

https://dspace.org/

Causes of death stratified by HIV status in the post-art era

Mabhikwa, Mellisa Nothando

2020
https://hdl.handle.net/10539/29983

Downloaded from a DSpace repository. https://dspace.org/

CAUSES OF DEATH STRATIFIED BY HIV STATUS IN THE POST-ART ERA

A research report submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements for the degree of Master of Science in Epidemiology

Student name: Mellisa Nothando Mabhikwa

Student number: 708821

Supervisor: Dr Denise Evans

Johannesburg, June 2020

DECLARATION

I, Mellisa Nothando Mabhikwa declare that this research report is my own unaided work. It is being submitted in partial fulfilment for the degree of Master of Science in Epidemiology in the field of Epidemiology and Biostatistics, Faculty of Health Sciences at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature: 

Date: 16 June 2020

DEDICATION

This research report is dedicated to the loving memory of Moses Mabhikwa, Benjamin Chipango, Alice Chipango and Sheila Musiyamanje.

ABSTRACT

Introduction: To date, HIV and AIDS has claimed the lives of more than 34 million people worldwide. Reports claim that HIV positive patients on antiretroviral therapy (ART) have similar life expectancies to their HIV negative counterparts, but there is a lack of data comparing causes of death between the two groups. This study aimed to describe the causes of death, and more specifically HIV-associated deaths, as reported in a government public tertiary hospital mortuary register.

Methods: A cross sectional study of adult deaths reported in the Helen Joseph Hospital (HJH) mortuary register between 2014 and 2017 was performed. The recorded causes of death during six winter and six summer months over the three year period are reported. Causes of death were reported by HIV status (positive, negative and unknown) and CD4 count category (< 200 cells/ml, 200 - 500 cells/ml and > 500 cells/ml) within the 12 months before death. A sub-analysis among HIV positive individuals known to have been on ART was performed to examine factors associated with death from a respiratory illness.

Results: A total of 1766 adults (52% male; median age 53.2 years IQR 39.7 - 67.4; 43% HIV positive) were included in the analysis. Overall, infections (n = 416, 23.5%) were the leading cause of death followed by AIDS (n = 215, 14.2%). Similarly, among HIV positive individuals infections (n = 194, 25.7%) and AIDS (n = 172, 22.8%) were the main causes of death. Among HIV negative individuals, infections were the leading cause of death (n = 53, 21.3%), followed by malignancies (n = 31, 12.4%) and heart or vascular illnesses (n = 28, 11.2%). In those with an unknown HIV status, infections were the leading cause of death (n = 169, 22.8 %) followed by heart or vascular illnesses (n = 67, 8.8%) and AIDS (n = 62, 8.1%), suggesting that these individuals were HIV positive. As expected, AIDS was the main cause of death among HIV positive individuals with a low CD4 count (< 200 cells/ml) while heart or vascular illness, bacterial infection, malignancy and renal failure were equally ranked after AIDS among those with a high CD4 count (> 500 cells/ml).

Of the HIV positive individuals who could be linked to the electronic ART register and known to be on ART (n = 339), 88% (n = 299) had a viral load reported within months before death. Of these 55% were virally suppressed (n = 163). Individuals who had completed TB treatment in less than a year had a two-fold risk of dying from a respiratory illness (aOR 2.03 95% CI 1.07, 3.83) compared to those who had completed TB treatment in more than one year prior to their death.

Discussion: Consistent with other reports bacterial infections and AIDS-related infections, of which TB is the biggest contributor, were the top causes of death between 2014 and 2017. HIV prevalence was lower than other reports possibly due to underreporting of AIDS-related deaths. Poor documentation of HIV status may result in misclassification and further underestimation of AIDS-related deaths. Individuals who have recently completed TB treatment have a high rate of respiratory infection including recurrent TB. It is therefore important to allocate appropriate surveillance resources and monitor HIV positive TB patients during and after TB treatment.

ACKNOWLEDGEMENTS

I express my deepest gratitude to God for using this project to deepen my faith and connection to Him. I also thank myself for pressing on in times of discouragement.

I would like to thank my mom, sister, nephew, Sekuru Farayi and extended family for their unwavering support.

Many thanks to Dr D. Evans for always giving constructive feedback. She, through her supervision and mentorship, provided opportunities for my personal and professional growth. I learnt many lessons beyond the academic purposes of this project from her.

I whole-heartedly thank Mrs Zodwa Ndhlovu who was always available to listen and help. Her calm and warm nature provided comfort in the difficult times. I also thank Professors Jonathan Levin and Matthew F. Chersich for their encouragement and Prof. Chrissie Rey whose wisdom I will forever draw on.

While this is my own work, remarkable individuals from the Health Economics and Epidemiology Research Office (HE^2RO) helped at various stages of this project. Firstly, Dr Ribka Berhanu for giving invaluable insight in the classification of causes of death as well as other clinical technicalities and Dr Onoya for her input in the final stages of the write up. Tembeka Sineke, Caroline Govathson and Lise Jamieson were a great help in handling SAS data. My days at the HE^2RO would not have been the same without the ever-jovial Cornelius Nattey, Patience Manjengwa, Nozipho Musakwa, Nelly Jinga, Lerato, Willie and Bra David. Thanks to Barbara Xhosa and Ross Greener who never tired of my multiple requests to use the “supercomputer”.

Helen Joseph Hospital Acting CEO Dr Punwasi, the mortuary manager and staff co-operated actively in the site evaluation and during data collection. I acknowledge Mr Thomas of the NHLS for performing the matching at their organisation.

FUNDING

The Centre for Scientific and Industrial Research (CSIR) in collaboration with the Department of Science and Technology (DST) funded this degree through the CSIR-DST Interbursary award. I acknowledge the financial support provided by the Wits Postgraduate Association through the Wits Postgraduate Merit Award.

This study has been made possible by the generous support of the American people and the President's Emergency Plan for AIDS Relief (PEPFAR) through US Agency for International Development (USAID) under the terms of Cooperative Agreements AID-674-A-12-00029 and 72067419CA00004 to Health Economics and Epidemiology Research Office, a Division of the Wits Health Consortium (Pty) Ltd and do not necessarily reflect the views of USAID or the United States Government.

TABLE OF CONTENTS

DECLARATION i
DEDICATION ii
ABSTRACT iii
ACKNOWLEDGEMENTS v
FUNDING vi
LIST OF FIGURES ix
LIST OF TABLES x
NOMENCLATURE xi
ABBREVIATIONS xiii
CHAPTER 1: INTRODUCTION 1
1.1 BACKGROUND 1
1.2 LITERATURE REVIEW 2
1.2.1 Literature review methods 2
1.2.2 Mortality data sources 2
1.2.3 Mortality reporting in South Africa 3
1.2.4 All-cause mortality 4
1.2.5 Research methodology 5
1.3 PROBLEM STATEMENT 6
1.4 JUSTIFICATION 7
1.5 AIMS AND OBJECTIVES 7
1.5.1 Aim 7
1.5.2 Objectives 7
CHAPTER 2: METHODS 8
2.1 STUDY SITE 8
2.2 STUDY DESIGN 9
2.3 STUDY POPULATION 9
2.4 STUDY PERIODS 10
2.5 STUDY SAMPLE 10
2.5.1 Sampling method 10
2.6 DATA SOURCES 11
2.6.1 The HJH Mortuary Register 11
2.6.2 TherapyEdge HIV™ 11
2.6.3 NHLS Corporate Data Warehouse 12
2.7 DATA COLLECTION 12

2.8 STUDY PROCEDURES....13
2.8.1 Probabilistic linkage....13
2.9 STUDY VARIABLES....14
2.9.1 Primary outcome....18
2.10 STATISTICAL METHODS....19
2.10.1 Accuracy of the matching algorithm....20
2.10.2 Comparison of causes of death by HIV status....20
2.10.3 Comparison of causes of death by latest CD4 count category at time of death....20
2.10.4 Demographic and clinical factors associated with death from a respiratory illness: A sub-analysis of known HIV positive individuals on ART in TE....20
2.10.5 The logistic regression model....21
2.10.6 Missing data....21
2.11 ETHICAL CONSIDERATIONS....22
CHAPTER 3: RESULTS....23
3.1 Demographic characteristics of the study sample by time period of death....23
3.2 Causes of death by study period....24
3.3 Feasibility and accuracy of matching mortuary register data to large HIV databases....28
3.4 Demographic characteristics of the study sample by HIV status....31
3.5 Causes of death stratified by HIV status....32
3.6 Causes of death by CD4+ cell category within 12 months before death among HIV positive individuals....34
3.7 Death from a respiratory illness: A sub-analysis of known HIV positive individuals on ART....36
3.7.1 Factors associated with death from a respiratory illness....37
CHAPTER 4: DISCUSSION....40
4.1 Study findings....40
4.1.1 Summary of study findings....40
4.1.2 Distribution of causes of death....40
4.1.3 Causes of death by HIV status....41
4.1.4 Feasibility of matching and accuracy of the probabilistic algorithm....42
4.1.5 Factors associated with death from a respiratory illness....43
4.2 Limitations....44
4.3 Conclusion and Recommendations....45
CHAPTER 5: REFERENCES....46
APPENDICES....52
Appendix 1: Plagiarism declaration form....52

Appendix 2: Ethics clearance certificate 53
Appendix 3: NHLS data access approval 54
Appendix 4: HJH ethics approval 55
Appendix 5: Data sharing agreement for TE data access through HE^{2}RO 56
Appendix 6: Turnitin Report 60
SUPPLEMENTARY MATERIAL 62

LIST OF FIGURES

Figure 1: Johannesburg administrative regions……8

Figure 2: Distribution of bacterial infections other than AIDS-related infections in the study population……...27

Figure 3: Proportions of AIDS infections in the study population…….27

Figure 4: Proportions of AIDS malignancies in the sample…….....28

Figure 5: Record linkage process of mortuary register patients TE……...……29

Figure S1: REDCap data collection electronic case report form……..…..62

Figure S2: Classification of deaths according to the Coding Causes of Death in HIV Protocol……...…63

LIST OF TABLES

Table 1: Winter and summer months from which study participants were drawn from 10

Table 2: Sources of data for the variables used in this study 15

Table 3: Description of demographic variables collected from the mortuary register 16

Table 4: Description of the clinical variables from TE and NHLS data which were used to supplement mortuary register data 17

Table 5: Case definitions for AIDS related illnesses 19

Table 6: Demographic characteristics of the study sample collected from the HJH mortuary register 24

Table 7: Causes of death among patients who died of any cause at HJH between May 2014 and July 2017 26

Table 8: Sensitivity and specificity of the matching algorithm 30

Table 9: Demographic characteristics of mortuary register patients matched to NHLS,

TE and those that were unmatched 31

Table 10: Demographic characteristics of the overall study population and by HIV status 32

Table 11: Causes of death in the study population stratified by HIV status 34

Table 12: Causes of death by latest CD4 count category +/- 12 months before death among patients known to be HIV positive 36

Table 13: Demographic and clinical factors associated with death from a respiratory illness among known HIV positive patients on ART 38

Table S1: Percentages of missing data per variable before and after matching mortuary register data to NHLS and TE 64

Table S2: Causes of death by sex 65

Table S3: Causes of death by age group 66

Table S4: The ten leading underlying natural causes of death in South Africa 2014-2016 67

NOMENCLATURE

Cause of death – any condition which leads or contributes to death.

Contributing causes – are other significant conditions contributing to death, but not part of the direct causal sequence.

Death certificate – official records of individual deaths, containing information on the cause of death certified by a physician, or other legally appointed official, and any other required demographic information.

External causes of death – Deaths due to accidents and violence including environmental events, circumstances and conditions as the cause of injury, poisoning and other adverse effects. Broad categories include accidents, suicides, medical misadventures or abnormal reactions, homicide, legal intervention and injury from war operations.

Garbage terms – Global Burden of Disease experts coined this term for mortality data which do not signify an underlying cause of death. This includes ill-defined signs and symptoms, intermediate causes of death (e.g. septicaemia), mechanisms of death (e.g. cardiac arrest) or partially specified causes (e.g. cancer with unknown site).

Immediate cause of death – is the final disease, injury or complication directly causing the death. It should be noted that the mechanism of death or terminal event (for example, heart failure, cardiac arrest, respiratory arrest) is not considered to be a valid underlying cause of death and should not be reported on the death certificate without stating the preceding disease or injury.

Intermediate cause – any cause between the underlying cause and the immediate cause of death or a complication of the underlying cause.

Manner of death – modality/intention surrounding the death e.g. natural, accident, intentional self-harm (including suicide), assault, homicide and undetermined.

Mechanism of death – the physiological disturbance in the body at the time of death e.g. metabolic acidosis, hypokalaemia, acute cardiac failure.

Multiple causes of death – all those diseases, morbid conditions or injuries which either resulted in or contributed to death and the circumstances of the accident or violence which produced any such injuries.

Primary cause of death – or underlying cause of death is the disease or injury which started the sequence of events leading directly to death or the circumstances of the accident or violence that produced the fatal injury. In the case of a violent death, the form of external violence or accident is antecedent to the injury entered, although the two events may be almost simultaneous.

Procedure-related death – the death of a person undergoing, or as a result of, a procedure of therapeutic, diagnostic or palliative nature, or of which any aspect of such a procedure has been a contributory cause and shall not be deemed to be a death from natural causes.

Vital statistics – data collected from continuous or periodic recording or registration of all vital events such as births, deaths, marriages and divorces.

