reports that approximately 10% of all recent deaths, and 43% of deaths in children aged 1 – 4 years for 1996 - 2006, remained unregistered (Paper 2).

A high level of incomplete death registration in children 1 – 4 years has also been suggested in the record linkage study, reporting a significantly lower matching likelihood among these children compared to other age groups. Other sub-groups missed include people in the lowest wealth quintile, non-citizens, persons with temporary residency status, and those with no education, pointing out vulnerable groups to be targeted to enhance equity and strengthen completeness (Paper 3).

### **7.7 Concluding call: Reforming a “data-rich, but information-poor” situation**

The thesis documents considerable progress from a CRVS system paralysed by disintegration, disparity and delay to one with potential to offer inclusive and timely sub-national all-cause mortality statistics that could be adjusted and used for demographic and health analysis. Additionally, cause-of-death data, certified and coded according to international standards into 3-digit ICD-10 codes, are available by regional and socio-demographic disaggregation for the past 17 years. Moreover, many years of mortality data, in many repositories, are available from many alternative data collections.

However, the findings in this thesis point to a data-rich, but information-poor mortality data situation. It is clear that inaccuracies in cause-of-death data constitute a substantial problem in South Africa’s CRVS system, and represent a considerable challenge to evidence-based prioritisation and suitable resource distribution in health. Data use in alternative collections, has been hindered by some shortcomings described in Paper 1, diluting a promising scenario

for improved population health monitoring and public health planning from an evidence base informed by multiple sources.

In addition, there is a critical need to carefully think about the satisfactory ratings for the country's mortality data and data-systems performance, as done in the assessments by Joubert et al<sup>131</sup> and Mikkelsen et al.<sup>351</sup> The overall ratings, in particular, may distort the true situation of specific, critical components of data and system performance. Joubert et al found six out of nine criteria satisfactory, however, a crucially important aspect of mortality data and health intelligence, i.e. cause-of-death data, have been identified with substantial weaknesses in multiple empirical and expertly-modelled studies reported in the very paper.<sup>131</sup> Mikkelsen and colleagues, using a clearly-valuable, single composite metric, the VSPI, rated South Africa's CRVS system performance and data utility as "high". However, the quality and utility of the cause-of-death data remain of limited quality and utility as indicated before.

Reporting the Global Burden of Disease 2010 results for mortality from 235 causes of death, Lozano et al write that a country's health information system is not providing accurate cause-of-death information by age and sex, "major reforms are required to provide health planners with this essential health intelligence".

Perhaps the time has come for the relevant South African role players and leaders to heed the comment about *major reforms* towards providing health planners with valid cause-of-death data. Given the population's particularly heavy<sup>93,358-360</sup> burden of disease and the associated cost, improving cause-of-death data will be an appropriate if not crucial investment in essential health intelligence.

This can be achieved through strong public and political awareness of the value of reliable and valid mortality data for better health solutions; bold vision to initiate and harmonise the activities and inputs of all role players—including the agencies committed to the CRVS Scaling Up Plan<sup>326</sup>; and unwavering authority to compel and support collaboration and coordination towards good-quality mortality data.<sup>316</sup>

Nine of the 13 health-related sub-goals proposed by the Open Working Group for the Sustainable Development Goals (SDGs) relate to mortality and causes of death: reduction of the global maternal mortality ratio to <70; reduction of neonatal mortality to <12/1000; reduction of under-5 mortality to <25/1000; 90% reduction in each of AIDS-related, tuberculosis, and malaria deaths; substantially reduce number of deaths from hazardous

chemicals, and air, water, and soil contamination; one-third reduction of premature mortality from NCDs; and halving the number of global traffic deaths.<sup>316</sup> Further, the goal of the *Global Civil Registration and Vital Statistics: Scaling-up Investment Plan 2015-2024*<sup>326</sup> is “universal civil registration of births, deaths and other vital events including reporting cause of death, and access to legal proof of registration for all individuals by 2030”. The first component of the plan is focused on strengthening national CRVS systems, including the improvement of death reporting and determining causes of death.<sup>326</sup> In addition, universal health coverage has been identified a target goal of the SDGs,<sup>361</sup> and South Africa’s national government has committed to ensure that universal health coverage is achieved.<sup>362</sup>

To achieve sustainable development and universal health coverage for all, it will be important to know who “all” are and what the risks to their health are.<sup>361</sup> It may therefore be imperative, rather than appropriate, to plan for major reforms towards improved completeness of vital-event registration, and reliable and valid cause-of-death data. By generating both the numerators and denominators for various population and public health indicators, improved CRVS systems should clearly be central in measuring the SDGs; appraising progress on the objectives of the CRVS Scaling-up Investment Plan; as well as calculating health and mortality risks and resource allocation for universal health coverage.

While the final content and shape of these imminent measure-intensive initiatives are still being deliberated, a window of opportunity exist, and should be seized,<sup>316</sup> to advocate, strategize and act for strengthening CRVS in South Africa. The fundamentally-important role of good-quality cause-of-death data in reducing inequity and improving health should be recognised by relevant sectors, spanning civil registration, health, statistics, education, and civil society, and strategic research and interventions aiming at breaking the cycle of poor-quality cause-of-death data should be prioritized.

## **7.8 Contribution to existing knowledge and relevance of the research**

This body of research includes the first mortality data source review in 20 years, documenting the availability, strengths and limitations of the data sources in the country. The data source review prior to this study, covered vital statistics from civil registration only, whereas the review in this thesis covered censuses, national surveys, mortality audits, disease registers, health and demographic surveillance systems, and the rapid mortality surveillance initiative.

A review was conducted to identify the criteria used for evaluating the quality of national mortality statistics, resulting in the first independent, comprehensive evaluation of national CRVS data for South Africa. The review highlighted cause-of-death information as a particularly weak part of CRVS data, and highlighted that although the completeness of death registration has reached satisfactory levels in the adult population, non-registration of deaths constitute a particular problem in children 1- 4 years old.

A number of studies have evaluated CRVS,<sup>8-10,17,126,150,152</sup> each with useful, at times overlapping evaluation criteria. Chapter 3 offers a unique table, Table 3.1, composed of all the criteria used in different country-level evaluations, contributing to knowledge a handy snapshot of these criteria, and their frequency of occurrence, over the past 25 years. Furthermore, the thesis contributes a comprehensive framework with 22 evaluation criteria that can be used at country level (Table 3.2). This framework brings together existing knowledge of criteria used before, and adds new knowledge in a novel presentation with an extended list of data attributes and criteria, coupled with simple assessment questions and suggested thresholds and responses that have possible utility as a road map in countries that are initiating or developing a CRVS system.

The linkage study between CRVS and HDSS data is the first in South Africa, and possibly in sub-Saharan Africa, to assess the feasibility and utility of linking such data. The resultant data are useful for assessing selected population and individual health measures, and hold potential to improve rural data quality. The study adds to the literature what is possibly the first study to successfully link HDSS and official routine CVSR mortality data, despite operational and methodological challenges. A method has been documented of how to overcome these and data confidentiality challenges. A promising, low-cost validation method that provided empirical evidence about the quality and utility of CRVS cause-of-death data, has shown considerable potential in a rural area. Evidence from the study advances knowledge about the quality of CRVS cause-of-death data in the area, and concur with findings from urban and national studies.

It furthermore illustrates the potential of HDSS sites as independent data sources against which to periodically evaluate the quality of CRVS cause-of-death data, and the performance of civil registration in terms of completeness of death registration.

The linkage-study analysis offers the first empirical evidence of large-scale, systematic misclassification of HIV/AIDS in CRVS data in a rural setting in South Africa. The extent of HIV deaths being misattributed in the CRVS data (89%), are similar to that in earlier studies (73 – 92%),<sup>121,123,162</sup> pointing to the reliability, and possible future use, of the HDSS surveillance data against which to assess data quality in CRVS.

The systematic and substantial under-recording of homicide and suicide deaths in the CRVS data are likely to impede the recognition of a serious and fatal injury problem, thereby obstructing the implementation of appropriate injury prevention and safety promotion strategies. In addition, Paper 4 highlights the systematic non-determination of the external causes of injury deaths in the official data—due to the absence of a field for the intent and manner of injury death on the official death notification form—and the resultant likelihood of being unable to plan for appropriate health care responses and prevention interventions. In contrast, the HDSS data hold particular value for providing important detail on external causes that would better inform interventions aimed at reducing or avoiding fatal injuries.

The linkage exercise revealed the importance of standardising variables and their labelling across data sources. Standardised labels of personal identifiers (such as gender, date of birth and death), and standardised geographic identifiers (such as town, city, tribal area, and district names) will simplify matching exercises and optimize comparative analysis and analytical techniques such as capture-recapture analysis.

Relevant to the country's cause-of-death certification approach, and congruent with research findings, recommendations have been made towards addressing the identified immediate and underlying challenges to good-quality cause-of-death data (Papers 2 and 4). The recommendations include the conduct of a large-scale national validation of CRVS cause-of-death data against medical, forensic and verbal autopsy records (Section 6.1.4). In view of the plausible changes in district-level cause-of-death profiles facilitated by the application of adjustment factors generated in the linkage study, a national validation study for South Africa has recently been conceptualized. A concept proposal, activity plan, and budget have been submitted for funding over a 4-year period. Local and international research partners and stakeholders have acknowledged the critical need for and potential value of such a study. Partial funding has been made available by an international donor in view of the project's potential to strengthen strategic information for the country.

In the concluding paper<sup>316</sup> of the recent four-paper Lancet CRVS series, the compilation and accessibility of research findings to policy-making, programming and practice are recommended. The linkage sub-study in this research project has been done in collaboration with Statistics South Africa, the official government agency responsible for collating, coding, and disseminating cause-of-death data, and the study findings of this project have been made available to key leaders in the agency.

Finally, the concluding paper<sup>316</sup> also made it clear that reliable data for deaths and causes of death are indispensable for effectively addressing the health and development challenges of the coming decades. The thesis and its publications made several locally-relevant recommendations towards enhancing such awareness; improving the quality of cause-of-death data; and suggest that the time perhaps has come for major reforms towards providing health planners with good-quality cause-of-death data.

## References

1. Naghavi M, Wang H, Lozano R, Davis A, Liang X, Zhou M, et al. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014; 385(9963): 117-71.
2. United Nations. Principles and recommendations for a vital statistics system. Revision 3. Final Draft. New York: United Nations, 2013. Accessed: 13 May 2014. Available: [http://unstats.un.org/unsd/demographic/standmeth/principles/unedited\\_M19Rev3en.pdf](http://unstats.un.org/unsd/demographic/standmeth/principles/unedited_M19Rev3en.pdf)
3. Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, Levin-Rector A, et al. Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859): 2071-94.
4. United Nations. Status of civil registration and vital statistics in the SADC region. New York, US: United Nations, 2010. Accessed: 17 Sep 2014. Available: <https://unstats.un.org/unsd/demographic/CRVS/Technical%20report%20SADC%20final%20v2.pdf>
5. Setel PW, Macfarlane SB, Szreter S, Mikkelsen L, Jha P, Stout S, et al. A scandal of invisibility: making everyone count by counting everyone. *Lancet*. 2007; 370(9598): 1569-77.
6. United Nations. Approaches to the collection of mortality data in the context of data needs. Data bases for mortality measurement: Proceedings of the United Nations/World Health Organization Working Group on Data Bases for Measurement of Levels, Trends, and Differentials in Mortality; Bangkok, 1981; ST/ESA/Series/A/84, [http://www.un.org/esa/population/publications/UN\\_1984\\_Data\\_Bases\\_for\\_Mortality\\_Measurement/UN1984\\_Databases\\_ch5\\_Approaches\\_Data\\_Needs.pdf](http://www.un.org/esa/population/publications/UN_1984_Data_Bases_for_Mortality_Measurement/UN1984_Databases_ch5_Approaches_Data_Needs.pdf). New York: United Nations; 1984.
7. United Nations. Principles for a vital statistics system: recommendations for the improvement and standardization of vital statistics. New York: United Nations, 1953. Accessed: 13 May 2014. Available: [http://unstats.un.org/unsd/demographic/standmeth/principles/Series\\_M19en.pdf](http://unstats.un.org/unsd/demographic/standmeth/principles/Series_M19en.pdf)
8. Mathers CD, Ma Fat D, Inoue M, Rao C, Lopez AD. Counting the dead and what they died from: an assessment of the global status of cause of death data. *Bull World Health Organ*. 2005; 83(3): 171-7C.
9. Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, Rao C, et al. Civil registration systems and vital statistics: successes and missed opportunities. *Lancet*. 2007; 370(9599): 1653-63.