ABBREVIATIONS

AIDS – Acquired Immunodeficiency Syndrome

ART – antiretroviral therapy

BMI – body mass index

CD4+ – cluster of differentiation 4

CDW – corporate data warehouse

CoDe – Coding Causes of Death in HIV Protocol

COPD – chronic obstructive pulmonary disease

CRVS – civil registration and vital statistics

CVD – cardiovascular disease

DHA – Department of Home Affairs

ELISA – enzyme linked immunosorbent assay

HAART – highly active antiretroviral therapy

HIV – human immunodeficiency virus

HJH – Helen Joseph Hospital

ICD-10 – International Classification of Diseases 10

NDoH – National Department of Health

NHLS – National Health Laboratory Service

NNRTI – non-nucleotide reverse transcriptase inibitor

PCP – pneumocystis pneumonia

PLHIV – people living with HIV

RNA – ribonucleic acid

SA – South Africa

StatsSA – Statistics South Africa

TB – tuberculosis

TE – Therapy-Edge HIV™

TLC – Themba Lethu Clinic

USA – United States of America

VL – Viral load

WHO – World Health Organisation

CHAPTER 1: INTRODUCTION

This chapter begins with a background of the research problem, followed by a review of the literature. Also presented in this chapter is a summary of the published literature on the relatively new and uncommon research methods used in this study (i.e. probabilistic linkage of patient records and the Coding Causes of Death in HIV protocol). Lastly, the aim and specific objectives of the research project are stated.

1.1 BACKGROUND

Information about frequency and causes of death broken down by age, sex and year provide a starting point for informed health policy debate. However, drawing meaningful conclusions about mortality in a population involves addressing many data challenges, which include reconciling marked discrepancies in cause of death classifications over time and across populations and adjusting for vital registration system data with coverage and quality issues appropriately.

The Department of Home Affairs (DHA) SA is responsible for mortality reporting under the governance of the Births and Deaths registration Act 1992 [Act No. 51 of 1992] (1). DHA works in collaboration with Statistics South Africa (StatsSA), which releases annual reports on vital statistics (1). StatsSA is a reliable source of mortality data; however, causes of death stratified by HIV status are not given in StatsSA reports. South Africa has one of the world's largest HIV treatment programs, with an estimated 2.6 million patients on treatment in 2016. Key indicators used to measure the success of this program include an increase in life expectancy and a reduction in HIV-related mortality for patients on antiretroviral therapy (ART). Tracking changes in the contribution of HIV to overall mortality over time is thus an important component of evaluating the national HIV treatment program. Much is known about causes of death in HIV positive patients but very few studies have considered the similarities and differences in causes of death among HIV positive and HIV negative patients in the post-ART era.

People Living with HIV/AIDS (PLHIV) continue to have an increased frequency of respiratory illness compared with the general population. There is evidence from multiple sources of impaired lung function in people with long-standing HIV infection (2, 3). Although most current evidence comes from cohorts in the United States of America (USA) or Europe and less is known about populations in other settings. Higher rates of respiratory pathology despite ART have been a consistent finding of studies among PLHIV. There are

likely to be several reasons for this, including higher rates of smoking and recreational drug use, lung injury occurring prior to ART and possible consequences of the ongoing immune activation and disordered immune responses associated with HIV. We sought to provide a comprehensive description of factors associated with death from a respiratory illness in HIV positive patients on ART in an urban setting.

Understanding how the causes of mortality differ by HIV status will allow appropriate implementation of preventative measures for mortality. The aim of this study was to describe causes of mortality, and then more specifically trends in causes of death and HIV status over time, among deceased individuals at the Helen Joseph Hospital (HJH) mortuary in Johannesburg, SA.

1.2 LITERATURE REVIEW

1.2.1 Literature review methods

We searched PubMed and Google Scholar for recently published journal articles using the search terms "causes of death", "all-cause mortality", "mortality", "HIV mortality", and "HIV deaths". We also drew on policy documents on the ART programme released by the South African National Department of Health (NDoH) and reports from relevant bodies such as StatsSA and the World Health Organisation (WHO).

1.2.2 Mortality data sources

Mortality statistics, including causes of death, are the foundation of public health planning, monitoring and evaluation of interventions (1). Globally, HIV/AIDS programmes have developed strategic plans to scale up prevention, treatment and care. It is critical to know how effective these programmes have been in reducing mortality. The overwhelming majority of low- and middle-income countries do not have reliable mortality statistics (4). Out of 119 countries reporting causes of death to the World Health Organization (WHO), only 34 countries – representing 15% of the world population – produce high-quality cause-of-death data, and almost all of these countries are in Europe and the Americas (4, 5). A further 85 countries – representing 65% of the world population – produce lower-quality cause-of-death data, while 74 countries, mostly in sub-Saharan Africa, lack such data altogether (5). South Africa and Egypt are the only countries in Africa who reported cause of death data to WHO (5). This information paradox – where information is lacking where it is needed most – hinders governments and country programmes to track progress of HIV programmes.

A well-functioning civil registration and vital statistics (CRVS) system is the best way to monitor mortality and causes of death (6, 7). Civil registration systems aim to register all the births, deaths with cause of death, and marriages. Many countries do not have civil registration systems that produce reliable vital statistics, as coverage tends to be low, especially for deaths, and information on causes are unreliable (8). Even though there is increasing awareness and commitment to strengthening CRVS systems, it will take considerable time before these systems can produce reliable mortality statistics (9). Therefore, alternative methods to collect information on causes of death, including HIV, need to be considered. Such methods include sample registration systems, household surveys, burial systems, clinical autopsy, analytical methods and hospital data (7).

In the absence of quality CRVS data, global and regional estimation methods are used to determine leading causes of death (7). In 2012, an estimated 1.6 million (range 1.4 to 1.9 million) people died from HIV-related causes, including 1.2 million (range 1.1 to 1.3 million) in sub-Saharan Africa (4). These estimates are based on country data that have been analysed in statistical models such as Spectrum which use standard methods developed to estimate the course of the HIV/AIDS epidemic (4). Therefore, no direct HIV/AIDS mortality data are used to obtain global, regional and country estimates of HIV/AIDS mortality. In countries with reliable mortality registration data, it is necessary to compare the plausibility of the model-based mortality estimates with results from death registration (8).

South Africa is one of the few countries with a high HIV burden that have a well-functioning national CRVS system (5). However, a national CRVS system does not automatically imply that reliable information on HIV/AIDS as a cause of death is available. Incorrect application of the standard principles and rules on death certification and coding may result in underreporting of HIV/AIDS deaths, as is the case in South Africa, where the accuracy of death certification and HIV coding is of questionable quality (10). Monitoring causes of death in HIV-positive people enables the appropriate targeting of interventions to improve the quality of inpatient care and reduce avoidable mortality.

1.2.3 Mortality reporting in South Africa

The DHA SA is responsible for mortality reporting under the governance of the Births and Deaths registration Act 1992 [Act No. 51 of 1992]. DHA works in collaboration with StatsSA, which releases annual reports on vital statistics. StatsSA is a reliable source of mortality data; however, all-cause mortality stratified by HIV status is not given in StatsSA

reports (1). More descriptive sources of mortality data are needed to better understand the causes of mortality in PLHIV in the post-ART era.

Health data for patients often exists in multiple sources and record linkage techniques have proved to be useful in creating longitudinal databases with a single patient record from multiple data sources (11). In a study to describe the persistent high burden of HIV disease among patients seeking care in South Africa's national HIV program, probabilistic matching was used to link CD4 count and viral load monitoring tests in the National Health Laboratory Services (NHLS) Corporate Data Warehouse (CDW) using patient name, surname and date of birth (12). The sensitivity of the method was 75% (12).

1.2.4 All-cause mortality

All-cause mortality is defined as death from any cause and it is an important outcome in epidemiological research. All-cause mortality is decreasing among adults in all world regions. Between 1990 and 2015, Africa experienced a 23% decline in adult mortality while its adult population doubled from 319 million to 640 million in the same time frame (13). In 2015, the percentage of deaths at ages 65 years and over reached 55% worldwide, up from around 41% in 1990. In Africa, with lower life expectancy at birth and higher mortality risks at all ages, the percentage of deaths at older ages (>65 years) remains low, from 16% in 1990 to 25% in 2015 (13). In Southern Africa, South Africa had the highest percentage of deaths in the 25-65 and over 65 age groups at 53% and 33% in 2015 respectively (13).

In 2016, ischaemic heart disease, stroke and chronic obstructive pulmonary disease, lower respiratory infections, Alzheimer's disease, lung cancer, diabetes mellitus, road injury, diarrhoeal disease and TB were the top ten causes of death worldwide (7). The top ten causes of death in South Africa for the same year were TB, diabetes mellitus, heart disease, cerebrovascular disease, HIV, hypertensive disease, pneumonia, other viral diseases, and chronic lower respiratory diseases. The rise in rank of non-communicable diseases as causes of death in SA reflects how rapid changes in social and economic development have resulted in the adoption of lifestyle practices which expose people to risk factors for non-communicable diseases (5). TB remained the number one cause of death in SA in the 2014-2016 year period, while HIV moved from sixth position in 2014 to fifth position in 2016 (1).

1.2.4.1 HIV mortality

Globally HIV related mortality has declined by 33% since 2010, while SA estimates have shown an unclear pattern over the years because of difficulties in accurately estimating HIV attributable deaths (14). With the increases in survival rate among PLHIV associated with

increased uptake of ART, the HIV positive population is growing older and facing new challenges in the form of co-morbidities such as diabetes mellitus, hypertensive diseases and cardiovascular diseases (CVD) (15). Studies have shown an increased risk of CVD among HIV positive patients on ART compared to their HIV negative counterparts (16, 17) and certain ART regimens (PI-based and stavudine-based) have been associated with increased risk of CVD (17). There are known toxicities associated with certain ART regimens, but the long-term effects of ART on mortality are not clear since it is barely 30 years since the introduction of ART (18).

1.2.4.2 Mortality from respiratory illnesses

HIV-associated lung complications are frequent causes of illness and death in industrialized nations (19). Previous epidemiologic analyses have demonstrated an increased risk of chronic obstructive pulmonary disease, as well as a significant burden of respiratory symptoms in HIV-infected patients (20). HIV infection causes alteration in several lines of host defences in the lung and respiratory tract that contribute to an increased risk for lung complications. These alterations include abnormalities in mucociliary function and soluble defence molecules, such as defensins within respiratory secretions (21). Within the lung parenchyma, innate and adaptive immune responses to pathogens may be impaired. For example, alveolar macrophages from HIV-infected individuals have been shown to be deficient in pathogen recognition (22). HIV also results in chronic stimulation and activation of inflammatory cells within the alveolar space (21). Most low- and middle-income countries including SA still show high rates of mortality from communicable respiratory diseases such as PCP and TB among the HIV positive population. The HIV-TB co-infection rate is estimated at 72%, and is proving to be a difficult challenge to solve as HIV infection is a risk factor for TB (23, 24).

1.2.5 Research methodology

1.2.5.1 Record Linkage

Record linkage is the process of identifying and merging records in different databases which refer to the same individual (11, 25, 26). The practice is gaining popularity as routine data sources are proving useful in supporting public health research (26). There are two types of record linkage techniques, namely deterministic and probabilistic matching. In deterministic record linkage, data in the different sources must be the same if it belongs to one individual whereas in probabilistic linkage, a set of rules are set to determine the probability that different records belong to the same individual (26). Probabilistic linkage is based on the notion that some identifiers have more discriminatory power than others (11). For example,

matching on Sex does not carry the same weight as matching on last name (27). Matching weights are a good indication of quality of matches, but statistical matching algorithms are not perfect (25). Linkage decisions of match, non-match, or leave for clerical review must still be made to complete the record linkage. This is done by setting threshold values to create three regions for weights falling into each region to be classified according to the rule for that region (26). Errors of this method are false linkages where records are matched when they should not be and missed matches where records that should be matched are not matched (27).

1.2.5.2 Cause of death classification systems

Surveillance of causes of death associated with HIV is critical for targeting interventions for HIV patients at various stages of disease (15). However, the heterogeneity of cause of death classification systems makes it difficult to compare results from different studies (28). The WHO International Classification of Diseases (ICD-10) system is widely used but does not cover some AIDS-defining diseases and adverse effects of ART (29). For example, some central nervous system diseases have a different aetiology in HIV infected patients and are therefore not covered by the ICD-10 and are at risk of misclassification (29). Following this realisation, a group of medical experts launched the Coding Causes of Death in HIV Protocol (CoDe), which is a standardised international coding and classification system for deaths (30). The CoDe process involves use of a case report form to collect sociodemographic and clinical data of the case followed by a matched review to determine the sequence of events that led to death. The death is then classified as related or not to HIV. The CoDe method has been used in Brazil and Europe (28).

1.3 PROBLEM STATEMENT

Cause of death statistics are essential for monitoring the health of populations and identifying public health priorities. A review of the quality of cause of death data from 115 WHO member countries in 2010 categorised South Africa as having low quality cause of death data based on a combination of a more than 20% of deaths being ill-defined deaths and less than 70% of deaths being registered (31). HIV reporting in SA has been poor because health care providers were reluctant to record HIV as cause of death due to confidentiality issues regarding the death certificate and some life insurance policies included exclusion clauses for HIV (32). The extensive misclassification of HIV as an underlying cause made it necessary for the use of modelling approaches to estimate the extent of HIV mortality (33). However,

data discrepancies between StatsSA and modelled estimates suggest that other sources of information are needed to understand the role of HIV in mortality.