10. Rao C, Lopez AD, Yang G, Begg S, Ma J. Evaluating national cause-of-death statistics: principles and application to the case of China. *Bull World Health Organ.* 2005; 83(8): 618-25.
11. United Nations. Strategy for conducting assessment of and planning for national civil registration and vital statistics systems. New York: United Nations, 2012. Accessed: 24 May 2014. Available: [http://www.uneca.org/sites/default/files/page\\_attachments/acs12-0026-ore-improving-crvs-systems-in-africa\\_rev\\_july26072012.pdf](http://www.uneca.org/sites/default/files/page_attachments/acs12-0026-ore-improving-crvs-systems-in-africa_rev_july26072012.pdf)
12. Bourne DE. Sources of South African mortality data, 1910 - 1992. *S Afr Med J.* 1995; 85(10): 1020-2.
13. Groenewald P, Msemburi W, Neethling I, Day C, Tuoane-Nkhazi M, Bradshaw D. Burden of Disease. In: Massyn N, Day C, Barron P, Haynes R, English R, Padarath A, editors. District Health Barometer 2011/12, <http://www.hst.org.za/publications/district-health-barometer-201112>. Durban, South Africa: Health Systems Trust; 2013. p. 131-50.
14. Hill K, Lopez AD, Shibuya K, Jha P, on behalf of the MoVE writing group. Interim measures for meeting needs for health sector data: births, deaths, and causes of death. *Lancet.* 2007; 370(9600): 1726-35.
15. United Nations. Handbook of vital statistics methods. New York: United Nations, 1955. Accessed: 25 May 2014. Available: [http://unstats.un.org/unsd/demographic/standmeth/handbooks/Series\\_F7en.pdf](http://unstats.un.org/unsd/demographic/standmeth/handbooks/Series_F7en.pdf)
16. United Nations. Handbook of vital statistics systems and methods: review of national practices. New York: United Nations, 1985. Accessed: 25 May 2014. Available: <http://unstats.un.org/unsd/pubs/gesgrid.asp?id=27>
17. Phillips D, Lozano R, Naghavi M, Atkinson C, Gonzalez-Medina D, Mikkelsen L, et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. *Popul Health Metr.* 2014; 12(1): 14.
18. World Health Organisation, The University of Queensland Health Information Systems Knowledge Hub. Rapid assessment of national civil registration and vital statistics systems. Geneva: World Health Organization, 2010. Accessed: 14 Dec 2013. Available: [http://www.uq.edu.au/hishub/docs/WP02/WP\\_02.pdf](http://www.uq.edu.au/hishub/docs/WP02/WP_02.pdf)
19. World Health Organisation, The University of Queensland Health Information Systems Knowledge Hub. Improving the quality and use of birth, death and cause-of-death information: guidance for a standards-based review of country practices. Geneva: World Health Organization, 2010. Accessed: 22 Dec 2012. Available: [http://www.uq.edu.au/hishub/docs/WP01/WP\\_01.pdf](http://www.uq.edu.au/hishub/docs/WP01/WP_01.pdf)
20. United Nations, World Health Organisation. Generating reliable cause-of-death information within a civil registration and vital statistics system: an overview of concepts, standards and implementation. New York: United Nations, 2012. Accessed: 24 May 2014. Available:



[http://www.uneca.org/sites/default/files/page\\_attachments/generating-reliable-causes\\_of\\_death.pdf](http://www.uneca.org/sites/default/files/page_attachments/generating-reliable-causes_of_death.pdf)

21. Wang H, Dwyer-Lindgren L, Lofgren KT, Rajaratnam JK, Marcus JR, Levin-Rector A, et al. Age-specific and sex-specific mortality in 187 countries, 1970-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859): 2071-94. Epub 2012/12/19.
22. United Nations. The Millenium Development Goals Report 2011. New York: United Nations, 2011. Accessed: 7 Feb 2012. Available: [http://www.un.org/millenniumgoals/pdf/\(2011\\_E\)%20MDG%20Report%202011\\_Book%20LR.pdf](http://www.un.org/millenniumgoals/pdf/(2011_E)%20MDG%20Report%202011_Book%20LR.pdf)
23. Hill K. Making deaths count. *Bull World Health Organ*. 2006; 84(3): 162.
24. United Nations Development Programme. Human Development Report 1990. New York: United Nations Development Programme; 1990.
25. United Nations Development Programme. Human Development Report 2013. The rise of the South: human progress in a diverse world. New York: United Nations Development Programme, 2013. Accessed: 28 May 2014. Available: <http://hdr.undp.org>
26. Commission on Information and Accountability for Women's and Children's Health. Keeping promises, measuring results. Geneva: World Health Organization, 2011. Accessed: 29 May 2014. Available: [http://www.everywomaneverychild.org/images/content/files/accountability\\_commission/final\\_report/Final\\_EN\\_Web.pdf](http://www.everywomaneverychild.org/images/content/files/accountability_commission/final_report/Final_EN_Web.pdf)
27. Pearce N, Ebrahim S, McKee M, Lamptey P, Barreto ML, Matheson D, et al. The road to 25x25: how can the five-target strategy reach its goal? *The Lancet Global Health*. 2014; 2(3): e126-e8.
28. United Nations Statistics Division. Topics of the Demographic Yearbook series: 1948 - 2004 New York: United Nations; c2013 [Updated: n.d.]. Available: <http://unstats.un.org/unsd/demographic/products/dyb/dybpubs.pdf> Cited: 30 May 2014.
29. United Nations Statistics Division. Demographic Yearbook. New York: United Nations, 2013. Accessed: 31 May 2014. Available: <http://unstats.un.org/unsd/demographic/products/dyb/default.htm>
30. United Nations. Population and vital statistics report. New York: United Nations, 2014. Accessed: 29 May 2014. Available: [http://unstats.un.org/unsd/demographic/products/vitstats/sets/Series\\_A\\_2014.pdf](http://unstats.un.org/unsd/demographic/products/vitstats/sets/Series_A_2014.pdf)
31. United Nations Population Division. United Nations Expert Group Meeting on Current Issues in the Estimation of Adult Mortality, 26-27 October 2006: Report of the Meeting. Report No. ESA/P/WP.203. New York: United Nations, 2007. Accessed: 12 Aug 2010. Available: [http://www.un.org/esa/population/meetings/EGM\\_AdultMort/EGM\\_AdultMort.pdf](http://www.un.org/esa/population/meetings/EGM_AdultMort/EGM_AdultMort.pdf)

32. United Nations Population Division. World Mortality Report 2005. New York: United Nations, 2006. Accessed: 12 Aug 2010. Available: <http://www.un.org/esa/population/publications/worldmortality/WMR2005.pdf>
33. World Health Organization. First consultation of the High Level Advisory Panel on Health Statistics. Geneva, Switzerland: World Health Organization, 2005. Accessed: 12 Aug 2010. Available: <https://www.who.int/healthinfo/statistics/programmeconsultationreport1.pdf>
34. Bhir A, Bhutta Z, Binka F, Black R, Bradshaw D, Garnett G, et al. Better health statistics are possible. *Lancet*. 2006; 367(9506): 190-3.
35. Health Information Systems Knowledge Hub, Institute for Health Metrics and Evaluation. Improving the quality and use of health information systems: essential strategic issues. Brisbane, Australia: The University of Queensland, 2009. Accessed: 20 Dec 2013. Available: <http://www.uq.edu.au/hishub/docs/WP05/HISHUB-WP5-Full-12-WEB12Oct12.pdf>
36. Mikkelsen L. Rapid assessment of vital statistics systems: evaluation of the application of the World Health Organization/Health Information Systems Knowledge Hub tool in 26 countries in the Asia–Pacific region. Brisbane, Australia: The University of Queensland, 2010. Accessed: 20 Dec 2013. Available: <http://www.uq.edu.au/hishub/docs/WP10/HISHUB-WP10-08-WEB-3Oct12%20A.pdf>
37. Editorial. Stumbling around in the dark. *Lancet*. 2005; 365(9476): 1983.
38. World Health Organization. Health Metrics Network: Framework and standards for the development of country health information systems. Geneva: World Health Organization; 2006.
39. Boerma JT. Civil and vital registration systems: A perspective from the health sector: Health Metrics Network/WHO. 6th Africa Symposium on Statistical Development (ASSD); 31 October – 2 November 2010; Cairo, Egypt.
40. Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Medicine*. 2010; 7(4).
41. Murray C, Lopez A, editors. The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and projected to 2020. Cambridge, MA: Harvard University Press on behalf of the WHO and World Bank; 1996.
42. Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJL. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet*. 2006; 367(9524): 1747-57.
43. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859): 2095-128. Epub 2012/12/19.

44. Lopez AD, AbouZahr C, Shibuya K, Gollogly L. Keeping count: births, deaths, and causes of death. *Lancet*. 2007; 370(9601): 1744-6.
45. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859): 2197-223. Epub 2012/12/19.
46. AbouZahr C, Cleland J, Coullare F, Macfarlane SB, Notzon FC, Setel P, et al. The way forward. *Lancet*. 2007; 370(9601): 1791-9.
47. Murray CJL, Lopez AD, Shibuya K, Lozano R. Verbal autopsy: advancing science, facilitating application. *Popul Health Metr*. 2011; 9(1): 18.
48. Murray C, Lopez A, Black R, Ahuja R, Ali S, Baqui A, et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. *Popul Health Metr*. 2011; 9(1): 27.
49. Lopez AD, Mathers C. Measuring the global burden of disease and epidemiological transitions, 2002 - 2030. *Ann Trop Med Parasitol*. 2006; 100: 481-99.
50. Statistics South Africa. Mortality and causes of death in South Africa, 2013: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa; 2014.
51. Statistics South Africa. Mortality and causes of death in South Africa, 2010: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa; 2013.
52. Bradshaw D, Groenewald P, Bourne DE, Mahomed H, Nojilana B, Daniels J, et al. Making COD statistics useful for public health at local level in the city of Cape Town. *Bull World Health Organ*. 2006; 84(3): 211-7C.
53. Bradshaw D, Dorrington R, Laubscher R. Rapid mortality surveillance report 2011. Cape Town: South African Medical Research Council; 2012.
54. Setel PW, Sankoh O, Rao C, Velkoff VA, Mathers C, Gonghuan Y, et al. Sample registration of vital events with verbal autopsy: a renewed commitment to measuring and monitoring vital statistics. *Bull World Health Organ*. 2005; 83(8): 611-7.
55. Dorrington RE, Moultrie TA, Timæus IM. Estimation of mortality using the South African Census 2001 data. Cape Town: Centre for Actuarial Research, University of Cape Town, 2004. Accessed: 11 Aug 2010. Available: [http://www.commerce.uct.ac.za/Research\\_Units/CARE/Monographs/Monographs/Mono11.pdf](http://www.commerce.uct.ac.za/Research_Units/CARE/Monographs/Monographs/Mono11.pdf)
56. Moultrie T, Dorrington RE. National Income Dynamics Study (NIDS): Demography: Analysis of the NIDS wave 1 dataset. Discussion Paper no. 9. Cape Town: University of Cape Town, 2009. Accessed: 8 March 2012. Available: <http://www.nids.uct.ac.za/home/index.php?Nids-Documentation/discussion-papers.html>