1.4 JUSTIFICATION

Causes of death can be used to inform prevention measures which reduce mortality and can be indicators of major causes of death in the population. Studying causes of death by HIV status or different CD4 cell count categories can help in public health planning including design of interventions to help reduce mortality in those at risk (HIV positive individuals with advanced disease). There has been evidence of ART inducing susceptibility to respiratory infections (2, 19, 20, 22). Understanding risk factors of death from a respiratory illness will allow public health planners to consider solutions such as ART adherence and TB treatment completion among HIV patients on ART. The mortuary register is a reliable source of complete mortality data from facilities with mortuaries which can be used to supplement other sources.

1.5 AIMS AND OBJECTIVES

1.5.1 Aim

The aim of this study was to describe causes of mortality, and then more specifically trends in causes of death and HIV status over time, among deceased individuals at the Helen Joseph Hospital (HJH) mortuary in Johannesburg, SA.

1.5.2 Objectives

1. Describe the causes of death among patients who died of any cause at HJH between May 2014 and July 2017.
2. Determine the HIV status of patients who died of any cause at HJH between May 2014 and July 2017, and in doing so assess the completeness of data and feasibility of matching patients in the mortuary register to large HIV databases (e.g. TherapyEdge-HIV™ (TE), NHLS database).
3. Compare the causes of death by
    a) HIV status (i.e. HIV positive, negative and unknown status)
    b) Most recent (+/-12 months) CD4+ cell count among patients known to be HIV positive.
4. Identify demographic and clinical factors associated with deaths from a respiratory illness: A sub-analysis among HIV positive patients on ART.

CHAPTER 2: METHODS

This chapter provides a detailed description of the study design and research techniques used to answer the study question. Descriptions of sources of data and definitions of variables are given. The chapter ends with an in-depth description of statistical methods employed to analyse the data as per the objectives outlined in the previous chapter.

2.1 STUDY SITE

Gauteng is the smallest province in area in SA, yet richest and largest in population. While Gauteng presents major economic opportunities (which attracts local immigrants from other provinces in South Africa and international immigrants from the rest of Africa and the world), the province has high crime rates including murder and fraud (34).

The province is divided into three metropolitan municipalities (City of Johannesburg, City of Ekurhuleni and City of Tshwane) and two district municipalities (Sedibeng and Westrand). The district municipalities are further divided into six local municipalities namely Emfuleni, Lesedi, Midvaal, Merafong City, Mogale City and Rand West City (34). Gauteng province has an estimated population of 14,700,000 of which 51% is female and median age is 27 years (34).

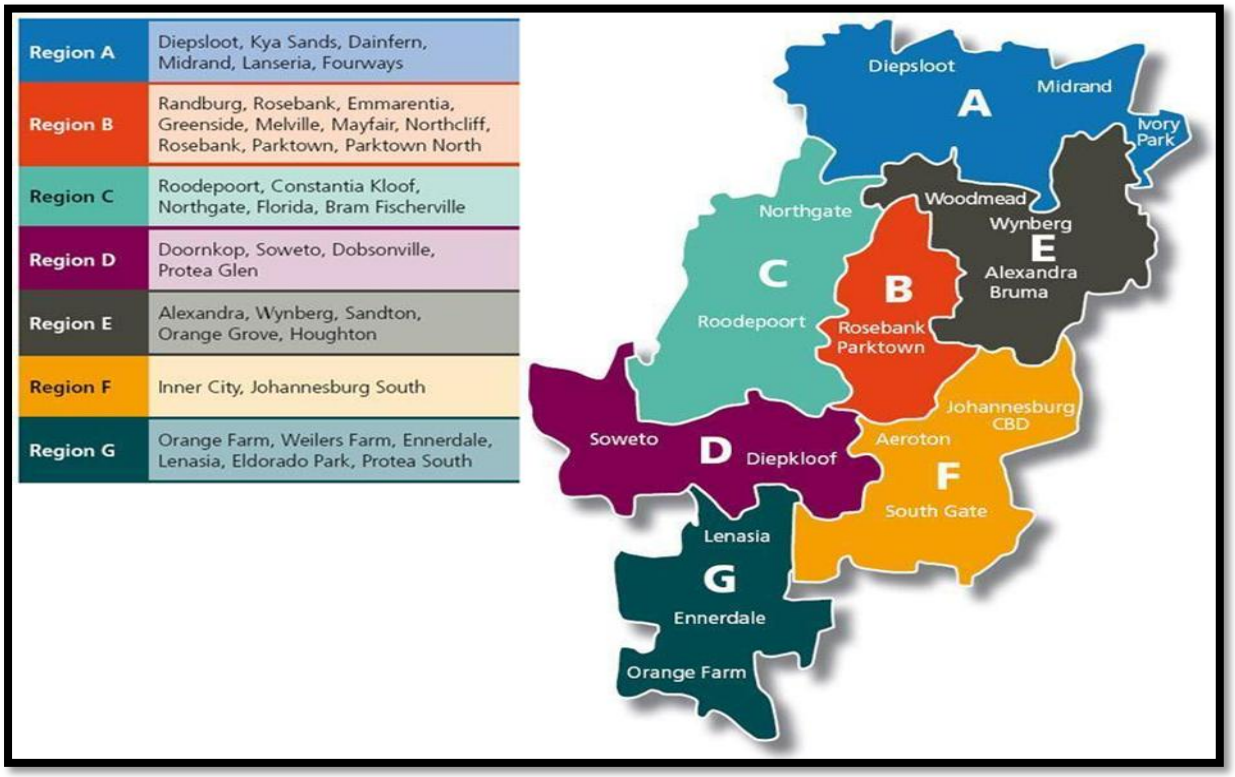


Figure 1: Johannesburg administrative regions http://www.pikitup.co.za/contact-us/.

The HJH is in a medium-density suburb called Westdene, which falls under the City of Johannesburg Metropolitan Municipality in the Gauteng Province of South Africa. The hospital serves the low to middle income residents of Regions B, C and D of the City of Johannesburg Metropolitan Municipality (Fig. 1). HJH has two inpatient wards, six surgical wards, a psychiatric unit, a ten bed ICU, a 12-bed high care/step down unit, a theatre complex comprising of nine functional units, speciality clinics including stoma unit, renal dialysis unit, pain clinic, endoscopy unit and breast clinic. At the back of the hospital is a mortuary where deceased patients are kept for collection by their next of kin or family and friends.

HJH also houses the Themba Lethu HIV Clinic (TLC), the largest antiretroviral therapy (ART) initiation and on-going treatment site in South Africa, with over 30 000 HIV-infected patients in HIV care and over 21 000 on ART (35, 36). After treatment initiation, patients are typically scheduled for medical follow-up visits at 1, 3, 6, and 12 months and then 6-monthly thereafter (35). Since the implementation of the 2010 national treatment guidelines, monitoring laboratory tests such as viral load tests are performed at 6 and 12 months post-initiation, and annually thereafter (37-40). Additionally, CD4+ cell count tests are performed 12 months following ART initiation, and annually thereafter. Complete blood counts, haemoglobin, creatinine clearance, and liver function tests are performed at least annually (37-40).

2.2 STUDY DESIGN

This study was a cross sectional study (an observational study where the outcome and exposures in the study participants are measured at the same time) of adult deaths reported in the Helen Joseph Hospital (HJH) mortuary register between 2014 and 2017.

2.3 STUDY POPULATION

The study population comprised of included deceased patients (outlined in section *2.4.1*: *Sampling method*) at the HJH mortuary who died between May 2014 and July 2017. Study data were extracted from the mortuary register in July 2018. HJH is an adult-oriented facility. Children from around and within and around Regions B, C and D of the City of Johannesburg Metropolitan Municipality are admitted to the Raheema Moosa Mother and Children Hospital (RMMCH) and are sent to the RMMCH mortuary when they die. RMMCH is located within a two-kilometre radius of the HJH, in the neighbouring suburb of Coronationville. As such, we excluded participants in the study population whose age was below 18 years.

2.4 STUDY PERIODS

The HJH mortuary register is kept for five year periods at a time, and at the time this study was conducted the 2013-2017 register was available. We selected three winter (May, June and July) and three summer months (October, November and December) between 2014 and 2017 to capture seasonal variations in deaths as there is increased mortality overall and due to unnatural causes in winter compared to the summer season (Table 1).

Table 1: Winter and summer months from which study participants were drawn.

	Month of death												
Time period	January	February	March	April	May	June	July	August	September	October	November	December	
Period 1 (2014)					■	■	■			■	■	■	
Period 2 (2015)					■	■	■			■	■	■	
Period 3 (2016)										■	■	■	
Period 3 (2017)					■	■	■						

Note: Shaded boxes represent the months and seasons from which study participants were selected within each year of the study time period.

2.5 STUDY SAMPLE

At the time the study was conducted, HJH had approximately 2200 medical admissions per month and between 150-180 deaths per month. According to StatsSA, 30% of all deaths in 2014 were HIV deaths (1). A sample size table for two proportions (available from http://www.statstodo.com/SSiz2Means_Tab.php, accessed on 04/02/2018) was used to calculate the sample size required to detect at least a 10% change in proportions of AIDS deaths with 80% power between any two time periods. Given that previous probabilistic linkage methods in SA indicated at least 75% sensitivity (41), the sample size was inflated to account for 25% missing data due to possible unsuccessful linkage leading to a final sample size of 1780 deaths.

2.5.1 Sampling method

Study participants were sampled from the HJH mortuary register. Deaths in the register are numbered chronologically, beginning on the 1st of January of each year and ending on the 31st of December. The first corpse received on the 1st of January of each year is assigned the number 1 up to number *x* for the last corpse received in that year on the 31st of December. A random sample of 1780 deaths over the study periods was required. There was a total of 18

months over the three study periods to draw study participants from, translating to at least 99 deaths from each month of the study periods. A set of 99 random numbers were generated using the *gen double x = (B - A) * runiform ( ) + A* formula in STATA 14 to select the study participants per selected month, where A = number assigned to the first corpse received at the mortuary in that month and B = number assigned to the last corpse received at the mortuary in that month. The list of 99 random numbers was used to select 99 deaths from the register per selected month.

2.6 DATA SOURCES

2.6.1 The HJH Mortuary Register

The primary data source was the standard paper-based mortuary register, provided by the NDoH and kept at the HJH mortuary. Records of the register are only kept for five years. Entries are numbered according to the date of death, with the number "01" assigned to the first patient received at the mortuary on the 1st of January in that year to the last deceased patient received on the 31st of December. For each deceased patient, demographic details (i.e. name or initial, surname, Sex, date of birth, date and time of death, next of kin), admission details (such as ward and attending practitioner) and cause of death details are entered by the attending health practitioner. Ideally the primary cause of death, which is the disease or injury which started the sequence of events leading directly to death or the circumstances of the accident or violence that produced the fatal injury, should be recorded in the mortuary register, but in some cases the immediate cause of death (the final disease, injury or complication directly causing the death) is reported.

2.6.2 TherapyEdge HIV™

TE is an electronic patient management system in which data on patients receiving HIV care is maintained in a longitudinal observational clinical database. Data are updated regularly with outpatient and inpatient clinical documentation (the clinician enters admission data such as date of admission and reason for admission if they know of a patient's admission, otherwise admission details are self-reported by the patient) and laboratory test results (results of laboratory tests done at the NHLS are automatically downloaded into TE daily). Demographics, laboratory test results, ART regimens and conditions are collected in the clinical records by the medical provider and support staff and data capturers enter data into TE.

TE often relies on family reports of vital status of patients (35), so the dates of death may not be accurate (in some cases only the month and year is given and data capturer enters "01" or "15" for the day because it is not known). Previous attempts to link TE to the national vital statistics showed that only about 70% of patients had a valid national ID which limited the linkage and ability to verify their vital status (42). If deaths were not reported by families, death status or death date could not be updated on TE– potentially underestimating death status. For these reasons it seemed more reasonable to use the death date in the mortuary register.

2.6.3 NHLS Corporate Data Warehouse

The NHLS is a national government institution which comprises of a network of 265 laboratories which provide diagnostic pathology services for the public sector in SA. NHLS diagnostic laboratories are found in provincial and district hospitals where they play a major role in the national HIV and AIDS treatment programme through HIV testing, CD4+ cell testing and viral load reporting to monitor the HIV treatment programme. Patients undergoing HIV treatment at public health centres in SA have routine blood samples sent to a NHLS laboratory for testing. Healthcare workers at public health facilities complete laboratory requisition forms which accompany each sample submitted to the NHLS Corporate Data Warehouse (CDW). All data captured on the requisition forms, including patient identifiers, name of facility, date of sample and tests requested, are sent to the CDW and are captured electronically in the NHLS information system in real time.

2.7 DATA COLLECTION

A random sample of 1780 deaths was selected from the mortuary register as described under sub-section *2.4.1: Sampling methods*. Patient demographics and cause of death data were captured from the mortuary register using a REDCap (43) electronic case report form (eCRF) hosted at University of the Witwatersrand on a mobile tablet (Figure S1). The eCRF was designed after performing a site evaluation visit at the mortuary to assess the variables collected in the mortuary register and plan data collection activities. Range checks on key variables where enabled to ensure good data quality on entry. For example, identifying variables such as name, surname and date of birth were set as mandatory required fields. Other quality checks included an alert to flag date of births occurring after date of death.

2.8 STUDY PROCEDURES

2.8.1 Probabilistic linkage

Prior to carrying out the matching exercise, the demographic data from the mortuary register was cleaned to improve the efficiency of the matching. We then created variables containing the first initial of the first name; first three letters of name and first three letters of surname from sub-stringing the name and surname respectively. Dates of birth were split into separate columns of day, month and year of birth and variables containing the month and year of birth only, month and year of birth swapped around were created.

2.8.1.1 Matching criteria used for linkage between the mortuary register and TE

Patients were matched to TE records using their full first name, full surname, full date of birth and sex (Fig. 5). For patients who still remained unmatched, we (i) used first three letters of the name and first three letters of the surname, full date of birth and sex, (ii) used first letter of name and first three letters of surname, month and year of birth only and sex (iii) full date of birth, sex and swapped first name with surname and vice versa (since it is common for people to have names which could be surnames and vice versa e.g. Lee Thomas) (25). During these iterations, low scoring matches (scoring between 0.6 and 0.7) were discarded and high scoring matches (scoring above 0.9) were considered exact matches. A thorough manual review of the matched records was performed and records where sex was the same and first name and surname were similar (allowing for spelling differences but similar phonetically) were considered as matches. Where dates of birth were not the same, we considered switching around the month of birth and day of birth. If the dates of birth were still not similar, it was considered a non-match.

2.8.1.2 Matching criteria used for linkage between the mortuary register and NHLS

A list of all 1780 deceased HJH patients with corresponding patient identifiers (first name, surname, date of birth, sex) was sent to the NHLS for matching of laboratory test records. We specified the criteria for matching and the matching was performed by the data analyst at NHLS CDW. An internal participant ID to identify each patient was included so that the data analyst could determine which records matched those we requested. The NHLS extracted data in August 2018.

2.8.1.3 Resolution of data discrepancies between data sources

For patients who were matched to both NHLS and TE, it was not uncommon to have disagreeing information with respect to sex and race. For example a patient could be recorded as male in NHLS, but female in TE. Such discrepancies in sex and race were settled using mortuary register data. Mortuary register variables are more accurate as they are entered

directly from the admission forms. Where sex or race were missing in the mortuary register, TE data took precedence as it is updated regularly from patient files and NHLS test results are uploaded into the TE database daily.

2.9 STUDY VARIABLES

The primary data source was the mortuary register from which demographic variables and cause of death data were obtained. Other variables i.e. clinical variables were obtained from the supplementary data sources TE and NHLS (Table 2). Individuals with positive HIV enzyme linked immunosorbent assay (ELISA) screening and confirmatory test results in NHLS were classified as HIV positive. Those with negative ELISA screening and confirmatory test results were classified as HIV negative. Where there were disparities between screening and confirmatory test results, the confirmatory result was used. Individuals were also classified as HIV positive if there was a record of any CD4 and viral load test results. Individuals who were linked to TE were classified as HIV positive because TE is an ART monitoring register, presence of an ART start date further confirmed the HIV positive status.

HIV viral load and CD4+ cell count test results were also used to confirm a positive HIV status.

HJH is an adult-oriented facility. Children from around and within and around Regions B, C and D of the City of Johannesburg Metropolitan Municipality are admitted to the Raheema Moosa Mother and Children Hospital (RMMCH) and are sent to the RMMCH mortuary when they die. RMMCH is located within a two-kilometre radius of the HJH, in the neighbouring suburb of Coronationville. As such, we excluded participants in the study population whose age was below 18 years.

Table 2: Sources of data for the variables used in this study.

Variable	Available in mortuary register	Available in NHLS data	Available in TE data
Sex	Yes	Yes	Yes
Race	Yes	No	Yes
Hospital ward	Yes	No	No
Date of birth	Yes	Yes; used to supplement mortuary register data where missing	Yes; used to supplement mortuary register data where missing
Date of death	Yes	No	Yes; family reported date of death
Time period of death	Calculated variable-sorting of dates of death into the time periods	No	No
Age	Yes	No	Calculated from date of birth and family reported date of death
Cause of death	Yes	No	No
HIV status	No	Yes	Yes
CD4+ cell count	No	Yes	Yes
Viral load count	No	Yes	Yes
WHO stage (most recent +/- 12 months)	No	No	Yes
Regimen type	No	No	Yes
Completed TB treatment in the last 12 months	No	No	Yes; calculated from date of TB treatment end date and date of death
Haemoglobin counts (most recent +/- 12 months)	No	No	Yes
BMI (most recent +/- 12 months)	No	No	Yes

Note: HIV – Human Immunideficiency virus; TB – tuberculosis; ART – antiretroviral treatment; WHO – World Health Organisation

Demographic variables and death details collected from the mortuary register are shown in Table 3. Name and surname were collected solely for the purposes of matching patients between data sources, after the matching process was complete these identifiers were removed from the dataset and kept in a separate, password protected linking file.

Table 3: Description of demographic variables collected from the mortuary register.

Variable	Variable source	Description
Death ID	mortuary register	A unique numeric identifier assigned on data entry and derived from a combination of the patient's allocated number in the mortuary register and their month and year of birth.
Sex	mortuary register	sex of patient, either male or female; categorical variable.
Date of birth	mortuary register	The date on which the patient was born; date variable.
Date of death	mortuary register	The date on which the patient died; date variable.
Time period of death	mortuary register	Created by grouping deaths which fell in each of the three time periods.
Age	mortuary register	age of patient at death, calculated from date of death and date of birth; continuous variable.
Age category	mortuary register	Age was categorised into 5-year age groups (except for the 18-19-year age band) since mortality data are often categorised into 5-year age groups with an open-ended interval for the elderly (65+) (44); categorical variable.
Race	mortuary register	Ethnicity of the patient black, white, coloured or Indian. Patients whose race had been recorded as "Muslim" or "Hindu" in the mortuary register were classified as "Indian" as the Islam and Hinduism religions are common among Indians in SA (45); categorical variable.

Note: ID – identification number; SA – South Africa

Clinical variables such as HIV status, viral load and CD4 count to supplement mortuary register data were obtained from TE and NHLS (Table 4) for individuals who had been successfully matched through record linkage. Availability of ART regimen and ART start date in TE was used to confirm that matched HIV positive individuals had been receiving treatment.

Table 4: Description of the clinical variables from TE and NHLS data which were used to supplement mortuary register data.

Variable	Source	Description
Ward	Mortuary register	If a patient had been admitted in HJH, the ward was recorded; categorical variable (medical, emergency, intensive care unit (ICU), psychiatry, surgical, theatre, orthopaedic).
HIV status	NHLS, TE	categorical variable for HIV negative, positive or unknown; For patients matched to TE: Patients were classified as HIV positive if they had a start date for triple combination ART in TE. For patients matched to NHLS: Patients were classified as HIV positive if they had either a positive HIV_1/_2 antibody/antigen confirmatory test result or a positive HIV_1/_2 rapid test result or they had CD4 or viral load test results indicative of HIV infection monitoring.
Completed TB treatment in the last 12 months?	TE	Binary variable (yes/no); calculated time from date of death and date of TB treatment completion then grouped patients who had completed TB treatment within one year from death under "yes" and "no" for those who had completed treatment in more than one year from death.
CD4+ cell count	NHLS, TE	Continuous variable; we collected CD4+ cell count per millilitre of blood results done +/- 12 months from date of death. Where data was missing for the specified time range, we took the last result available.
CD4+ category	NHLS, TE	Categorical variable; CD4 counts were grouped into CD4+ < 200 cells/ml, 200 ≥ CD4+ cells/ml < 500 cells/ml and CD4+ ≥ 500 cells/ml for advanced HIV disease, moderate and healthy respectively (12).
Viral load	NHLS, TE	Continuous variable, in units of copies per ml. We were interested in tests done +/- 12 months from death and where data was missing, we extended the time range to include the latest viral load result available.
Viral load category	NHLS, TE	Most recent +/- 12 months; categorical variable; suppressed VL < 1000 copies/ml, unsuppressed VL ≥ 1000 copies/ ml (12).
ART regimen type	TE	Categorical variable of the type of ART regimen which study patients were taking at time of death, whether standard first line ART or second line ART.
Haemoglobin counts	TE	Haemoglobin cell counts (continuous variable; most recent +/- 12 months).
Haemoglobin category	TE	The continuous haemoglobin counts were categorised into normal, mild and severe anaemia

		For females: normal Hb >120g/l; mild anaemia 110 - 119 g/l severe anaemia Hb <80 g/l For males: normal Hb >130 g/l; mild anaemia 110 – 129 g/l severe anaemia Hb <80 g/l.
BMI	TE	BMI (continuous variable; most recent +/- 12 months).
BMI category	TE	The continuous BMI values were categorised underweight (BMI < 18.5 kg/m^2, normal weight (18.5 - 25 kg/m^2, overweight (25 - 30 kg/m^2) and obese (BMI > 30 kg/m^2).

Note: HJH – Helen Joseph Hospital; HIV – Human Immunodeficiency virus; NHLS – National Health Laboratory Service; TE – TherapyEdge-HIV™; TB – tuberculosis; ART – antiretroviral treatment; VL – viral load; WHO – World Health Organisation; Hb – haemoglobin; BMI – body mass index

2.9.1 Primary outcome

The primary outcome of interest in this study is the cause of death recorded in the HJH mortuary register. In cases were more than one cause of death was recorded, the other causes were entered in separate columns.

Since causes of death were assigned by attending medical practitioners at death, we categorized the causes of death recorded in the mortuary register according to the CoDe classification system (Fig. S2) which is a standardised international coding and classification system for deaths (30, 46). The classification was performed by two independent reviewers, including a clinician. Where there was disagreement between the two reviewers, the justified opinion of the clinician was favoured.

Where there was more than one cause of death, we relied on the clinical expertise of the clinician to decipher the chronological sequence of illnesses and select the most appropriate underlying cause. Causes of death which occurred frequently within the broad categories of bacterial infections (other than 01.1 AIDS infection), AIDS infections and AIDS malignancies frequently were described in detail in the Results section (Figs. 2-4)

We defined AIDS related illnesses according to the Centre for Disease Control (CDC) case definition (47) which describes AIDS defining illnesses as serious, life threatening diseases which occur in advanced HIV infection (Table 5). Any other bacterial infections not mentioned in Table 5 below were classified as Bacterial infections (other than 01.1: AIDS infection).

Table 5: Case definitions and clinical presentations of illnesses which are considered AIDS-related.

Illness/ Infection	**Characteristics**
Candidiasis	Oesophageal, bronchi, trachea, or lungs
Cervical cancer	Invasive
Coccidioidomycosis	Disseminated or extrapulmonary
Cryptococcosis	Extrapulmonary
Cryptosporidiosis	Chronic intestinal (>1 month's duration)
Cytomegalovirus disease/ retinitis	Other than liver, spleen or lymph nodes
Lymphoma	Burkitt's, immunoblastic, primary of brain
Mycobacterium avium complex or *Mycobacterium kansasii, Mycobacterium tuberculosis, Mycobacterium* other species	any site (pulmonary or extrapulmonary)
Pneumocystis jiroveci pneumonia; *Pneumocystis carinii* pneumonia	disseminated or extrapulmonary,
Encephalopathy	HIV-related (including AIDS Dementia complex)
Herpes simplex virus	Chronic ulcers (>1 month's duration) bronchitis; pneumonitis or esophagitis
Histoplasmosis	At a site other than exclusively pulmonary or cervical lymph nodes
Isosporiasis	Chronic intestinal, persisting for >1 month
Kaposi's sarcoma	Human herpes virus 8-related; (mucocutaneous or visceral)
Leukoencephalopathy	Progressive multifocal
Pneumonia	Recurrent bacterial (2 documented episodes within one year of each other)
Salmonella	Septicemia, recurrent (2 documented episodes within one year of each other) and non-typhoid
Toxoplasmosis of brain	Encephalitis onset at age > 1 month
Wasting syndrome	HIV-related; weight loss (over 10% of baseline) with no other cause and 30 days or more of either diarrhoea or weakness with fever

Source: Centers for Disease Control and Prevention. Revised surveillance case definition for HIV infection—United States, 2014. MMWR Recomm Rep. 2014;63(RR-03):1-10

2.10 STATISTICAL METHODS

Descriptive statistics were used to describe the demographic characteristics of the study participants. We used the mean and standard deviation for continuous normally distributed

data. The probability-probability plot (P-P plot) was used to check for normality of data. For continuous data which was not normally distributed the median and interquartile range were used. We calculated proportions and percentages for categorical data. Continuous variables such as age, CD4+ cell counts, viral load counts, BMI and haemoglobin counts were categorised to examine possible non-linear associations and to group participants according to clinically acceptable categories.

2.10.1 Accuracy of the matching algorithm

A ten percent sample of patients in the mortuary register was manually matched to the TE dataset in Microsoft Excel. Variables used for matching were name, surname and date of birth, and matches had to agree on all three variables otherwise they would be classified as non-matches. Two-by-two tables were used to calculate sensitivity, specificity, positive predictive and negative predictive values of the probabilistic matching algorithm using the manual matching as the gold standard.

For records matched to NHLS only, we manually reviewed a ten percent sample checking whether name, surname and date of birth truly matched. The proportion of matches that were indeed true matches on all three variables was reported.

2.10.2 Comparison of causes of death by HIV status

Proportions of deaths in each category of cause of death were calculated for the overall study sample and then by HIV status. The numerator was count of deaths for each category on the CoDe classification and the denominator was the total number of eligible participants.

2.10.3 Comparison of causes of death by latest CD4 count category at time of death

We grouped the latest available CD4+ cell counts into the following categories: CD4+ cell count < 200 cells per ml, CD4+ cell count 200 - 500 cells per ml and CD4+ cell count > 500 cells per ml to examine the association between low or high CD4+ cell counts and causes of death, since CD4+ cell count has previously been used as an indicator of immune function (12). We calculated proportions of deaths per cause and ranked the causes according to frequency of deaths on a scale of one to ten.