57. Department of Health (South Africa), South African Medical Research Council, Macro International Inc. South Africa Demographic and Health Survey 1998. Full Report. Pretoria: Department of Health, 2002. Accessed: 18 July 2012. Available: <http://www.measuredhs.com/publications/publication-FR131-DHS-Final-Reports.cfm>
58. Department of Health (South Africa), South African Medical Research Council, Macro International Inc. South Africa Demographic and Health Survey 2003. Full Report. Pretoria: Department of Health, 2007. Accessed: 18 July 2012. Available: <http://www.measuredhs.com/publications/publication-FR206-DHS-Final-Reports.cfm>
59. Ndirangu J, Newell M, Tanser F, Herbst A, Bland R. Decline in early life mortality in a high HIV prevalence rural area of South Africa: evidence of HIV prevention or treatment impact? *AIDS*. 2010; 24: 593 - 602.
60. Herbst AJ, Cooke GS, Bärnighausen T, KanyKany A, Tanser F, Newell M. Adult mortality and antiretroviral treatment roll-out in rural KwaZulu-Natal, South Africa. *Bull World Health Organ*. 2009; 87(10): 754-62.
61. Kahn K, Collinson M, Hargreaves J, Clark S, Tollman S. Socio-economic status and child mortality in a rural sub-district of South Africa. *Measuring health equity in small areas - Findings from demographic surveillance sites*. 2005: 67-86.
62. Kahn K, Garenne ML, Collinson MA, Tollman SM. Mortality trends in a new South Africa: Hard to make a fresh start. *Scand J Public Health*. 2007; 35(Suppl 69): 26-34.
63. Tollman SM, Kahn K, Sartorius B, Collinson MA, Clark SJ, Garenne ML. Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study. *Lancet*. 2008; 372(9642): 893-901.
64. Hill K, Choi Y. Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors. The Adult Mortality in Developing Countries Workshop; 8 - 11 July 2004; Marconi Center, Marin County, California (Paper).
65. Dorrington RE, Timæus IM, Moultrie TA. Death distribution methods for estimating adult mortality: Sensitivity analysis with simulated data errors, revisited. Fifth African Population Conference; 10-14 Dec 2007; Arusha, Tanzania (Paper).
66. United Nations. Handbook on civil registration and vital statistics systems: Management, operation and maintenance. Studies in Methods. New York: United Nations; 1998.
67. Shryock HS, Siegel JS, and Associates. The methods and materials of demography. Condensed edition. San Diego, California: Academic Press; 1976.
68. Rowland DT. Demographic methods and concepts. New York: Oxford University Press; 2006.
69. United Nations Statistics Division. Revising current international standards on civil registration and vital statistics: proposed outline. New York: United Nations, 2011.

70. Department of Government Communication and Information System. The Land and its People. In: Burger D, editor. South Africa Yearbook 2009/10. Pretoria: Government of the Republic of South Africa; 2010.
71. 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*. 2009; 374(9692): 817-34.
72. Khalfani AK, Zuberi T, Bah S, Lehohla P. Population statistics. In: Zuberi T, Sibanda A, Udjo EO, editors. The Demography of South Africa. Armonk, New York and London, England: M.E. Sharpe; 2005.
73. Statistics South Africa. Mid-year population estimates, 2013. Pretoria: Statistics South Africa, 2013.
74. Statistics South Africa. The South Africa I know, the home I understand. Pretoria: Statistics South Africa, 2012. Accessed: 13 June 2014. Available: [http://www.statssa.gov.za/Census2011/Products/Census\\_2011\\_Pictorial.pdf](http://www.statssa.gov.za/Census2011/Products/Census_2011_Pictorial.pdf)
75. Dorrington RE. Alternative South African mid-year estimates, 2013. Cape Town: University of Cape Town, 2013.
76. Hale W, Joubert J, Kalula S. Aging populations and chronic illness. In: Markle W, Fisher M, Smego R, editors. Understanding Global Health, 2nd edition. New York and others: McGraw-Hill Education; 2013. p. 363-86.
77. Joubert JD, Bradshaw D. Population ageing and health challenges in South Africa. In: Steyn K, Fourie J, Temple N, editors. Chronic diseases of lifestyle in South Africa: 1995-2005, <http://www.mrc.ac.za/chronic/cdl1995-2005.pdf>. Cape Town: South African Medical Research Council; 2006. p. 204-19.
78. Statistics South Africa. Census 2011. Pretoria: Statistics South Africa, 2012. Accessed: 14 June 2014. <http://www.statssa.gov.za/publications/P03014/P030142011.pdf>
79. Kok P, Collinson M, Van Tonder L, Roux N, Garenne M. Migration and urbanisation. Pretoria: Statistics South Africa, 2006. Accessed: 14 June 2014. Available: <http://www.statssa.gov.za/PublicationsHTML/Report-03-04-02/html/Report-03-04-02.html>
80. Swart L-A, Laher H, Seedat M, Gantchev G. Urban and rural differences in child injury deaths in South Africa: A one-year review. *African Safety Promotion Journal*. 2012; 10(2): 28-40.
81. South African Cities Network. Towards resilient cities: A reflection on the first decade of a democratic and transformed local government in South Africa 2001–2010. Johannesburg: 2011. Accessed: 15 June 2014. Available: <http://www.sacities.net/knowledge-centre/research/publications/25-what-we-do/socr/607-towards-resilient-cities>
82. Schellack N, Meyer JC, Gous AGS, Winters C. Health and economic context. *S Afr Med J*. 2011; 101(8): 558-61.

83. The World Bank Group. Data: GDP per capita (current US\$) c2014.
84. The World Bank World dataBank: World Development Indicators and Global Development Finance [Internet]. The World Bank. 2010 [cited 23 September 2010]. Available from: <http://data.worldbank.org/country/south-africa>.
85. The World Bank Group. Databank: The GINI Index New York, USA: The World Bank; c2015. Available: <http://data.worldbank.org/indicator/SI.POV.GINI> Cited: 16 April 2015.
86. The World Bank Group. Data: Unemployment, total c2014. Available: <http://data.worldbank.org/indicator/SL.UEM.TOTL.ZS> Cited: 13 June 2014.
87. The World Bank. South Africa at a glance [Internet]. New York: The World Bank; 2014 [Updated: 15 March 2014]. Available: [http://devdata.worldbank.org/AAG/zaf\\_aag.pdf](http://devdata.worldbank.org/AAG/zaf_aag.pdf) Cited: 14 June 2014.
88. Kimuna SR, Makiwane M. Older people as resources in South Africa: Mpumalanga households. *J Aging Soc Policy*. 2007; 19(1): 97-114. Epub 2007/03/10.
89. Booyesen F. Social grants as safety net for HIV/AIDS-affected households in South Africa. *Journal of Social Aspects of HIV/AIDS Research Alliance (SAHARA)*. 2004; 1(1): 45-56. Epub 2007/07/03.
90. Tamasane T, Head J. The quality of material care provided by grandparents for their orphaned grandchildren in the context of HIV/AIDS and poverty: a study of Kopanong municipality, Free State. *Journal of Social Aspects of HIV/AIDS Research Alliance (SAHARA)*. 2010; 7(2): 76-84. Epub 2011/03/17.
91. Case A, Menendez A. Does money empower the elderly? Evidence from the Agincourt demographic surveillance site, South Africa. *Scandinavian Journal of Public Health Supplement*. 2007; 69(August): 157-64. Epub 2007/09/06.
92. UNICEF. South Africa Annual Report 2012. New York: UNICEF, 2012. Accessed: 15 June 2014. Available: [http://www.unicef.org/southafrica/SAF\\_resources\\_annual2012.pdf](http://www.unicef.org/southafrica/SAF_resources_annual2012.pdf)
93. Bradshaw D, Groenewald P, Laubscher R, Nannan N, Nojilana B, Norman R, et al. Initial burden of disease estimates for South Africa, 2000. *S Afr Med J*. 2003; 93(9): 682-8.
94. Mayosi BM, Lawn JE, van Niekerk A, Bradshaw D, Abdool Karim SS, Coovadia HM. Health in South Africa: changes and challenges since 2009. *Lancet*. 2012; 380(9858): 2029-43. Epub 2012/12/04.
95. Chopra M, Lawn JE, Sanders D, Barron P, Karim SSA, Bradshaw D, et al. Achieving the health Millennium Development Goals for South Africa: challenges and priorities. *Lancet*. 2009; 374(9694): 1023-31.
96. Cooper D, Morroni C, Orner P, Moodley J, Harries J, Cullingworth L, et al. Ten years of democracy in South Africa: Documenting transformation in reproductive health policy and status. *Reprod Health Matters*. 2004; 12(24): 70-85.

97. Van Rensburg HC, Harrison D. History of Health Policy. In: Trust HS, editor. South African Health Review 1995. Durban: Health Systems Trust; 1995.
98. National Department of Health (South Africa). Strategic plan for the prevention and control of non-communicable diseases 2013-2017. Pretoria: National Department of Health, 2013. Accessed: 24 Sep 2013. Available: <http://www.health-e.org.za/wp-content/uploads/2013/09/NCDs-STRAT-PLAN-CONTENT-8-april-proof.pdf>
99. Bradshaw D, Pillay-Van Wyk V, Laubscher R, Nojilana B, Groenewald P, Nannan N, et al. Cause of death statistics for South Africa: challenges and possibilities for improvement. Cape Town: South African Medical Research Council, 2010. Accessed: 14 April 2012. Available: [http://www.mrc.ac.za/bod/cause\\_death\\_statsSA.pdf](http://www.mrc.ac.za/bod/cause_death_statsSA.pdf)
100. Nannan N, Timæus IM, Laubscher R, Bradshaw D. Levels and differentials in childhood mortality in South Africa, 1977-1998. *J Biosoc Sci.* 2007; 39(4): 613-32.
101. Nannan N, Dorrington RE, Laubscher R, Zinyakatira N, Prinsloo M, Darikwa T, et al. Under-5 mortality statistics in South Africa: Shedding some light on the trend and causes, 1997–2007. Cape Town: South African Medical Research Council, 2012. Accessed: 18 Dec 2012. Available: <http://www.mrc.ac.za/bod/MortalityStatisticsSA.pdf>
102. INDEPTH Network. INDEPTH Projects [Updated: 22 Oct 2014]. Available: <http://www.indepth-network.org/> Cited: 23 Oct 2014.
103. Chandramohan D, Shibuya K, Setel P, Cairncross S, Lopez AD, Murray CJL, et al. Should data from Demographic Surveillance Systems be made more widely available to researchers? *PLoS Med.* 2008; 5(2): e57.
104. Kahn K, Collinson MA, Gómez-Olivé FX, Mokoena O, Twine R, Mee P, et al. Profile: Agincourt Health and Socio-demographic Surveillance System. *Int J Epidemiol.* 2012; 41(4): 988-1001.
105. Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, et al. Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt Health and Demographic Surveillance System. *Scand J Public Health.* 2007; 35(Suppl 69): 8-20.
106. Bradshaw D, Kielkowski D, Sitas F. New birth and death registration forms - A foundation for the future, a challenge for health workers? *S Afr Med J.* 1998; 88(8): 971-4.
107. Republic of South Africa. Births and Deaths Registration Act 51 of 1992. Government Gazette No. 13953. Cape Town: Government Printer; 1992.
108. Groenewald P, Pieterse D. Getting the facts of death right. *S Afr Med J.* 2007; 97(11): 1056-8.
109. Republic of South Africa. Births and Deaths Registration Act, 1992: Regulations on the registration of births and deaths, 2014. In: Regulation Gazette no. 10135. Pretoria: Department of Home Affairs; 2014.

110. Statistics South Africa. Mortality and causes of death in South Africa, 2008: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa; 2010.
111. South Africa Government Services. Application to give notice of death n.d. Available: [http://www.services.gov.za/services/content/Home/ServicesForPeople/Traveloutsid eSA/Applicationtogivenoticeofdeath/en\\_ZA](http://www.services.gov.za/services/content/Home/ServicesForPeople/Traveloutsid eSA/Applicationtogivenoticeofdeath/en_ZA) Cited: 25 Sep 2011.
112. Department of Home Affairs. Registering a death. n.d. Available: <http://www.dha.gov.za/> Cited: Accessed 25/09/2011.
113. Republic of South Africa. Inquest Act No. 58 of 1959. Cape Town: Government Printer; 1959.
114. Statistics South Africa. Mortality and causes of death in South Africa, 2007: Findings from death notification. Statistical release P0309.3. Pretoria: Statistics South Africa; 2009.
115. World Health Organisation. ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth revision. 2nd ed. Geneva: World Health Organization; 2004.
116. Statistics South Africa. Cause of death certification: A guide for completing the Notice of Death/Stillbirth (DHA-1663). Pretoria, South Africa: Statistics South Africa, 2012. Accessed: 8 OCT 2014. Available: <http://www.sahealthinfo.org/bod/deathtraining/CODcertificationguideline.pdf>
117. Byass P, Kahn K, Fottrell E, Mee P, Collinson M, Tollman S. Using verbal autopsy to track epidemic dynamics: the case of HIV-related mortality in South Africa. *Popul Health Metr.* 2011; 9(1): 46.
118. Byass P, Chandramohan D, Clark SJ, Ambruoso L, Fottrell E, Graham WJ, et al. Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Glob Health Action.* 2012; 5( ): 19281.
119. Bah S. Multiple forces working in unison: the case of rapid improvement of vital statistics in South Africa post-1996. *World Health & Population.* 2009; 11(1): 50-9. Epub 2009/12/31.
120. Nojilana B, Groenewald P, Bradshaw D, Reagon G. Quality of cause of death certification at an academic hospital in Cape Town, South Africa. *S Afr Med J.* 2009; 99(9): 648-52.
121. Yudkin PL, Burger EH, Bradshaw D, Groenewald P, Ward AM, Volmink J. Deaths caused by HIV disease under-reported in South Africa. *AIDS.* 2009; 23(12): 1600-2.
122. Burger EH, Van der Merwe L, Volmink J. Errors in the completion of the death notification form. *S Afr Med J.* 2007; 97(11): 1077-81.



123. Burger EH, Groenewald P, Bradshaw D, Ward AM, Yudkin PL, Volmink J. Validation study of cause of death statistics in Cape Town, South Africa, found poor agreement. *J Clin Epidemiol*. 2012; 65(3): 309-16. Epub 2011/12/16.
124. Mathers C, Lopez A, Stein C, Fat DM, Rao C, Inoue M, et al. Deaths and disease burden by cause: Global Burden of Disease estimates for 2001 by World Bank country groups. Disease Control Priorities Project Working Paper No. 18. Annex A, Table 6. New York, NY: World Bank, 2004, revised 2005. Accessed: 11 Mar 2012. Available: <http://www.dcp2.org/file/33/wp18.pdf>
125. Mathers C, Lopez A, Murray C. The burden of disease and mortality by condition: data, methods, and results for 2001. In: Lopez A, Mathers C, Ezzati M, Jamison D, Murray CJ, editors. *Global Burden of Disease and Risk Factors*. New York, NY: Oxford University Press; 2006. p. 100-6, Table 3A.3.
126. França E, de Abreu DX, Rao C, Lopez AD. Evaluation of cause-of-death statistics for Brazil, 2002–2004. *Int J Epidemiol*. 2008; 37(4): 891-901.
127. Obermeyer Z, Rajaratnam JK, Park CH, Gakidou E, Hogan MC, Lopez AD, et al. Measuring adult mortality using sibling survival: a new analytical method and new results for 44 countries, 1974–2006. *PLoS Med*. 2010; 7(4): e1000260.
128. Health Metrics Network. MOVE-IT: Monitoring of vital events [Internet]. Geneva: Health Metrics Network, WHO; 2012. Available: [http://www.who.int/healthmetrics/move\\_it/en/index.html](http://www.who.int/healthmetrics/move_it/en/index.html) Cited: 12 July 2012.
129. Health Metrics Network. The case for investment in civil registration and vital statistics systems. Geneva: World Health Organization, 2012. Accessed: 10 Apr 2014. Available: [http://www.who.int/healthmetrics/resources/CRVS\\_investment\\_case.pdf?ua=1](http://www.who.int/healthmetrics/resources/CRVS_investment_case.pdf?ua=1)
130. Joubert J, Rao C, Bradshaw D, Dorrington RE, Vos T, Lopez AD. Characteristics, availability and uses of vital registration and other mortality data sources in post-democracy South Africa. *Glob Health Action*. 2012; 5(19263).
131. Joubert J, Rao C, Bradshaw D, Vos T, Lopez AD. Evaluating the quality of national mortality statistics from civil registration in South Africa, 1997–2007. *PLoS ONE*. 2013; 8(5): e64592. doi:10.1371/journal.pone.0064592.
132. Bah S. The making and unmaking of a national but stratified vital statistics system in the Republic of South Africa and the new making of a more comprehensive vital statistics system. *South Afr J Demogr*. 1999; 7(1): 45-50.
133. Freund B. Forced resettlement and the political economy of South Africa. *Review of African Political Economy*. 1984, [http://www.jstor.org/stable/4005648\(29\)](http://www.jstor.org/stable/4005648(29)): 49-63.
134. Kok P, O'Donovan M, Bouare O, Van Zyl J. Post-apartheid patterns of internal migration in South Africa. Pretoria: Human Sciences Research Council Publishers; 2003.

135. Zulu EM, Sibanda A. Racial differences in household structure. In: Zuberi T, Sibanda A, Udjo EO, editors. *The Demography of South Africa*. Armonk, New York and London: M.E. Sharpe; 2005.
136. Dorrington RE, Bradshaw D, Wegner T. Estimates of the Level and Shape of Mortality Rates in South Africa around 1985 and 1990 derived by applying indirect demographic techniques to reported deaths. MRC Technical Report. Cape Town: South African Medical Research Council; 1999.
137. Kaufman C. Reproductive control in South Africa. Working Paper No. 97. New York: The Population Council, Policy Research Division., 1997. Available:
138. Murray C. Class, gender and the household: the developmental cycle in Southern Africa. *Development and Change*. 1987; 18(2): 235-49.
139. Murray C. Migrant labour and changing family structure in the rural periphery of southern Africa. *Journal of Southern African Studies*. 1980; 6(2): 139-56.
140. Platzky L, Walker C, Surplus People Project (South Africa). *The surplus people: forced removals in South Africa*. Johannesburg: Ravan Press; 1985.
141. Sharp J. Relocation and the problem of survival in Qwaqwa: A report from the field. *Social Dynamics*. 1982; 8(2): 11-29.
142. Republic of South Africa. Constitution of the Republic of South Africa Act 200 of 1993. Cape Town: Government Printer; 1993.
143. National Health Information System. Evaluation of the new death notification form (BI 1663). Pretoria, South Africa: Department of Health; 1999.
144. Republic of South Africa. Births and Deaths Registration Amendment Act 18 of 2010. Government Gazette No. 33851. Cape Town: Government Printer; 2010.
145. Health Metrics Network. Assessing the National Health Information System: An assessment tool, Version 4.00. Geneva: World Health Organization, 2008. Accessed: 28 May 2014. Available: [http://www.who.int/healthmetrics/tools/Version\\_4.00\\_Assessment\\_Tool3.pdf](http://www.who.int/healthmetrics/tools/Version_4.00_Assessment_Tool3.pdf)
146. Health Metrics Network. Strengthening country health information systems: Assessment and monitoring tool. Version 2.00. Geneva, Switzerland: World Health Organization, 2007. Accessed: 20 Dec 2013. Available: [http://www.who.int/healthmetrics/tools/hmn\\_assessment\\_tool\\_guide\\_ver2.pdf](http://www.who.int/healthmetrics/tools/hmn_assessment_tool_guide_ver2.pdf)
147. World Health Organisation. Country health information systems: a review of the current situation and trends. Geneva, Switzerland: World Health Organisation, 2011. Accessed: 22 Dec 2012. Available: [http://www.who.int/healthmetrics/news/chis\\_report.pdf](http://www.who.int/healthmetrics/news/chis_report.pdf)
148. Mikkelsen L. Assessing the quality of vital statistics systems: lessons from national evaluations in Sri Lanka and the Philippines. Brisbane, Australia: Health Information Systems Knowledge Hub, School of Population Health, The University

- of Queensland, 2009. Accessed: 13 May 2014. Available: <http://www.uq.edu.au/hishub/docs/WP08/HISHUB-WP08-09-WEB-15Oct12.pdf>
149. The University of Queensland Health Information Systems Knowledge Hub. Tools and Guidelines c2014. Available: <http://www.uq.edu.au/hishub/tools-and-guidelines> Cited: 14 Oct 2014.
  150. Ruzicka LT, Lopez AD. The use of cause-of-death statistics for health situation assessment: national and international experiences. *World Health Stat Q.* 1990; 43(4): 249-58.
  151. World Health Organisation. World Health Statistics Annual. Geneva: World Health Organization, 1998 & 1989.
  152. Mahapatra P, Rao CPV. Cause of death reporting systems in India: a performance analysis. *National Medical Journal of India.* 2001; 14(3): 154-62.
  153. Naghavi M, Makela S, Foreman K, O'Brien J, Pourmalek F, Lozano R. Algorithms for enhancing public health utility of national causes-of-death data. *Popul Health Metr.* 2010; 8: 9.
  154. Statistics South Africa. Causes of death in South Africa, 1997 to 2007, extracted from death notification forms. Provided by Statistics South Africa, Pretoria, South Africa. 1997-2007.
  155. ASSA2008 AIDS and Demographic Models [Internet]. Actuarial Society of South Africa. c2011 [cited 10 Nov 2012]. Available from: <http://aids.actuarialsociety.org.za/ASSA2008-Model-3480.htm>.
  156. Johansson L, Pavilion G, Anderson R, Glenn D, Griffiths C, Hoyert D, et al. Counting the dead and what they died of. *Bull World Health Organ.* 2006; 84(3): 254.
  157. Walker S. Capturing health information--a coding perspective. *Bull World Health Organ.* 2006; 84(3): 254-5.
  158. Mathers CD, Ma Fat D, Inoue M, Rao C, Lopez AD. Authors' response. *Bulletin of the World Health Organisation.* 2006; 84(3): 255-6.
  159. AbouZahr C, Mikkelsen L, Rampatige R, Lopez AD. Mortality statistics: a tool to improve understanding and quality. Brisbane, Australia: The University of Queensland HIS Hub, 2010. Accessed: 14 December 2013. Available: <http://www.uq.edu.au/hishub/docs/WP13/HISHub-WP13-11-WEB%202Oct12.pdf>
  160. Westwood AT. Childhood deaths due to HIV--the role of the new death certificate. *S Afr Med J.* 2000; 90(9): 877.
  161. Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D. Identifying deaths from AIDS in South Africa. *AIDS.* 2005; 19(2): 193-201.
  162. Birnbaum JK, Murray CJL, Lozano R. Exposing misclassified HIV/AIDS deaths in South Africa. *Bull World Health Organ.* 2011; 89(4): 278-85. Epub 2011/04/12.