2.10.4 Demographic and clinical factors associated with death from a respiratory illness: A sub-analysis of known HIV positive individuals on ART in TE

HIV-positive people appear to continue to have an increased frequency of respiratory illness compared with the general population (19). We performed a sub-analysis restricted to HIV positive individuals on ART in TE to examine factors associated with death from a respiratory illness. Individuals were included in the analysis if they had been successfully

matched between the mortuary register and TE and had an ART start date in TE. We created a binary outcome variable for whether a participant had died from a respiratory illness. Respiratory illnesses included lung embolus, TB, PCP, chronic obstructive pulmonary disease (COPD), lung cancer or bronchogenic malignancy.

2.10.5 The logistic regression model

We used logistic regression to identify demographic and clinical factors associated with deaths from respiratory illnesses. Logistic regression uses a logit function to find the best fitting, parsimonious yet biologically plausible model to describe the relationship between a binary outcome and explanatory variables.

Variable selection

Continuous variables were categorised to examine possible non-linear associations. We considered known risk factors of respiratory illness which were available in the data including Sex, age, viral load and TB treatment completion as *a priori* variables. We could not use smoking status or alcohol use variables although they were available in TE because there is no indication of the extent to which the individual smoked or used alcohol in terms of quantity, frequency and duration, all of which are necessary to evaluate the risk of respiratory illness. We analysed associations between individual explanatory variables and the outcome (death from a respiratory illness) using the Chi-square test of association. Variables that were associated with the outcome after univariable analysis (*p-value* < 0.2) and *a priori* variables were added to the multivariable/adjusted model. We used the highest frequency categories as reference groups.

The Pearson's goodness of fit test was conducted to measure the degree of fit of the data. A model to which the variables fit the data is expected to have a *p-value* greater than 0.05. All tests of hypothesis were two-tailed and statistically significant association was set at 95% confidence interval (*p-value* < 0.2 for the univariable analysis and *p-value* < 0.05 for the adjusted analysis).

2.10.6 Missing data

Missing data is unavoidable particularly when using retrospective data and has the potential to undermine the validity of research results (48). Complete case analysis was used in the data analysis to avoid biases resulting from missing data. In the descriptive analysis, a category within all categorical variables for those observations which had missing data was added. In the results of the regression model, we presented proportions of the total sample

with available data. The denominator was the total number of participants with data for the respective variable and the enumerator was the number of participants with the characteristic of interest (e.g. for the categorical variable viral load the proportion would be presented as follows: n with viral load <1000/total n with a viral load test result).

2.11 ETHICAL CONSIDERATIONS

This study was reviewed and approved by the University of the Witwatersrand Human Ethics Research Committee (Medical) under protocol number M180349 (Appendix 2). Written approvals to access the mortuary register, NHLS data and identified TE data were obtained from the HJH Ethics Committee, NHLS research committee and Right to Care, data gatekeeper of TE (Appendices 3-5) respectively. This was a cross sectional study of programmatic data and a waiver of informed consent was granted to review these records post-humously.

CHAPTER 3: RESULTS

This chapter presents tables and figures to describe the results of the descriptive and statistical analyses. The results follow the order of the study objectives stated in the Introduction chapter.

3.1 Demographic characteristics of the study sample by time period of death

The total sample size was 1780 and we excluded 14 individuals who were below the age of 18. The final analytic sample size was 1766 consisting of 910 (51.5%) males. The median age was 53.2 (IQR 39.7- 67.4) and most of the individuals were black (n = 1086, 61.5%). The three time periods were equally represented.

Equal numbers of participants were included for each of time period one, two and three (Table 6). The proportion of individuals who died from the emergency ward increased from 9.8% (n =56) in time Period 1 to 12.2% (n = 69) in Period 3. Deaths from the medical ward decreased from 78.6% (n = 445) in Period 2 to 71.7% (n = 406) in Period 3.

Table 6: Demographic characteristics of the study sample collected from the HJH mortuary register.

Characteristic	Level	Overall n = 1766 n, %	Period 1 (2014) n = 590	Period 2 (2015) n = 587	Period 3 (2016/2017) n = 589
Sex	Male	909 (51.5)	294 (49.8)	283 (48.2)	279 (47.4)
	Female	856 (48.4)	295 (50.0)	304 (51.8)	310 (52.6)
Age (years)	median (IQR)	53.2 (39.7- 67.4)	52.1 (39.8- 67.8)	54.2 (39.7-68.4)	53.0 (39.6- 66.1)
Age group	18-29	124 (7.0)	47 (8.1)	43 (7.5)	36 (6.2)
	30-49	635 (36.0)	220 (37.7)	198 (34.3)	217 (37.2)
	50+	1007 (57.0)	317 (54.3)	336 (58.2)	330 (56.6)
Race	Black	1057 (59.8)	369 (64.5)	352 (64.2)	368 (65.5)
	White	369 (20.9)	106 (18.7)	153 (27.9)	110 (19.6)
	Coloured	182 (10.3)	72 (12.7)	36 (6.4)	75 (13.4)
	Indian	40 (2.3)	23 (4.0)	8 (1.5)	9 (1.6)
	Missing	118 (6.7)	20 (1.1)	38 (6.5)	27 (4.6)
Ward	Emergency	156 (8.8)	56 (9.8)	31 (5.5)	69 (12.2)
	ICU	56 (3.7)	19 (3.3)	12 (2.1)	24 (4.1)
	Medical	1291 (73.1)	440 (74.6)	445 (78.6)	406 (71.7)
	Orthopaedic	12 (0.7)	4 (0.7)	5 (0.9)	3 (0.5)
	Psychiatry	3 (0.2)	0 (0.0)	1 (0.2)	2 (0.4)
	Surgical	168 (9.5)	50 (8.8)	70 (12.4)	48 (8.5)
	Theatre	8 (0.5)	4 (0.7)	2 (0.4)	2 (0.3)
	Missing	71 (4.0)	17 (2.9)	21 (3.6)	33 (5.6)

Note: IQR – interquartile range; ICU – intensive care unit

3.2 Causes of death by study period

Bacterial infections, AIDS infections, heart or vascular diseases, malignancies other than AIDS malignancies, stroke, renal failure, other causes, unclassifiable causes, accident or other violent death and unknown causes were the top ten causes of death in the whole study sample (Table 7). Bacterial infections and AIDS infections ranked first and second top causes of death consistently across all study periods. There was no difference in the total number of deaths by sex (Table S2) however accidents or other violent deaths were 70% higher among males than females. Most deaths occurred in the 50+ age group (Table S3) however accidents

or other violent deaths were five percent higher among the 18-29 year age group than the 50+ age group.

Table 7: Causes of death among individuals who died of any cause at HJH between May 2014 and July 2017.

Code	Overall n, %	Rank	Period 1 (2014) n = 590	Rank	Period 2 (2015) n = 587	Rank	Period 3 (2016/2017) n = 589	Rank
1. AIDS								
1.1 Infection	237 (13.4)	2	95 (16.1)	2	72 (12.2)	2	70 (11.9)	2
1.2 Malignancy	18 (1.0)		4 (0.7)		6 (1.0)		4 (0.7)	
2.Infection								
2.1 Bacterial	302 (17.1)	1	98 (16.6)	1	93 (15.8)	1	110 (18.7)	1
2.2 Others	55 (3.1)		20 (3.4)	10	15 (2.6)		18 (3.1)	
2.3 Unknown aetiology	59 (3.3)		14 (2.4)		21 (3.6)		25 (4.3)	
4.Malignancy (other than 1.2)	127 (7.2)	4	42 (7.1)	4	53 (9.0)	3	36 (6.1)	5
5.Diabetes mellitus	35 (2.0)		17 (2.9)		10 (1.7)		8 (1.4)	
8. MI or other ischemic heart disease	29 (1.6)		8 (1.4)		11 (1.9)		10 (1.7)	
9. Stroke	102 (5.8)	5	29 (4.9)	7	40 (6.8)	5	33 (5.6)	6
10. Gastro-intestinal haemorrhage	21 (1.2)		8 (1.4)		7 (1.2)		6 (1.0)	
13. Chronic obstructive lung disease	57 (3.2)		23 (3.9)	9	23 (3.9)	10	11 (1.9)	
14. Liver failure	42 (2.4)		15 (2.5)		9 (1.5)		18 (3.1)	
15. Renal failure	99 (5.6)	6	40 (6.8)	5	26 (4.4)	9	33 (5.6)	6
16. Accident or other violent death	80 (4.5)	9	15 (2.5)		28 (4.8)	8	37 (6.3)	4
20. Haematological disease	30 (1.7)		13 (2.2)		7 (1.2)		10 (1.7)	
24. Heart or vascular	141 (8.0)	3	51 (8.6)	3	40 (6.8)	5	50 (8.5)	3
25. Respiratory disease	22 (1.2)		7 (1.2)		10 (1.7)		5 (0.9)	
26. Digestive system disease	57 (3.2)		14 (2.4)		20 (3.4)		23 (3.9)	10
90. Other causes	91 (5.2)	7	24 (4.1)	8	41 (7.0)	4	26 (4.4)	9
91. Unclassifiable causes	90 (5.1)	8	30 (5.1)	6	30 (5.1)	7	32 (5.4)	8
92. Unknown	72 (4.1)	10	23 (3.9)		26 (4.4)		23 (3.9)	10

Note: Causes of death in each study period are ranked on a scale of one to ten according to frequency of deaths per cause. MI - myocardial infarction

Bacterial infections other than AIDS-related infections were the top cause of death in all study periods. Pneumonia (n=164, 62%) was the most common bacterial infection in the

bacterial infections other than 1.1 AIDS category (Fig. 2), followed by meningitis (n=43, 16%).

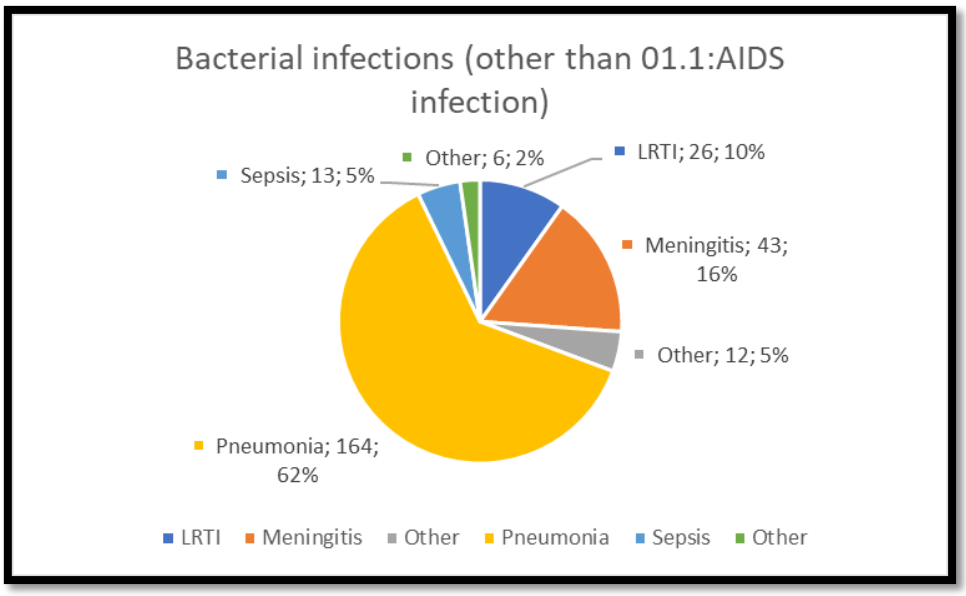


Figure 2: Distribution of bacterial infections other than AIDS-related infections in the study population.

Note: LRTI - lower respiratory tract infection

Of all AIDS related infections, TB was the most common at 73% (n = 172) followed by PCP at 16% (Fig. 3). TB-PCP co-infection accounted for only four percent of the AIDS-related infections whereas CCM accounted for seven percent.

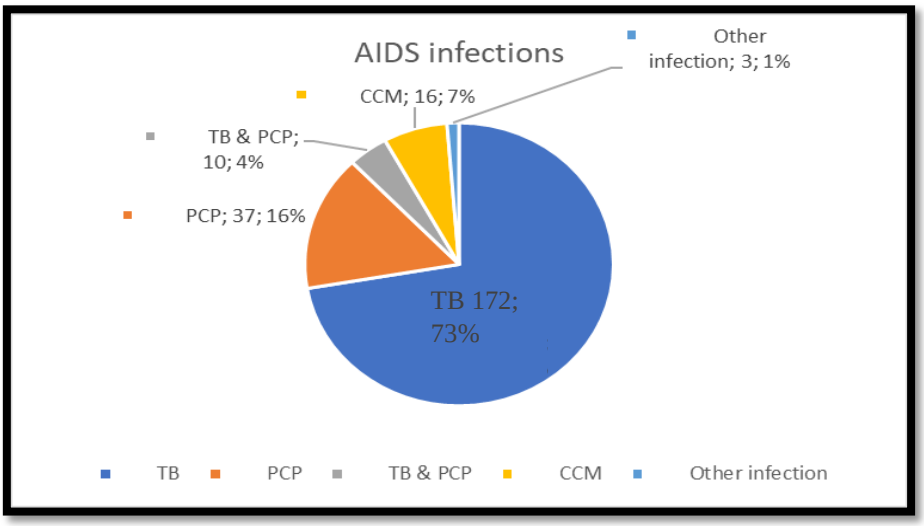


Figure 3: Proportions of AIDS infections in the study population.