163. Pillay-van Wyk V, Msemburi W, Laubscher R, Dorrington RE, Groenewald P, Matzopoulos R, et al. Second National Burden of Disease Study South Africa: national and subnational mortality trends, 1997–2009. *Lancet*. 2013; 381(Supplement 2): S113.
164. Johansson LA. Targeting non-obvious errors in death certificates. Uppsala, Sweden: Uppsala University; 2008.
165. Van den Broeck J, Argeseanu Cunningham S, Eeckels R, Herbst K. Data cleaning: detecting, diagnosing, and editing data abnormalities. *PLoS Med*. 2005; 2(10): e267.
166. Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What can we conclude from death registration? Improved methods for evaluating completeness. *PLoS Med* [Internet]. 2010 14 Aug 2012; 7(4). Available: <http://dx.doi.org/10.1371/journal.pmed.1000262>.
167. Hill KH. Estimating census and death registration completeness. *Asian and Pacific Population Forum*. 1987; 1(3): 8-13.
168. Bennett NG, Horiuchi S. Mortality estimation from registered deaths in less developed countries. *Demography*. 1984; 21(2): 217-33.
169. Bennett NG, Horiuchi S. Estimating the completeness of death registration in a closed population. *Population Index*. 1981; 47(2): 207-21.
170. Machedmedze T. Old age mortality in South Africa. MPhil thesis. [Partial fulfillment of course work degree]. Cape Town: University of Cape Town; 2009.
171. Dorrington RE, Bradshaw D. Maternal mortality in South Africa: lessons from a case study in the use of deaths reported by households in censuses and surveys. *Journal of Population Research*. 2011; 28(1): 49-73.
172. Darikwa TB, Dorrington RE. The level and trends of child mortality in South Africa, 1996-2006. *J Afr Population Stud*. 2011; 25(1, Supplement): 158-72.
173. Roos LL, Wajda A. Record linkage strategies. Part I: Estimating information and evaluating approaches. *Methods Archive*. 1991; 30(2): 117-23.
174. Jutte DP, Roos LL, Brownell MD. Administrative record linkage as a tool for public health research. *Annu Rev Public Health*. 2011; 32: 91-108. Epub 2011/01/12.
175. Holman CD, Bass AJ, Rosman DL, Smith MB, Semmens JB, Glasson EJ, et al. A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system. *Aust Health Rev*. 2008; 32(4): 766-77. Epub 2008/11/05.
176. Howe GR, Lindsay J. A generalized iterative record linkage computer system for use in medical follow-up studies. *Comput Biomed Res*. 1981; 14(4): 327-40.
177. Ferrante A, Boyd J. A transparent and transportable methodology for evaluating Data Linkage software. *J Biomed Inform*. 2012; 45(1): 165-72.

178. Newman TB, Brown AN. Use of commercial record linkage software and vital statistics to identify patient deaths. *J Am Med Inform Assoc.* 1997; 4(3): 233-7.
179. Horon IL. Underreporting of maternal deaths on death certificates and the magnitude of the problem of maternal mortality. *Am J Public Health.* 2005; 95(3): 478-82. Epub 2005/02/25.
180. Herrchen B, Gould JB, Nesbitt TS. Vital statistics linked birth/infant death and hospital discharge record linkage for epidemiological studies. *Comput Biomed Res.* 1997; 30(4): 290-305. Epub 1997/08/01.
181. Klijs B, Nusselder WJ, Mackenbach JP. Nationwide individual record linkage study showed poor agreement of causes of death and hospital diagnoses at individual level but reasonable agreement at population level. *J Clin Epidemiol.* 2014; 67(2): 160-8.
182. Zingmond DS, Ye Z, Ettner SL, Liu H. Linking hospital discharge and death records—accuracy and sources of bias. *J Clin Epidemiol.* 2004; 57(1): 21-9.
183. Coady SA, Sorlie PD, Cooper LS, Folsom AR, Rosamond WD, Conwill DE. Validation of death certificate diagnosis for coronary heart disease: the Atherosclerosis Risk in Communities (ARIC) Study. *J Clin Epidemiol.* 2001; 54(1): 40-50.
184. Li B, Quan H, Fong A, Lu M. Assessing record linkage between health care and Vital Statistics databases using deterministic methods. *BMC Health Serv Res.* 2006; 6(1): 48.
185. Hufanga S, Carter KL, Rao C, Lopez AD, Taylor R. Mortality trends in Tonga: an assessment based on a synthesis of local data. *Population Health Metrics.* 2012; 10(14).
186. Carter K, Williams G, Tallo V, Sanvictores D, Madera H, Riley I. Capture-recapture analysis of all-cause mortality data in Bohol, Philippines. *Popul Health Metr.* 2011; 9(1): 9.
187. Rao C, Kosen S, Bisara D, Usman Y, Adair T, Djaja S, et al. Tuberculosis mortality differentials in Indonesia during 2007 - 2008: evidence for health policy and monitoring. *The International Journal of Tuberculosis and Lung Disease.* 2011; 15(12): 1608-14.
188. Newcombe HB, Kennedy JM, Axford SJ, James AP. Automatic Linkage of Vital Records. *Science.* 1959; 130(3381): 954-9.
189. Newcombe HB, Kennedy JM. Record linkage: making maximum use of the discriminating power of identifying information. *Communications of the ACM (Association for Computing Machinery).* 1962; 5(11): 563-6.
190. Newcombe HB. Handbook of record linkage: methods for health and statistical studies, administration, and business. Oxford, England: Oxford University Press; 1988.

191. Howe GR. Use of computerized record linkage in cohort studies. *Epidemiol Rev.* 1998; 20(1): 112-21.
192. Winkler WE. Record linkage software and methods for merging administrative lists. Statistical Research Report Series. Washington D.C., USA: Statistical Research Division, U.S. Bureau of the Census, 2001. Accessed: 11 Dec 2011. Available: <http://www.vrdc.cornell.edu/info7470/2011/Readings/rr2001-03.pdf>
193. Fellegi IP, Sunter AB. A theory for record linkage. *Journal of the American Statistical Association.* 1969; 64(328): 1183-210.
194. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med.* 1995; 14(5-7): 491-8.
195. Winkler WE. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In: American Statistical Association: pp. 354-9.
196. Zhu VJ, Overhage MJ, Egg J, Downs SM, Grannis SJ. An empiric modification to the probabilistic record linkage algorithm using frequency-based weight scaling. *J Am Med Inform Assoc.* 2009; 16(5): 738-45. Epub 2009/07/02.
197. Herzog TH, Scheuren F, Winkler WE. Record linkage. *Wiley Interdisciplinary Reviews: Computational Statistics.* 2010; 2(5): 535-43.
198. Roos LL, Jr., Wajda A, Nicol JP. The art and science of record linkage: methods that work with few identifiers. *Comput Biol Med.* 1986; 16(1): 45-57. Epub 1986/01/01.
199. Millar T, Domingo-Salvany A, Eastwood C, Hay G. Glossary of terms relating to capture-recapture methods. *J Epidemiol Community Health.* 2008; 62(8): 677-81.
200. Clark DE. Practical introduction to record linkage for injury research. *Inj Prev.* 2004; 10(3): 186-91.
201. Wajda A, Roos LL. Simplifying record linkage: Software and strategy. *Comput Biol Med.* 1987; 17(4): 239-48.
202. Kabudula C, Clark B, Gomez-Olive F, Tollman S, Menken J, Reniers G. The promise of record linkage for assessing the uptake of health services in resource constrained settings: a pilot study from South Africa. *BMC Med Res Methodol.* 2014; 14(1): 71.
203. Qayad MG, Zhang H. Accuracy of public health data linkages. *Matern Child Health J.* 2009; 13(4): 531-8.
204. Bohensky M, Jolley D, Sundararajan V, Evans S, Pilcher D, Scott I, et al. Data Linkage: A powerful research tool with potential problems. *BMC Health Serv Res.* 2010; 10(1): 346.
205. Méray N, Reitsma JB, Ravelli ACJ, Bonsel GJ. Probabilistic record linkage is a valid and transparent tool to combine databases without a patient identification number. *J Clin Epidemiol.* 2007; 60(9): 883.e1-e11.

206. Quantin C, Binquet C, Bourquard K, Pattisina R, Gouyon-Cornet B, Ferdynus C, et al. Which are the best identifiers for record linkage? *Informatics for Health and Social Care*. 2004; 29(3-4): 221-7.
207. Van den Brandt PA, Schouten LJ, Goldbohm RA, Dorant E, Hunen PMH. Development of a record linkage protocol for use in the Dutch Cancer Registry for epidemiological research. *Int J Epidemiol*. 1990; 19(3): 553-8.
208. Schrauder A, Claus H, Elias J, Vogel U, Haas W, Hellenbrand W. Capture-recapture analysis to estimate the incidence of invasive meningococcal disease in Germany, 2003. *Epidemiol Infect*. 2007; 135(4): 657-64. Epub 2006/08/30.
209. El Adssi H, Debouverie M, Guillemin F, on behalf of the LORSEP Group. Estimating the prevalence and incidence of multiple sclerosis in the Lorraine region, France, by the capture-recapture method. *Multiple Sclerosis Journal*. 2012; 18(9): 1244-50.
210. Domingo-Salvany A, Hartnoll RL, Maguire A, Suelves JM, Antó JM. Use of capture-recapture to estimate the prevalence of opiate addiction in Barcelona, Spain, 1989. *Am J Epidemiol*. 1995; 141(6): 567-74.
211. Bobo WV, Pathak J, Kremers HM, Yawn BP, Brue SM, Stoppel CJ, et al. An electronic health record driven algorithm to identify incident antidepressant medication users. *J Am Med Inform Assoc*. 2014; Published Online First, 18 Apr 2014.
212. International Population Data Linkage Network. Data linkage centres c2014 [Updated: n.d.]. Available: <http://www.ipdln.org/data-linkage-centres> Cited: 29 Aug 2014.
213. Black N, Barker M, Payne M. Cross sectional survey of multicentre clinical databases in the United Kingdom. *Br Med J*. 2004; 328(7454): 1478.
214. Kelman CW, Bass AJ, Holman CDJ. Research use of linked health data — a best practice protocol. *Aust N Z J Public Health*. 2002; 26(3): 251-5.
215. Evans SM, Bohensky M, Cameron PA, McNeil JJ. A survey of Australian clinical registries: can quality of care be measured? *Intern Med J*. 2011; 41(1a): 42-8.
216. Roos LL, Brownell M, Lix L, Roos NP, Walld R, MacWilliam L. From health research to social research: privacy, methods, approaches. *Soc Sci Med*. 2008; 66(1): 117-29.
217. Population Health Research Network [Internet]. The Australian Government. c2011 [cited 14Jul 2014]. Available from: <http://www.phrn.org.au/>.
218. Baldwin JA. The Oxford Record Linkage Study as a medical information system. *J R Soc Med*. 1972; 65(3): 237-9.
219. Zöller B, Ji J, Sundquist J, Sundquist K. Risk of coronary heart disease in patients with cancer: a nationwide follow-up study from Sweden. *Eur J Cancer*. 2012; 48(1): 121-8.