Note: TB – tuberculosis; PCP - pneumocystis pneumonia; CCM - cryptococcal meningitis; AIDS – Acquired Immunodeficiency Syndrome

Overall, AIDS defining cancers or AIDS malignancies were not very common in the sample (n =18). However, 50% of all AIDS malignancies were due to Kaposi's sarcoma (Fig. 4) and 22% were due to cervical cancer.

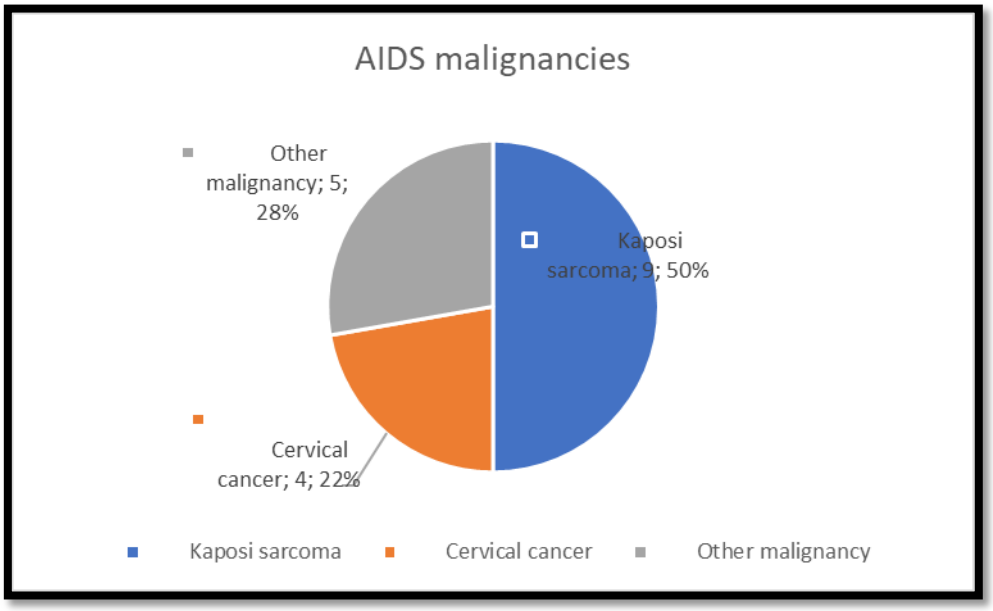


Figure 4: Proportions of AIDS malignancies in the study sample.

3.3 Feasibility and accuracy of matching mortuary register data to large HIV databases

Matching of the mortuary register patients to the NHLS database was done by the data analyst at the NHLS-CDW and a total of 686 matches were found. Matching of records between the mortuary register and TE database was done by the investigator using the -*reclink*- package in STATA Statistical Software: Release 14 and a total of 339 matches were found (Fig. 5). Missingness for sex and race dropped by 0.1% and 1.7% respectively after matching (Table S1). For the clinical variables which supplemented mortuary register data, reduction in missingness ranged from seven to twenty percent while HIV status was confirmed for 43% of the individuals.

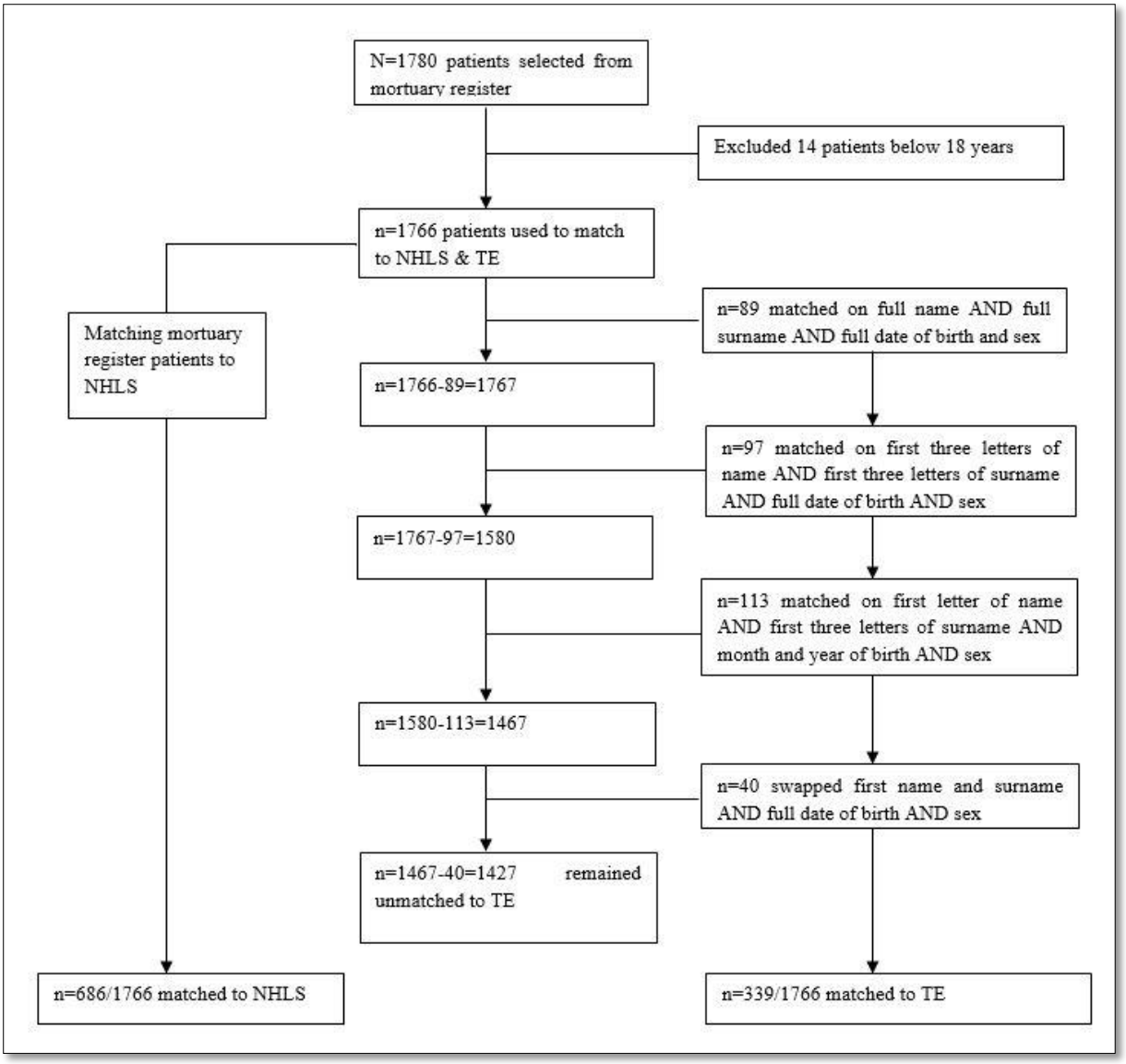


Figure 5: Record linkage process of mortuary register individuals to NHLS and TE records.

Note: NHLS – National Health Laboratory Service; TE - TherapyEdge-HIV™

The record linkage algorithm "reclink" in Stata had a sensitivity of 92.9% and specificity of 50%. Table 8 below shows the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the algorithm.

Table 8: Sensitivity and specificity of the matching algorithm.

		Algorithm output		
		Match	Non-match	Total
True match status (manually verified)	Match	26	3	29
	Non-match	2	3	5
	Total	27	6	33

Note: Sensitivity = 26/27 = 92.86%; Specificity = 3/6 = 50%; Positive predictive value (PPV) = 26/29 = 89.7% and Negative predictive value (NPV) = 3/5 = 60%

The NHLS matches were manually reviewed for accuracy (through reading through matched records to check if identifying variables truly matched), and the results were satisfactory. There was no significant difference in the sex of participants that were matched to TE, NHLS or those that were unmatched (Table 9). Unmatched individuals were older (median age 59.6 IQR: 42.2-72.6) compared to those matched to TE (median age 47.5 IQR: 38.7-64.1) and NHLS (median age 46.7 IQR: 38.4-59.5). Higher proportions of unmatched individuals had been admitted to the emergency and surgical wards before they died compared to those matched to TE and NHLS.

Table 9: Demographic characteristics of mortuary register individuals matched to NHLS, TE and those that were unmatched

Characteristic	NHLS n = 686	TE n = 339	Mortuary register Unmatched n = 762	p- value
Sex				0.50
Male	357 (52.0)	179 (52.8)	351 (46.1)	
Female	329 (48.0)	160 (47.2)	410 (53.9)	
Age (years) (median, IQR)	46.7 (38.4, 59.5)	47.5 (38.7, 64.1)	59.6 (42.2, 72.6)	< 0.001
Age group				< 0.001
18 - 29	47 (6.9)	21 (6.2)	52 (7.0)	
30 - 49	331 (48.3)	169 (49.9)	190 (25.7)	
50+	308 (44.5)	149 (44.0)	498 (67.3)	
Race				< 0.001
Black	533 (77.7)	271 (80.0)	343 (45.0)	
White	77 (11.2)	44 (13)	221 (29.0)	
Coloured	59 (8.6)	23 (6.8)	94 (12.3)	
Indian	9 (1.3)	1 (0.3)	23 (3.0)	
Missing	8 (1.2)	0	81 (10.6)	
Time period of death				0.31
Period 1 (2014)	243 (35.4)	125 (36.9)	226 (29.7)	
Period 2 (2015)	201 (29.3)	110 (32.5)	222 (29.2)	
Period 3 (2016/2017)	242 (35.3)	104 (30.7)	313 (41.1)	
Admission ward before death				<0.001
Emergency	25 (3.6)	22 (6.6)	98 (13.5)	
ICU	19 (32.8)	15 (4.5)	37 (5.1)	
Medical	579 (84.4)	268 (80.7)	495 (68.1)	
Orthopaedic	5 (0.7)	0	6 (0.8)	
Surgical	43 (6.3)	27 (8.1)	83 (11.4)	
Theatre	3 (0.4)	0	5 (0.7)	
Missing	12 (1.7)	0	59 (7.7)	

Note: NHLS – National Health Laboratory Service, TE – TherapyEdge^{TM}; IQR – interquartile range; ICU – intensive care unit

3.4 Demographic characteristics of the study sample by HIV status

In the study sample, 43.1% (n = 762) had an unknown HIV status, 42.8% (n = 755) were HIV positive and 14.1% (n = 249) were HIV negative (Table 10). Ten percent more males were HIV negative than females and the HIV negative participants were much older (61.3 years IQR 50.7-70.1). Medical admissions were most common across the strata than any other ward. HIV deaths were highest in period two (41.1%) and lowest in period three (26.0%).

Table 10: Demographic characteristics of the overall study sample and by HIV status.

			HIV status		
		Overall	**Positive**	**Negative**	**Unknown**
		n = 1766	**n = 755**	**n = 249**	**n = 762**
Characteristic	**Level**	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Sex	Male	910 (51.5)	361 (47.8)	138 (55.4)	410 (53.8)
	Female	856 (48.4)	394 (52.2)	110 (44.2)	351 (46.1)
Age (years)	median (IQR)	53.2 (39.7-67.4)	45.6 (37.2- 58.5)	61.3 (50.7-70.1)	59.6 (42.2-72.6)
Age group	18-29	124 (7.0)	60 (7.9)	14 (5.6)	50 (6.6)
	30-49	635 (36.0)	398 (52.7)	127 (51.0)	190 (24.9)
	50+	1007 (57.0)	297 (39.3)	108 (43.4)	498 (65.4)
Race	Black	1086 (61.5)	597 (79.1)	130 (52.2)	359 (47.1)
	White	369 (20.9)	87 (11.5)	61 (24.5)	221 (29.0)
	Coloured	182 (10.3)	41 (5.4)	47 (18.9)	94 (12.3)
	Indian	40 (2.3)	9 (1.2)	8 (3.2)	23 (3.0)
	Missing	89 (5.0)			
Time period of death	Period 1 (2014)	590 (33.4)	249 (33.0)	115 (46.2)	226 (29.7)
	Period 2 (2015)	587 (33.2)	310 (41.1)	55 (22.1)	222 (29.1)
	Period 3 (2016-17)	589 (33.2)	196 (26.0)	78 (31.7)	313 (41.1)
Ward	Emergency	156 (8.8)	47 (6.2)	11 (4.4)	98 (12.9)
	ICU	56 (3.7)	23 (3.0)	6 (2.4)	37 (4.9)
	Medical	1291 (73.1)	604 (80)	192 (77.1)	495 (65.0)
	Orthopaedic	12 (0.7)	3 (0.4)	3 (1.2)	6 (0.8)
	Psychiatry	3 (0.2)	0	0	3 (0.4)
	Surgical	168 (9.5)	57 (7.5)	28 (11.2)	83 (10.9)
	Theatre	9 (0.5)	2 (0.3)	2 (0.8)	5 (0.7)
	Missing	61 (3.5)			

Note: IQR – interquartile range; ICU – intensive care unit

3.5 Causes of death stratified by HIV status

AIDS infections (n = 162, 21.5%), bacterial infections (n = 149, 19.7%), malignancies (n = 48, 6.4%), heart or vascular illnesses (n = 46, 6.1%) and renal failure (n = 38, 5.0%) were the top five causes of death in the HIV positive sample. Among HIV negative individuals, bacterial infections were the leading cause of death (n = 43, 17.3%), followed by

malignancies (n = 31, 12.4%) and heart or vascular illnesses (n = 28, 11.2%). In the unknown HIV status group, bacterial infections were the leading cause of death (n = 110, 14.4%) followed by heart or vascular illnesses (n = 67, 8.8%) and AIDS (n = 58, 7.6%) as shown in Table 11 below.