220. British Columbia Linked Health Database (BCLHD) [Internet]. University of British Columbia. 2014 [cited 14 Jul 2014]. Available: <http://www.chspr.ubc.ca/data>.
221. Kurland LT, Molgaard CA. The patient record in epidemiology. *Sci Am*. 1981; 245(4): 54-63. Epub 1981/10/01.
222. Hawkins BS, Collaborative Ocular Melanoma Study Group. The Collaborative Ocular Melanoma Study (COMS) randomized trial of pre-enucleation radiation of large choroidal melanoma: IV. Ten-year mortality findings and prognostic factors. COMS report nu. 24. *Am J Ophthalmol*. 2004; 138(6): 936-51. Epub 2005/01/05.
223. Bangdiwala SI, Cohn R, Hazard C, Davis CE, Prineas RJ. Comparisons of cause of death verification methods and costs in the Lipid Research Clinics Program Mortality Follow-up Study. *Control Clin Trials*. 1989; 10(2): 167-87. Epub 1989/06/01.
224. Li JZ, Chen ML, Wang S, Dong J, Zeng P, Hou LW. A long-term follow-up study of serum lipid levels and coronary heart disease in the elderly. *Chinese Medical Journal (English)*. 2004; 117(2): 163-7. Epub 2004/02/21.
225. Ederer F, Geisser MS, Mongin SJ, Church TR, Mandel JS. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison. The Minnesota Study. *J Clin Epidemiol*. 1999; 52(5): 447-52. Epub 1999/06/09.
226. Blakely T, Woodward A, Salmond C. Anonymous linkage of New Zealand mortality and Census data. *Aust N Z J Public Health*. 2000; 24(1): 92-5. Epub 2000/04/25.
227. Holian J. Live birth and infant death record linkage: methodological and policy issues. *J Health Soc Policy*. 2000; 12(1): 1-10. Epub 2000/11/07.
228. Amin J, Law MG, Bartlett M, Kaldor JM, Dore GJ. Causes of death after diagnosis of hepatitis B or hepatitis C infection: a large community-based linkage study. *Lancet*. 2006; 368(9539): 938-45.
229. Johansson LA, Westerling R. Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics. *Int J Epidemiol*. 2000; 29(3): 495-502.
230. Johansson LA, Westerling R. Comparing hospital discharge records with death certificates: can the differences be explained? *J Epidemiol Community Health*. 2002; 56(4): 301-8. Epub 2002/03/16.
231. Mühlhauser I, Sawicki P, Blank M, Overmann H, Richter B, Berger M. Reliability of causes of death in persons with Type I diabetes. *Diabetologia*. 2002; 45(11): 1490-7.
232. Pattaraarchachai J, Rao C, Polprasert W, Porapakkham Y, Pao-in W, Singwerathum N, et al. Cause-specific mortality patterns among hospital deaths in Thailand: validating routine death certification. *Popul Health Metr*. 2010; 8(12).
233. Polprasert W, Rao C, Adair T, Pattaraarchachai J, Porapakkham Y, Lopez AD. Cause-of-death ascertainment for deaths that occur outside hospitals in Thailand: application of verbal autopsy methods. *Popul Health Metr*. 2010; 8(13).



234. Rampatige R, Gamage S, Peiris S, Lopez AD. Assessing the reliability of causes of death reported by the vital registration system in Sri Lanka: Medical records review in Colombo [online]. *Health Information Management Journal*. 2013; 42(3): 20-8.
235. World Health Organisation. Civil registration and vital statistics (CRVS) Geneva, Switzerland: World Health Organization; c2014. Available: <http://www.who.int/healthinfo/anacod/en/>; [http://www.who.int/healthinfo/civil\\_registration/en/](http://www.who.int/healthinfo/civil_registration/en/) Cited: 26 Sep 2014.
236. Dorrington RE, Bourne D, Bradshaw D, Laubscher R, Timaeus IM. The impact of HIV/AIDS on adult mortality in South Africa. Tygerberg, Cape Town: South African Medical Research Council, 2001. Accessed: 12 Sep 2010. Available: <http://www.mrc.ac.za/bod/complete.pdf>
237. Dorrington RE, Johnson L, Budlender D. ASSA2008 AIDS and Demographic Models: User Guide. Available: <http://aids.actuarialsociety.org.za/ASSA2008-Model-3480.htm> Accessed: 20 Oct 2011. Cape Town: Centre for Actuarial Research, University of Cape Town; 2010.
238. Dorrington RE, Bradshaw D. HIV/AIDS profile in the provinces of South Africa, 2002. *Southern African Journal of HIV Medicine*. 2003, [http://www.scopus.com/inward/record.url?eid=2-s2.0-0037305433&partnerID=40&md5=efa90b8d788383e987131660a6f00ca6\(10\):](http://www.scopus.com/inward/record.url?eid=2-s2.0-0037305433&partnerID=40&md5=efa90b8d788383e987131660a6f00ca6(10):) 46-7.
239. Dorrington RE, Johnson L, Bradshaw D, Daniel T. The Demographic Impact of HIV/AIDS in South Africa. National and Provincial Indicators for 2006. Cape Town: Centre for Actuarial Research, South African Medical Research Council and Actuarial Society of South Africa, 2006. Accessed: 19 Nov 2012. Available: <http://www.mrc.ac.za/bod/DemographicImpactHIVIndicators.pdf>
240. Peach HG, Brumley DJ. Death certification by doctors in non-metropolitan Victoria. *Aust Fam Physician*. 1998; 27(3): 178-82.
241. Harrison B, Stephenson P, Mohan G, Nasser S. An ongoing Confidential Enquiry into asthma deaths in the Eastern Region of the UK, 2001-2003. *Prim Care Respir J*. 2005; 14(6): 303-13. Epub 2006/05/17.
242. Fox CS, Evans JC, Larson MG, Lloyd-Jones DM, O'Donnell CJ, Sorlie PD, et al. A comparison of death certificate out-of-hospital coronary heart disease death with physician-adjudicated sudden cardiac death. *The American Journal of Cardiology*. 2005; 95(7): 856-9.
243. Lahti RA, Penttilä A. The validity of death certificates: routine validation of death certification and its effects on mortality statistics. *Forensic Sci Int*. 2001; 115(1-2): 15-32.
244. Sidenius KE, Munch EP, Madsen F, Lange P, Viskum K, Soes-Petersen U. Accuracy of recorded asthma deaths in Denmark in a 12-months period in 1994/95. *Respir Med*. 2000; 94(4): 373-7.

245. Ederer F, Geisser MS, Mongin SJ, Church TR, Mandel JS. Colorectal cancer deaths as determined by expert committee and from death certificate: a comparison. The Minnesota Study. *J Clin Epidemiol*. 1999; 52(5): 447-52.
246. Mlangeni DA, Harris MD, Franklin L, Hunt P, Karas JA, Enoch DA. Death certificates provide a poor estimation of attributable mortality due to *Clostridium difficile* when compared to a death review panel using defined criteria. *J Hosp Infect*. 2011; 77(4): 370-1.
247. Moy CS, Albert DM, Diener-West M, McCaffrey LD, Scully RE, Willson JK. Cause-specific mortality coding. Methods in the Collaborative Ocular Melanoma Study (COMS). COMS report no. 14. *Control Clin Trials*. 2001; 22(3): 248-62. Epub 2001/06/01.
248. Sington JD, Cottrell BJ. Analysis of the sensitivity of death certificates in 440 hospital deaths: a comparison with necropsy findings. *J Clin Pathol*. 2002; 55(7): 499-502.
249. The United Kingdom Heart Attack Study Collaborative Group. The falling mortality from coronary heart disease: a clinicopathological perspective. *Heart*. 1998; 80(2): 121-6.
250. Engel LW, Strauchen JA, Chiazze L, Heid M. Accuracy of death certification in an autopsied population with specific attention to malignant neoplasms and vascular diseases. *Am J Epidemiol*. 1980; 111(1): 99-112.
251. Comstock RD, Mallonee S, Jordan F. A comparison of two surveillance systems for deaths related to violent injury. *Inj Prev*. 2005; 11(1): 58-63.
252. Browne ML, Lewis-Michl EL, Stark AD. Investigation and reporting practices for drownings: Implications for injury prevention research in New York State. *Am J Forensic Med Pathol*. 2002; 23(4): 398-401.
253. Saito I, Ozawa H, Aono H, Ikebe T, Yamashita T, Makino Y. Trends in fatal coronary heart disease among people aged 25–74 years in Oita City, Japan, from 1987–1998. *J Clin Epidemiol*. 2002; 55(5): 469-76.
254. Iribarren C, Crow RS, Hannan PJ, Jacobs DR, Jr., Luepker RV. Validation of death certificate diagnosis of out-of-hospital sudden cardiac death. *Am J Cardiol*. 1998; 82(1): 50-3. Epub 1998/07/22.
255. Rauscher GH, Sandler DP. Validating cancer histories in deceased relatives. *Epidemiology*. 2005; 16(2): 262-5. Epub 2005/02/11.
256. Jones K, Berrill WT, Bromly CL, Hendrick DJ. A confidential enquiry into certified asthma deaths in the North of England, 1994-96: influence of co-morbidity and diagnostic inaccuracy. *Respir Med*. 1999; 93(12): 923-7. Epub 2000/02/01.
257. Ostbye T, Taylor DH, Jr., Clipp EC, Scoyoc LV, Plassman BL. Identification of dementia: agreement among national survey data, medicare claims, and death certificates. *Health Serv Res*. 2008; 43(1 Pt 1): 313-26. Epub 2008/01/24.