Table 11: Causes of death in the study sample stratified by HIV status (positive, negative and unknown).

	Overall		**Positive**		**Negative**		**Unknown**	
	n = 1766		**n = 755**		**n = 249**		**n = 762**	
Code	**n (%)**	**Rank**	**n (%)**	**Rank**	**n (%)**	**Rank**	**n (%)**	**Rank**
1. AIDS								
1.1 Infection	237 (13.4)	2	162 (21.5)	1	17 (6.8)	4	58 (7.6)	3
1.2 Malignancy	18 (1.0)		14 (1.3)		0		4 (0.5)	
2. Infection								
2.1 Bacterial	302 (17.1)	1	149 (19.7)	2	43 (17.3)	1	110 (14.4)	1
2.2 Others	55 (3.1)		28 (1.6)		1 (0.4)		26 (3.4)	
2.3 Unknown aetiology	59 (3.3)		17 (1.0)		9 (3.6)		33 (4.3)	
4. Malignancy (other than 1.2)	127 (7.2)	4	48 (6.4)	3	31 (12.4)	2	48 (6.3)	6
5. Diabetes mellitus	35 (2.0)		7 (0.9)		8 (3.2)		20 (2.6)	
8. MI or other ischemic heart disease	29 (1.6)		9 (1.2)		4 (1.6)		16 (2.1)	
9. Stroke	102 (5.8)	5	31 (4.1)	9	16 (6.4)	6	55 (7.2)	4
10. Gastro-intestinal haemorrhage	21 (1.2)		6 (0.8)		3 (1.2)		12 (1.6)	
13. Chronic obstructive lung disease	57 (3.2)		9 (1.2)		15 (6.0)	7	33 (4.3)	
14. Liver failure	42 (2.4)		22 (2.9)		5 (2.0)		15 (2.0)	
15. Renal failure	99 (5.6)	6	38 (5.0)	5	17 (6.8)	4	44 (5.8)	7
16. Accident or other violent death	80 (4.5)	9	20 (2.6)		9 (3.6)	9	51 (6.7)	5
20. Haematological disease	30 (1.7)		21 (2.7)	10	1 (0.4)		8 (1.0)	
24. Heart or vascular	141 (8.0)	3	46 (6.1)	4	28 (11.2)	3	67 (8.8)	2
25. Respiratory disease	22 (1.2)		12 (1.6)		5 (2.0)		5 (0.7)	
26. Digestive system disease	57 (3.2)		15 (2.0)		8 (3.2)		34 (4.5)	10
90. Other causes	91 (5.2)	7	36 (4.8)	6	13 (5.2)	8	42 (5.5)	9
91. Unclassifiable causes	90 (5.1)	8	36 (4.8)	6	11 (4.4)	10	43 (5.6)	8
92. Unknown	72 (4.1)	10	33 (4.4)	8	5 (2.0)		34 (4.5)	10

Note: MI – myocardial infarction; HIV – Human Immunodeficiency Virus; AIDS – Acquired Immunodeficiency Virus. Codes with less than 20 observations were added to the 90. Other causes, and these included codes 3. Chronic viral hepatitis, 6. Pancreatitis, 7. Lactic acidosis, 11. Primary pulmonary hypertension, 12. Lung embolus, 17. Suicide, 18. Euthanasia, 19. Substance abuse (active), 21. Endocrine diseases, 22. Psychiatric diseases, 23. CNS disease, 27. Skin and motor system disease, 28. Urogenital disease, 29. Obstetric complications and 30. Congenital disorders

3.6 Causes of death by CD4+ cell category within 12 months before death among HIV positive individuals

AIDS infections were the top cause of death among individuals who had a CD4+ cell < 200 cells/ml (n = 64, 28.7%) followed by bacterial infections (n = 47, 21.1%), other infections (n = 12, 5.4%), malignancies (n = 12, 5.4%) and heart or vascular illnesses (n = 12, 5.4%).

Malignancies (n = 9, 6.8%), stroke (n = 8, 6.0%) and renal failure (n = 8, 6.0%) were the most common causes of death after bacterial infections (n = 30, 22.6%) and AIDS infections (n = 25, 18.8%) among individuals who had a CD4+ cell count between 200 and 500 cells/ml within 12 months before death. As expected, AIDS was the main cause of death among individuals with a low CD4 count (< 200 cells/ml) while heart or vascular illness, bacterial infection, malignancy and renal failure were equally ranked after AIDS among those with a high CD4 count (> 500 cells/ml).The ranking or ordering of cause of death for those who had a missing CD4+ cell result followed a similar trend to that of individuals who were in the 200-500 cells/ml category (Table 12).

Table 12: Causes of death by most recent CD4 count category +/- 12 months before death among individuals known to be HIV positive

Code	< 200 cells/ml n=223 n (%)	Rank	200 - 500 cells/ml n=133 n (%)	Rank	> 500 cells/ml n=52 n (%)	Rank	Missing n=347 n (%)	Rank
1.AIDS								
1.1 Infection	64 (28.7)	1	25 (18.8)	2	19 (36.5)	1	54 (15.6)	2
1.2 Malignancy	4 (1.8)		3 (2.3)		0		3 (0.9)	
2.Infection								
2.1 Bacterial	47 (21.1)	2	30 (22.6)	1	4 (7.7)	2	68 (20.0)	1
2.2 Others	12 (5.4)	3	5 (3.8)	10	1 (1.9)	10	10 (2.9)	
2.3 Unknown aetiology	3 (1.3)		3 (2.3)		2 (3.8)	8	9 (2.6)	
4. Malignancy (other than 1.2)	12 (5.4)	3	9 (6.8)	3	4 (7.7)	2	23 (6.6)	3
5. Diabetes mellitus	1 (0.4)		3 (2.3)		0		3 (0.9)	
8. MI or other ischemic heart disease	2 (0.8)		2 (1.5)		0		5 (1.4)	
9. Stroke	10 (4.5)	7	8 (6.0)	4	0		13 (3.7)	9
10. Gastro-intestinal haemorrhage	1 (0.4)		1 (0.8)		0		4 (1.2)	
13. Chronic obstructive lung disease	1 (0.4)		0		0		8 (2.3)	
14. Liver failure	4 (1.8)		7 (5.3)	6	1 (1.9)		10 (2.9)	
15. Renal failure	11 (4.9)	6	8 (6.0)	4	4 (7.7)	2	15 (4.3)	7
16. Accident or other violent death	2 (0.9)		4 (3.0)		0		14 (4.0)	8
20. Haematological disease	9 (4.0)	9	2 (1.5)		1 (1.9)	10	9 (2.6)	
24. Heart or vascular	12 (5.4)	3	6 (4.5)	8	4 (7.7)	2	20 (5.8)	5
25. Respiratory disease	6 (2.7)		0		3 (5.8)	6	3 (0.9)	
26. Digestive system disease	1 (0.4)		3 (2.3)		0		11 (3.2)	10
90. Other causes	8 (3.6)	10	6 (4.5)	8	2 (3.8)	8	21 (6.1)	4
91. Unclassifiable causes	10 (4.5)	7	7 (5.3)	6	3 (5.8)	6	16 (4.6)	6
92. Unknown	3 (1.4)		1 (0.8)		4 (7.7)		25 (7.2)	

Note: Code 90. Other causes included codes 3. Chronic viral hepatitis, 6. Pancreatitis, 7. Lactic acidosis, 11. Primary pulmonary hypertension, 12. Lung embolus, 17. Suicide, 18. Euthanasia, 19. Substance abuse (active), 21. Endocrine diseases, 22. Psychiatric diseases, 23. CNS disease, 27. Skin and motor system disease, 28. Urogenital disease, 29. Obstetric complications and 30. Congenital disorders. Myocardial infarction (MI)

3.7 Death from a respiratory illness: A sub-analysis of known HIV positive individuals on ART

We did a sub-analysis of deaths from a respiratory illness among known HIV positive individuals on ART because research shows an increased risk of respiratory illnesses in this population, possibly due to complications of ART. A total of 339 HIV positive individuals were known to be on ART in this sample, these are the individuals who were matched to TE and had reliable records of receiving ART. Fifty three percent of the sample was female (n = 179) and just over a third of the sample were in the 30 - 49 year age group (n = 169, 33.7%). Of those who had a viral load result within 12 months before death (n = 299), 55% were

virally suppressed (n = 163). For those who had TB data, 80.7% had completed TB treatment within 12 months before death (n = 92). Table 13 summarises the demographic and clinical characteristics of the known HIV positive individuals on ART. The majority of the sample had a CD4+ cell count below 200 cells/ml (n = 204/315, 64.7%).

3.7.1 Factors associated with death from a respiratory illness

In the adjusted model, those in the 50+ age group were less likely to die from a respiratory illness compared to those in the 30 - 49 year age group (aOR 0.36 95% CI 0.20, 0.65). In the unadjusted model, individuals who died between 2016/2017 (Period 3) were less likely to die from a respiratory illness (uOR 0.48 95% CI 0.25, 0.92) than those who died during an early period (Period 1, 2014) (Table 13). Individuals who recently completed TB treatment (within 12 months before death) were more likely to die from a respiratory illness (aOR 2.03 95% CI 1.07, 3.83) when compared to those who had completed TB treatment more than 12 months before death.

Table 13: Demographic and clinical factors associated with death from a respiratory illness among known HIV positive individuals on ART.

Characteristic		n = 339 n (%)	uOR (95% CI)	p-value	aOR (95% CI)
Sex	Male	46/160 (28.9)	1.00		1.00
	Female	37/179 (20.7)	0.65 (0.39, 1.06)	0.09	0.76 (0.43, 1.32)
Age group at death	20-29	4/21 (19.1)	0.46 (0.15, 1.44)	0.18	0.56 (0.17, 1.82)
	30-49	57/169 (33.7)	1.00		1.00
	50+	22/149 (14.8)	0.34 (0.19, 0.59)	< 0.01	0.36 (0.20, 0.65)
Race	Black	65/271 (24.0)	1.00		
	White	13/44 (29.6)	1.33(0.66, 2.69)	0.43	
	Coloured	5/23 (21.7)	0.88 (0.31, 2.46)	0.81	
	Indian	1/1 (100)	-		
Study period	Period 1 (2014)	36/125 (28.8)	1.00		1.00
	Period 2 (2015)	30/110 (27.3)	0.93 (0.52, 1.64)	0.80	0.86 (0.44, 1.66)
	Period 3 (2016/2017)	17/104 (16.4)	0.48 (0.25, 0.92)	0.03	0.72 (0.36, 1.48)
TB history	Completed TB treatment in the last 12 months	30/92 (32.6)	0.47 (0.28, 0.78)	< 0.01	2.03 (1.07, 3.83)
	Completed TB treatment more than 12 months before death	9/22 (40.9)	1.00		1.00
	No history of TB	44/225 (19.6)	0.28 (0.25, 1.33)	0.19	0.72, (0.54, 1.12)
CD4 count (most recent in+/- 12 months)	< 200 cells/ml	48/204 (23.5)	0.52 (0.22, 1.22)	0.13	0.51 (0.21,1.25)
	200 - 500 cells/ml	19/84 (22.6)	0.49 (0.19, 1.26)	0.14	0.47 (0.17, 1.24)
	≥ 500 cells/ml	10/27 (37.0)	1.00		1.00
Viral load (most recent +/- 12 months)	< 1000 copies/ml	38/163 (23.3)	1.00		1.00
	≥ 1000 copies/ml	32/136 (23.5)	1.00 (0.61, 1.66)	0.99	0.83 (0.48, 1.45)
	Missing	13/40 (32.5)	1.12 (1.00, 1.23)	0.87	1.05 (1.00, 1.36)
Regimen type at death	First line regimen	74/284 (26.1)	1.00		1.00
	Second line regimen	9/55 (16.4)	0.93 (0.45, 1.93)	0.85	0.98 (0.95, 2.05)
Time on ART category	< 6 months	15/39 (38.5)	1.00		1.00
	6 – 24 months	9/38 (23.7)	0.62 (0.31, 1.24)	0.17	0.36 (0.07, 1.73)
	> 24 months	59/207 (28.5)	0.59 (0.37, 0.92)	0.02	0.34 (0.12, 0.94)
BMI category (most recent +/- 12 months)	Underweight	21/78 (26.9)	1.00		
	Normal weight	26/121 (21.5)	0.74 (0.38, 1.44)	0.38	
	Overweight	10/55 (18.2)	0.60 (0.26, 1.41)	0.24	
	Obese	7/27 (25.9)	0.95 (0.35, 2.57)	0.92	
	Missing	40/136 (29.4)	0.99 (0.96, 1.46)	0.95	

Haemoglobin category (most recent +/- 12 months)	Normal	27/108 (25.0)	0.79 (0.36, 1.73)	0.56
	Mild & Moderate	13/44 (29.6)	1.00	
	Severe anaemia	4/23 (17.4)	0.50 (0.14, 1.77)	0.28
	Missing	39/164 (23.8)	0.72 (0.21, 1.95)	0.56

Note: adjusted for sex, age group, time period, completed TB treatment in the last 12 months, CD4+ cell category & viral load category +/- 12 months from death (*a priori* variables). aOR – adjusted odds ratio; uOR – unadjusted odds ratio; CI – confidence interval

CHAPTER 4: DISCUSSION

This chapter delves into the meaning, importance and relevance of the results of the study. It is focused on explaining and evaluating study findings in relation to the literature review and research questions. This chapter ends with an argument in support of the overall conclusion of the study.