258. Wang L, Yang G, Ma J, Rao C, Wan X, Dubrovsky G, et al. Evaluation of the quality of cause of death statistics in rural China using verbal autopsies. *J Epidemiol Community Health*. 2007; 61(6): 519-256.
259. Yeo L, Lynch C, Hardiman O. Validating population-based registers for ALS: how accurate is death certification? *J Neurol*. 2010; 257(8): 1235-9. Epub 2010/02/13.
260. Azizova TV, Fedirko V, Tsareva Y, Tretyakov F, Lassen CF, Friis S, et al. Mayak workers study cohort. An inter-institutional comparison of causes of death in the cause-of-death register of Ozyorsk in the Russian Federation. *Methods Inf Med*. 2012; 51(2): 144-9. Epub 2012/02/09.
261. Chugh SS, Jui J, Gunson K, Stecker EC, John BT, Thompson B, et al. Current burden of sudden cardiac death: multiple source surveillance versus retrospective death certificate-based review in a large U.S. community. *J Am Coll Cardiol*. 2004; 44(6): 1268-75. Epub 2004/09/15.
262. Goraya TY, Jacobsen SJ, Belau PG, Weston SA, Kottke TE, Roger VL. Validation of death certificate diagnosis of out-of-hospital coronary heart disease deaths in Olmsted County, Minnesota. *Mayo Clin Proc*. 2000; 75(7): 681-7. Epub 2000/07/25.
263. Lydon-Rochelle MT, Cardenas V, Nelson JL, Tomashek KM, Mueller BA, Easterling TR. Validity of maternal and perinatal risk factors reported on fetal death certificates. *Am J Public Health*. 2005; 95(11): 1948-51. Epub 2005/10/01.
264. Roulson J, Benbow EW, Hasleton PS. Discrepancies between clinical and autopsy diagnosis and the value of post mortem histology; a meta-analysis and review. *Histopathology*. 2005; 47(6): 551-9.
265. Kircher T, Nelson J, Burdo H. The autopsy as a measure of accuracy of the death certificate. *N Engl J Med*. 1985; 313(20): 1263-9. Epub 1985/11/14.
266. Ravakhah K. Death certificates are not reliable: revivification of the autopsy. *South Med J*. 2006; 99(7): 728-33. Epub 2006/07/27.
267. Shojanian KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time: a systematic review. *JAMA*. 2003; 289(21): 2849-56. Epub 2003/06/05.
268. Rao C, Yang G, Hu J, Ma J, Xia W, Lopez AD. Validation of cause-of-death statistics in urban China. *Int J Epidemiol*. 2007; 36(3): 642-51.
269. de Carvalho MH, Alvarez-Hernandez G, Denman C, Harlow SD. Validity of underlying cause of death statistics in Hermosillo, Mexico. *Salud Publica Mex*. 2011; 53(4): 312-9. Epub 2011/10/12.
270. Rampatige R, Mikkelsen L, Prado BH, Riley I, Lopez AD. Assessing the reliability of hospital-based cause-of-death statistics: evidence-based guidelines for country application. Brisbane, Australia: Health Information Systems Knowledge Hub, School of Population Health, The University of Queensland, 2013. Accessed: 22 Oct 2014. Available: [www.uq.edu.au/hishub](http://www.uq.edu.au/hishub)

271. Jorgensen IM, Bulow S, Jensen VB, Dahm TL, Prah P, Juel K. Asthma mortality in Danish children and young adults, 1973-1994: epidemiology and validity of death certificates. *The European Respiratory Journal*. 2000; 15(5): 844-8. Epub 2000/06/15.
272. May KL. Death certificates in asthma and COPD patients (survey of statistical data in Warsaw). *Monaldi Arch Chest Dis*. 2002; 57(5-6): 253-7. Epub 2003/06/20.
273. Reid D, Hendrick V, Aitken T, Berrill W, Stenton S, Hendrick D. Age-dependent inaccuracy of asthma death certification in Northern England, 1991-1992. *Eur Respir J*. 1998; 12(5): 1079-83.
274. Smyth ET, Wright SC, Evans AE, Sinnamon DG, MacMahon J. Death from airways obstruction: accuracy of certification in Northern Ireland. *Thorax*. 1996; 51(3): 293-7. Epub 1996/03/01.
275. Andresen EM, Lee JA, Pecoraro RE, Koepsell TD, Hallstrom AP, Siscovick DS. Underreporting of diabetes on death certificates, King County, Washington. *Am J Public Health*. 1993; 83(7): 1021-4.
276. Murray CJL, Lopez AD, Black R, Ahuja R, Ali S, Baqui A, et al. Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. *Popul Health Metr*. 2011; 9(1): 27.
277. Fauveau V. Assessing probable causes of death without death registration or certificates: a new science? *Bull World Health Organ*. 2006; 84(3): 246-7.
278. Gajalakshmi V, Peto R, Kanaka S, Balasubramanian S. Verbal autopsy of 48 000 adult deaths attributable to medical causes in Chennai (formerly Madras), India. *BMC Public Health*. 2002; 2: 7. Epub 2002/05/17.
279. Yang G, Rao C, Ma J, Wang L, Wan X, Dubrovsky G, et al. Validation of verbal autopsy procedures for adult deaths in China. *Int J Epidemiol*. 2006; 35(3): 741-8.
280. Chandramohan D, Maude G, Rodrigues L, Hayes R. Verbal autopsies for adult deaths: their development and validation in a multicentre study. *Trop Med Int Health*. 1998; 3: 436 - 46.
281. World Health Organisation. WHO Technical Consultation on Verbal Autopsy Tools Final Report: Review of the literature and currently-used verbal autopsy tools. Geneva, Switzerland: World Health Organisation, 2005. Accessed: 29 April 2011. Available: [http://www.who.int/healthinfo/statistics/mort\\_verbalautopsy.pdf](http://www.who.int/healthinfo/statistics/mort_verbalautopsy.pdf)
282. Lozano R, Lopez AD, Atkinson C, Naghavi M, Flaxman AD, Murray CJ, et al. Performance of physician-certified verbal autopsies: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr*. 2011; 9: 32. Epub 2011/08/06.
283. Mobley CC, Boerma JT, Titus S, Lohrke B, Shangula K, Black RE. Validation study of a verbal autopsy method for causes of childhood mortality in Namibia. *J Trop Pediatr*. 1996; 42(6): 365-9. Epub 1996/12/01.

284. Quigley MA, Armstrong Schellenberg JR, Snow RW. Algorithms for verbal autopsies: a validation study in Kenyan children. *Bull World Health Organ.* 1996; 74(2): 147-54. Epub 1996/01/01.
285. Iriya N, Manji KP, Mbise RL. Verbal autopsy in establishing cause of perinatal death. *East Afr Med J.* 2002; 79(2): 82-4. Epub 2002/10/17.
286. Setel PW, Whiting DR, Hemed Y, Chandramohan D, Wolfson LJ, Alberti KG, et al. Validity of verbal autopsy procedures for determining cause of death in Tanzania. *Tropical Medicine & International Health.* 2006; 11(5): 681-96. Epub 2006/04/28.
287. Marsh DR, Sadruddin S, Fikree FF, Krishnan C, Darmstadt GL. Validation of verbal autopsy to determine the cause of 137 neonatal deaths in Karachi, Pakistan. *Paediatr Perinat Epidemiol.* 2003; 17(2): 132-42. Epub 2003/04/05.
288. Benara SK, Singh P. Validity of causes of infant death by verbal autopsy. *Indian J Pediatr.* 1999; 66(5): 647-50. Epub 2000/05/08.
289. Fauveau V, Koenig MA, Chakraborty J, Chowdhury AI. Causes of maternal mortality in rural Bangladesh, 1976-85. *Bull World Health Organ.* 1988; 66(5): 643-51. Epub 1988/01/01.
290. Chandramohan D, Rodrigues LC, Maude GH, Hayes RJ. The validity of verbal autopsies for assessing the causes of institutional maternal death. *Stud Fam Plann.* 1998; 29(4): 414-22. Epub 1999/01/27.
291. Gajalakshmi CK, Shanta V, Rama R. Registration of cancer mortality data in a developing area: Chennai (Madras, India) experience. *Cancer Causes Control.* 1998; 9(2): 131-6. Epub 1998/05/13.
292. Rao C, Porapakkham Y, Pattaraarchachai J, Polprasert W, Swampunyalert N, Lopez AD. Verifying causes of death in Thailand: rationale and methods for empirical investigation. *Popul Health Metr.* 2010; 8(1): 11.
293. Porapakkham Y, Rao C, Pattaraarchachai J, Polprasert W, Vos T, Adair T, et al. Estimated causes of death in Thailand, 2005: implications for health policy. *Popul Health Metr.* 2010; 8(1): 14.
294. Magro L, Conforti A, Del Zotti F, Leone R, Iorio ML, Meneghelli I, et al. Identification of severe potential drug-drug interactions using an Italian general-practitioner database. *Eur J Clin Pharmacol.* 2008; 64(3): 303-9. Epub 2007/11/10.
295. Omland G, Ruths S, Diaz E. Use of hormonal contraceptives among immigrant and native women in Norway: data from the Norwegian Prescription Database. *BJOG: An International Journal of Obstetrics and Gynaecology.* 2014; Epub ahead of print. Epub 2014/06/17.
296. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *The British Journal of General Practice.* 2010; 60(572): e128-36. Epub 2010/03/06.

297. De Angelis R, Sant M, Coleman MP, Francis S, Bailly P, Pierannunzio D, et al. Cancer survival in Europe 1999–2007 by country and age: results of EURO-CARE-5—a population-based study. *The Lancet Oncology*. 2014; 15(1): 23-34.
298. Coleman MP, Forman D, Bryant H, Butler J, Rachet B, Maringe C, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the international cancer benchmarking partnership): An analysis of population-based cancer registry data. *Lancet*. 2011; 377(9760): 127-38.
299. Swaminathan R, Lucas E, Sankaranarayanan R. Cancer survival in Africa, Asia, the Caribbean and Central America: database and attributes. *IARC Sci Publ*. 2011, (162): 23-31. Epub 2011/06/17.
300. Day C, Gray A. Health and related indicators. In: Padarath A, English R, editors. *South African Health Review 2012/13*, <http://www.hst.org.za/publications/south-african-health-review-2012/13>. Durban, South Africa: Health Systems Trust; 2013.
301. Grandin W, Westwood T, Lagerdien K, King MS. Deaths at Red Cross Children's Hospital, Cape Town 1999-2003--a study of death notification forms. *S Afr Med J*. 2006; 96(9 Pt 2): 964-8.
302. Kahn K, Tollman S, Garenne M, Gear J. Validation and application of verbal autopsies in a rural area of South Africa. *Trop Med Int Health*. 2000; 5: 824 - 31.
303. Stephen C, Bamford L, Patrick M, Wittenberg D, editors. *Saving Children 2009: Five Years of Data. A sixth survey of child healthcare in South Africa*. Pretoria: Tshepesa Press, MRC, CDC; 2011.
304. Department of Health (South Africa). *The South African National Tuberculosis Control Programme: Practical Guidelines*. Pretoria: Department of Health, 2004. Accessed: 15 Feb 2012. Available: <http://www.kznhealth.gov.za/chrp/documents/Guidelines/Guidelines%20National/Tuberculosis/SA%20TB%20Guidelines%202004.pdf>
305. Department of Health (South Africa). *Tuberculosis Strategic Plan for South Africa, 2007-2011*. Pretoria: Department of Health, 2007. Accessed: 15 Feb 2012. Available: <http://www.info.gov.za/view/DownloadFileAction?id=72544>
306. Albrecht C. Overview of the South African cancer research environment as a basis for discussions concerning the activation of CARISA (Cancer Research Initiative of South Africa). 2006. Accessed: 12 Feb 2012. Available: <http://www.sahealthinfo.org/cancer/overviewdocument.pdf>
307. Somdyala NIM, Bradshaw D, Gelderblom WCA. *Cancer incidence in selected municipalities of the Eastern Cape Province, 2003–2007*. Eastern Cape Province Cancer Registry Technical Report. Cape Town: South African Medical Research Council, 2013. Accessed: 10 July 2014. Available: <http://www.mrc.ac.za/bod/CancerIncidenceReport2013.pdf>
308. Donson H, editor. *A profile of fatal injuries in South Africa, 2008*. Annual Report for South Africa based on the National Injury Mortality Surveillance System (NIMSS). Cape Town: Medical Research Council; 2010.

309. Sankoh O, Byass P. The INDEPTH Network: filling vital gaps in global epidemiology. *Int J Epidemiol.* 2012; 41(3): 579-88.
310. Burger EH, Groenewald P, Rossouw A, Bradshaw D. Medical certification of death in South Africa – moving forward. *S Afr Med J.* 2015; 105(1): 27-30.
311. Massyn N, Day C, Barron P, Padarath A, English R, editors. District Health Barometer 2011/12. Durban: Health Systems Trust; 2013.
312. World Health Organization. Special tabulation lists for mortality and morbidity: Mortality Tabulation List 1. International Statistical Classification of Diseases and Health Related Problems, Tenth Revision. Vol. 1. 2nd ed. Geneva: World Health Organization; 2004. p. 1163-6.
313. Horton R. Counting for health. *Lancet.* 2007; 370(9598): 1526.
314. Okonjo-Iweala N, Osafo-Kwaako P. Improving health statistics in Africa. *Lancet.* 2007; 370(9598): 1527-8.
315. United Nations, African Union Commission, African Development Bank. Africa programme on accelerated improvement of civil registration and vital statistics (APAI-CRVS). New York: Economic Commission for Africa, United Nations; 2011.
316. AbouZahr C, de Savigny D, Mikkelsen L, Setel PW, Lozano R, Lopez AD. Towards universal civil registration and vital statistics systems: the time is now. *Lancet.* 2015 Online 10.1016/S0140-6736(15)60170-2.
317. Bryman A. Triangulation. In: Encyclopedia of Social Science Research Methods [Internet]. Thousand Oaks: SAGE Publications, Accessed: 7 Sep 2015. Available: <http://srmo.sagepub.com/view/the-sage-encyclopedia-of-social-science-research-methods/n1031.xml>.
318. Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull.* 1959; 56(2): 81-105.
319. Denzin NK. The research act in Sociology. Chicago, USA: Aldine; 1970.
320. AbouZahr C. Assessing and monitoring the performance of health information systems: metrics and models. Brisbane, Australia: The University of Queensland, 2013. Accessed: 17 Dec 2013. Available: [http://www.uq.edu.au/hishub/docs/WP29/WP29\\_health%20information%20systems\\_web2.pdf](http://www.uq.edu.au/hishub/docs/WP29/WP29_health%20information%20systems_web2.pdf)
321. Groenewald P, Bradshaw D, Dorrington R, Bourne D, Laubscher R, Nannan N. Identifying deaths from AIDS in South Africa: An update [4]. *AIDS.* 2005; 19(7): 744-5.
322. GSMA Mobile Identity Team. Mobile birth registration in Sub-Saharan Africa: a case study of Orange Senegal and Uganda Telecom solutions. Uganda: GSMA, Uganda Telecom, 2013. Accessed: 20 Aug 2015. Available: <http://www.gsma.com/personaldata/wp-content/uploads/2013/05/Mobile-Birth-Registration-in-Sub-Saharan-Africa.pdf>

323. World Health Organisation. Verbal autopsy standards: ascertaining and attributing cause of death. Geneva: World Health Organisation; 2007.
324. World Health Organisation, Health Metrics Network, INDEPTH Network. Verbal Autopsy Standards: 2012 WHO Verbal Autopsy Instrument. Release Candidate 1. Geneva: World Health Organization, 2012. Accessed: 4 March 2013. Available: [http://www.who.int/healthinfo/statistics/WHO\\_VA\\_2012\\_RC1\\_Instrument.pdf](http://www.who.int/healthinfo/statistics/WHO_VA_2012_RC1_Instrument.pdf)
325. Leitao J, Chandramohan D, Byass P, Jakob R, Bundhamcharoen K, Choprapawon C, et al. Revising the World Health Organization verbal autopsy instrument to facilitate routine cause-of-death monitoring. *Glob Health Action*. 2013; 6: 21518.
326. World Bank, World Health Organisation. Global civil registration and vital statistics: scaling up investment plan 2015-2024. Washington, DC, USA: World Bank Group, 2014. Accessed: 19 Aug 2015. Available: <http://documents.worldbank.org/curated/en/2014/05/19581045/global-civil-registration-vital-statistics-scaling-up-investment-plan-2015-2024>
327. James SL, Flaxman AD, Murray CJL, the Population Health Metrics Research Consortium (PHMRC). Performance of the Tariff Method: validation of a simple additive algorithm for analysis of verbal autopsies. *Popul Health Metr*. 2011; 9: 31.
328. Byass P, Calvert C, Miro-Nakiyingi J, Lutalo T, Michael D, Crampin A, et al. InterVA-4 as a public health tool for measuring HIV/AIDS mortality: a validation study from five African countries. *Glob Health Action*. 2013; 6: 22448. Epub 2013/10/22.
329. Murray C, Lozano R, Flaxman A, Serina P, Phillips D, Stewart A, et al. Using verbal autopsy to measure causes of death: the comparative performance of existing methods. *BMC Med*. 2014; 12(1): 5.
330. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJL, the Population Health Metrics Research Consortium (PHMRC). Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. *Popul Health Metr*. 2011; 9: 29. Epub 2011/08/06.
331. Byass P, Chandramohan D, Clark S, D'Ambruso L, Fottrell E, Graham W, et al. Strengthening standardised interpretation of verbal autopsy data: the new InterVA4 tool. *Glob Health Action*. 2012; 5: 1 - 8.
332. Lakkireddy DR, Basarakodu KR, Vacek JL, et al. Improving death certificate completion: a trial of two training interventions. *J Gen Intern Med*. 2007; 22(544): e8.
333. Hanzlick R. Improving the accuracy of death certificates. *Journal of the Medical Association of America*. 1995; 274(537): e8.
334. Aung E, Rao C, Walker S. Teaching cause-of-death certification: lessons from international experience. *Postgrad Med J*. 2010; 86(1013): 143-52.
335. Pieterse D, Groenewald P, Bradshaw D, Burger L, Reagon G, Rodhe J. Death certificates: Let's get it right! . *S Afr Med J*. 2009; 99(9): 643-4.



336. Lopez A, Thomason J. Civil registration and vital statistics—everybody's business but nobody's business. *Lancet*. 2013; 381(9874): 1275-6.
337. AbouZahr C, de Savigny D, Mikkelsen L, Setel PW, Lozano R, Nichols E, et al. Civil registration and vital statistics: progress in the data revolution for counting and accountability. *Lancet*. 2015 Online 10.1016/S0140-6736(15)60173-8.
338. The Lancet. CRVS systems: a cornerstone of sustainable development (Editorial). *Lancet*. 2015; 385(9981): 1917.
339. World Health Organization. Annual Epidemiological and Vital Statistics, 1939-1946. Part I, Vital Statistics and Causes of Death. Part II, Causes of Death from Notifiable Diseases. New York: Columbia University Press, 1952. Available:
340. World Health Organisation. World Health Statistics 2014. Geneva: World Health Organization., 2014. Accessed: 27 May 2014. Available: [http://www.who.int/gho/publications/world\\_health\\_statistics/2014/en/](http://www.who.int/gho/publications/world_health_statistics/2014/en/)
341. United Nations. Principles and Recommendations for a Vital Statistics System. New York, USA: United Nations, 2014. Accessed: 14 July 2015. Available: <https://unstats.un.org/unsd/demographic/standmeth/principles/M19Rev3en.pdf>
342. World Health Organization. Annual epidemiological and vital statistics 1939-1946. Part I: Vital statistics and causes of death. Geneva: World Health Organization; 1952.
343. World Health Organisation. Improving mortality statistics through civil registration and vital statistics systems: guidance for country strategies and partner support. Outcome of a technical meeting, 4 - 5 November 2014. Geneva: World Health Organisation; 2014.
344. World Health Organization. Global Burden of Disease: 2004 Update. Geneva, Switzerland: World Health Organization, 2008.
345. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859): 2095-128.
346. Murray C, Lopez A. The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. . Cambridge, MA, USA: Harvard School of Public Health on behalf of the World Health Organization and the World Bank; Distributed by Harvard University Press; 1996.
347. Black R, Cousens S, Johnson H, Lawn J, Rudan I, Bassani D, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet*. 2010.
348. Shibuya K, Scheele S, Boerma T. Health statistics: time to get serious. *Bull World Health Organ*. 2005; 83: 722-.
349. Shibuya K. Counting the dead is essential for health. *Bull World Health Organ*. 2006; 84(3): 170-1.

350. Ronsmans C, Graham WJ. Maternal mortality: who, when, where, and why. *Lancet*. 2006; 368(9542): 1189-200.
351. Mikkelsen L, Phillips DE, AbouZahr C, Setel PW, de Savigny D, Lozano R, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet*. 2015 Online 10.1016/S0140-6736(15)60171-4.
352. Phillips DE, AbouZahr C, Lopez AD, Mikkelsen L, de Savigny D, Lozano R, et al. Are well functioning civil registration and vital statistics systems associated with better health outcomes? *Lancet*. 2015 Online 10.1016/S0140-6736(15)60172-6.
353. AbouZahr C, de Savigny D, Mikkelsen L, Setel PW, Lozano R, Lopez AD. Towards universal civil registration and vital statistics systems: the time is now. *Lancet*. 10.1016/S0140-6736(15)60170-2.
354. Botha JL, Bradshaw D. African vital statistics - A black hole? *S Afr Med J*. 1985; 67(24): 977-81.
355. Van der Merwe S, Yach D, Metcalf CA. Peering into the black hole--the quality of black mortality data in Port Elizabeth and the rest of South Africa. *S Afr Med J*. 1991; 79(8): 419-22.
356. Dorrington RE. African mortality rates - an initial estimate. *Transactions of the Actuarial Society of South Africa*. 1989; 7(1): 169-202.
357. Pillay-van Wyk V, Bradshaw D, Groenewald P, Laubscher R. Improving the quality of medical certification of cause of death: the time is now! *S Afr Med J*. 2011; 101(9): 626.
358. Matzopoulos R, Prinsloo M, Pillay-van Wyk V, Gwebushe N, Mathews S, Martin LJ, et al. Injury-related mortality in South Africa: a retrospective descriptive study of postmortem investigations. *Bull World Health Organ*. 2015; 93(5): 303-13. Epub 2015/08/01.
359. Bradshaw D, Laubscher R, Dorrington R, Bourne D, Timaeus I. Unabated rise in number of adult deaths in South Africa. *S Afr Med J*. 2008; 94: 278.
360. Pillay-Van Wyk V, Laubscher R, Msemburi W, Dorrington RE, Groenewald P, Vos T, et al. Second South African National Burden of Disease Study: Data cleaning, validation and SA NBD List. Cape Town, South Africa: 2015. Accessed: 8 Sep 2015. Available: <http://www.mrc.ac.za/bod/reports.htm>
361. Lo S, Horton R. Everyone counts—so count everyone. *Lancet*. 2015;10.1016/S0140-6736(15)60305-1.
362. Marten R, McIntyre D, Travassos C, Shishkin S, Longde W, Reddy S, et al. An assessment of progress towards universal health coverage in Brazil, Russia, India, China, and South Africa (BRICS). *Lancet*. 2014;[http://dx.doi.org/10.1016/S0140-6736\(14\)60075-1\(0\)](http://dx.doi.org/10.1016/S0140-6736(14)60075-1(0)). Epub 30 April 2014.

## Notes

---

a The HDI, a composite statistic used to rank countries by level of human development, is compiled from data on life expectancy at birth, education, and per-capita gross domestic product. Life expectancy being part of this statistic exemplifies health and mortality being two sides of the same coin, and illustrates the role of mortality beyond the health sector.

b Cancer, cardiovascular diseases, chronic respiratory diseases, and diabetes.

c The DALY is a time-based measure that combines years of life lost due to premature mortality and time lived in health states less than ideal health.<sup>9</sup> Mahapatra P, Shibuya K, Lopez AD, Coullare F, Notzon FC, Rao C, et al. Civil registration systems and vital statistics: successes and missed opportunities. *Lancet*. 2007; 370(9599): 1653-63.

d “GDP per capita” is the gross domestic product of a country divided by its midyear population. GDP is the sum of gross value added by all resident producers in the economy, plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources.