4.1 Study findings

4.1.1 Summary of study findings

Bacterial infections, AIDS infections and heart or vascular diseases, were the top causes of death in the study sample. AIDS infections, bacterial infections, malignancies, heart or vascular illnesses and renal failure were the top five causes of death in the HIV positive sample. Among HIV negative individuals, bacterial infections were the leading cause of death, followed by malignancies and heart or vascular illnesses. In the unknown HIV status group, bacterial infections were the leading cause of death followed by heart or vascular illnesses and AIDS. AIDS was the main cause of death among individuals with a low CD4 count (< 200 cells/ml) while heart or vascular illness, bacterial infection, malignancy and renal failure were equally ranked after AIDS among those with a high CD4 count (> 500 cells/ml). Individuals who had completed TB treatment within 12 months before death were more likely to die from a respiratory illness when compared to those who had completed TB treatment more than 12 months before death.

4.1.2 Distribution of causes of death

In terms of sex and age distribution, our sample is representative of the adult population of SA. Approximately 50% of the study sample was female which is in line with StatsSA mid-year estimates showing that 51.2% of the SA population is female (49). The percentage of black Africans in this study is lower (61%) than the national proportion (81%) (49). We found a 43% HIV prevalence in this sample which is much lower than the 58% HIV prevalence reported in a recent study of mortality at an urban hospital in Gauteng (14). The same study reported that infections and parasites, respiratory and circulatory conditions were the top causes of death which is similar to the top causes of death in our overall sample (14). The ranks of causes of death reported in our present study differ with ranks at the national level (Table 7 vs Table S4). Diabetes mellitus does not feature in the top ten causes of death across all of our study periods whereas according to StatsSA national estimates it was in the top three in 2014, 2015 and 2016. Use of different cause of death classification systems and data sources may account for these differences. StatsSA mortality reports are generated using data from death notification forms with ICD-10 coded causes of death.

After infections and AIDS, heart or vascular conditions and malignancies came in as the third and fourth most common causes of death in the sample. National mortality statistics show that TB is the leading cause of death, which in this study made a large proportion of the infections (1). National data is also indicating the rise of non-communicable diseases such as CVD, diabetes and cancers (1).

When compared to global estimates for top causes of death, results of this study reflect some similarities with results of both low and high income countries. UN reports indicate that diarrhoeal diseases, HIV/AIDS and TB were among the top causes of death in low income countries in the year 2016 (13). In the high income countries, non-communicable diseases like ischaemic heart disease, stroke, Alzheimer's disease, colon and rectum cancers occupy positions in the top ten causes of death (13). In this study, AIDS and infections ranked highest with non-communicable diseases such as malignancies and vascular conditions ranking lower, but still in the top ten.

4.1.3 Causes of death by HIV status

Deaths attributed to AIDS accounted for 14% of all deaths and 23% of deaths among HIV positive individuals. A recent study in a medical ward at a large public sector hospital in Johannesburg similar to HJH reported an HIV-attributable fraction of 36% in the whole sample and 60% in those that were HIV-positive (50). StatsSA estimates the number of deaths attributable to AIDS in 2017 as 25% of all deaths in SA. Deaths from AIDS infections seemingly decreased from 16.1% in time period one to 11.9% in time period three (Table 7). It is apparent that AIDS deaths are underestimated in this study primarily because of the underreporting of AIDS as a cause of death in the mortuary register. The documented reluctance to report AIDS as a cause of death by medical practitioners because of concerns about the privacy of the death certificate during study time periods (10, 51) and previously life insurance policies included exclusion clauses for HIV (51) are factors that contributed to the underreporting of AIDS deaths. Studies on the quality of death notification in SA have reported up to 93% of AIDS death misattributed to other causes (32,52, 53). In this study, AIDS deaths were mostly misreported as TB, PCP, cryptococcal meningitis and retroviral disease (RVD) as these are the conditions which were common among the HIV positive individuals. We hypothesized that there would be a decrease in AIDS deaths over the study time periods due to the availability of ART. Our results reflect a decrease in AIDS deaths as we expected but there was misclassification of AIDS deaths as bacterial infection (a

proportion of individuals who had an AIDS death were HIV negative or had and unknown HIV status).

Accident or other violent deaths were much higher in the unknown HIV status group (6.7%) compared to the HIV positive (2.6%) or HIV negative (3.6%) groups. Additionally, accident or other violent deaths were five times higher in males than in females (Table S2). Current mortality trends show increasing rates of road traffic accidents and suicides among young adult males (1). StatsSA reports that assault is the second most common unnatural cause of death and young males are more likely to be victims of assault (1).

4.1.4 Feasibility of matching and accuracy of the probabilistic algorithm

The performance of our probabilistic record linkage was satisfactory in terms of the probability that a true match was correctly identified by our algorithm (sensitivity 92%) and the probability that a match identified by our algorithm was truly a match (PPV 90%). These values are similar to those reported in a study where national HIV cohort was created from NHLS data using similar probabilistic linkage methods (42). Only 57% of the sample (from the mortuary register) was matched to at least one of the electronic databases NHLS and TE. This means that HIV prevalence could be underestimated because HIV status is ascertained from NHLS and an electronic ART register from a single site. Over half of whites in this study (65%) were unmatched because the prevalence of HIV in this group was very low (Table 10). In addition, the majority of the white population use private sector facilities therefore would not have records in the NHLS which caters for public sector facilities (63).

In a cross-sectional study to assess the accuracy and completeness of SA national laboratory CD4+ cell and viral load data, 90% of adults on ART who visited a hospital-based HIV clinic in Durban were matched to NHLS records (55). The huge disparity in proportions of records matched is because they matched individuals receiving ART at the facility, who would have had routine CD4+ cell count and viral load counts done at the NHLS and recorded under uniform identifiers i.e. name, surname and sex since all individuals where from the same facility (55).

Typographical errors and patient mobility between facilities may result in patient specimens being associated with different sets of identifying data, which impacted the results of our matching exercise (42). We used handwritten and electronic data sources and it is known that where records are hand-written, implementation and maintenance of electronic records is not impervious to error (56). Incomplete demographic fields, variability in names arising from

hearing errors, use of nicknames, translations (e.g. Mpho vs Gift), first name/surname inversions, use of middle names and abbreviations also add to complications in record linkage (42). Variation in dates of birth can arise from typographical and hearing errors, from month-day inversions, from false reports or misremembering, or through provision of an age rather than an exact date of birth. Sex may be listed incorrectly due to typographical error, due to the large proportion of androgynous names in this setting and may also change if some individuals transition to a different sex (42).

4.1.5 Factors associated with death from a respiratory illness

The 339 individuals matched to TE and subsequently used in the sub-analysis of factors associated with death from a respiratory illness are not a fair representation of the TLC patient population. This subset of individuals is those who had fallen very ill to the point that warranted hospital admission. Since 2009, TLC has been 'down-referring' stable patients (i.e. those on ART for at least 11 months, with an undetectable viral load in the previous 10 months, stable weight, a CD4 count > 200 cells per ml, < 5% weight loss over the last three visits and no opportunistic infections) to Crosby and Rex clinics for monitoring and treatment (57). Patients who deteriorate and fail to respond to treatment at the down-referral site are 'up-referred' back to TLC for care and continued treatment until they are eligible to be down-referred again (57). In addition, HJH is a public healthcare facility in a region with many other well-equipped public healthcare facilities which ill patients may opt for. Many HIV patients receive ART in public sector facilities but also receive treatment for other conditions in the private sector facilities (58). Individuals who died outside of hospital admission (mild to moderate illness) and those who were admitted in private sector facilities at death are not represented in this study.

Indeed, ART is known to reduce the risk of TB infection by 65% across all CD4 strata (59) and already ART has been made easily accessible through the public health sector but there is still incidence of TB among HIV positive individuals on ART (60). Individuals who had completed TB treatment within 12 months before death had two-fold higher odds of dying from a respiratory illness (aOR 2.03 95% CI 1.07, 3.83) than individuals who had completed TB treatment in more than 12 months before death. These estimates show that HIV positive people are at higher risk of death within 12 months of completing TB treatment. These findings support the importance of monitoring HIV positive TB patients receiving TB treatment and also close follow-up of persons with prior TB episodes (64). Surveillance resources should be devoted to these individuals because, despite successful treatment, they

have a high risk of recurrent disease. This is because people who have already experienced a TB episode may still live in a high risk environment (crowded living space, place of work, school, or transport situation) where other infectious cases continue to be present. Lung damage or immunological deficiency caused by the previous episode, smoking and drug abuse may also contribute to increased risk of re-infection (64).

4.2 Limitations

In interpreting the results of this study, certain limitations must be taken into account. Underestimation of AIDS infections in this study is likely because of two reasons. Firstly, HIV status is based on linkage to NHLS and TE for a single ART site of which the matching to clinical data sources did not yield perfect results due to lack of a unique identifier common in all three data sources. We tried to mitigate this by using multiple pieces of identifying data (i.e. name, surname and date of birth) and examining possible combinations of date of birth that could have arisen from typographical errors. We developed a simple algorithm using conventional patient identifiers which may be useful in record linkage in similar settings, where patient data lacks a unique identifier such as national ID number due to confidentiality issues.

Secondly, there was no record of whether an infection was recurrent or not in the mortuary register, which is a key characteristic in the case definition of most AIDS-related infections. We relied on expert knowledge to classify conditions as accurately as possible, taking into account all causes of death recorded for each individual and arriving to a medically plausible final cause of death.

Drug resistant TB is an important current topic however it is beyond the scope of this report. While we knew which patients had died from TB, we had no further information indicating the nature of the TB infection.

Missing data are common in retrospective studies such as the present study, wherein routinely collected data are subsequently used for a different purpose. We used complete case analysis (i.e. exclusion of those participants who have missing data in the variables of interest from the analysis). While complete case analysis reduces sample size and subsequently reduced statistical power, it is unbiased if the data are missing at random as was the case in this study (62).

Strengths of this study include a large sample size and availability of supplementary data sources. In addition, we propose a novel method for confirming the HIV status of deceased

individuals in mortuary registers using routine HIV monitoring data from a facility-level longitudinal database and a national-level data source.

4.3 Conclusion and Recommendations

Infectious diseases such as pneumonia and TB remain the major causes of mortality among HIV positive individuals in the post-ART era. This indicates that the current TB treatment programs need to be strengthened, with a particular focus on prevention among the HIV positive population, to address mortality associated with HIV-TB co-infection. Prevention programs for opportunistic infections among HIV positive individuals should be expanded. Efforts must be made to improve the cause of death reporting by medical practitioners so that good quality mortality data is available for use to reliably inform health policy.

CHAPTER 5: REFERENCES

1. Statistics South Africa . Mortality and causes of death in South Africa, 2016. 2016. accessed 13 June 2018 http://www.statssa.gov.za/publications/P03093/P030932016.pdf.
2. Collini P, Morris A. Maintaining lung health with longstanding HIV. Current Opinion in Infectious Diseases. 2016;29 (1):31-8.
3. Varkila MRJ, Vos AG, Barth RE, Tempelman HA, Deville WLJ, Coutinho RA, et al. The association between HIV infection and pulmonary function in a rural African population. PLoS One. 2019;14 (1):e0210573.
4. 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 (London, England). 2012;380 (9859):2095-128.
5. World Health Organisation. Global status report on noncommunicable diseases 2010 Accessed 12 March 2019 https://www.who.int › chp › ncd_global_status_report.
6. 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;386 (10001):1386-94.
7. 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;386 (10001):1407-18.
8. 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;386 (10001):1395-406.
9. 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;386 (10001):1373-85.
10. 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.
11. Hagger-Johnson G, Harron K, Goldstein H, Aldridge R, Gilbert R. Probabilistic linkage to enhance deterministic algorithms and reduce data linkage errors in hospital administrative data. Journal of Innovation in Health Informatics. 2017; 24 (2):891.
12. Carmona S, Bor J, Nattey C, Maughan-Brown B, Maskew M, Fox MP, et al. Persistent High Burden of Advanced HIV Disease Among Patients Seeking Care in South Africa's

National HIV Program: Data From a Nationwide Laboratory Cohort. Clinical Infectious Diseases. 2018; 66 (suppl_2):S111-s7.

13. United Nations. World Mortality Report. 2017. accessed 21 December 2019 https://www.un.org/en/development/desa/population/publications/pdf/mortality/World-Mortality-2017-Data-Booklet.pdf
14. L C Long, D H Evans, S Rosen, C Sauls, A T Brennan, I M Sanne, et al. Can routine inpatient mortality data improve HIV mortality estimates? Inpatient mortality at an urban hospital in South Africa. South African Medical Journal. 2018. 108(10):870-875.
15. Croxford S, Kitching A, Desai S, Kall M, Edelstein M, Skingsley A, et al. Mortality and causes of death in people diagnosed with HIV in the era of highly active antiretroviral therapy compared with the general population: an analysis of a national observational cohort. The Lancet Public Health. 2017;2 (1):e35-e46.
16. Ryom L, Lundgren JD, El-Sadr W, Reiss P, Kirk O, Law M, et al. Cardiovascular disease and use of contemporary protease inhibitors: the D:A:D international prospective multicohort study. The Lancet HIV. 2018;5 (6):e291-e300.
17. Lundgren J, Mocroft A, Ryom L. Contemporary protease inhibitors and cardiovascular risk. Current Opinion in Infectious Diseases. 2018;31 (1):8-13.
18. Chawla A, Wang C, Patton C, Murray M, Punekar Y, de Ruiter A, et al. A Review of Long-Term Toxicity of Antiretroviral Treatment Regimens and Implications for an Aging Population. Infectious Diseases and Therapy. 2018;7 (2):183-95.
19. Drummond MB, Kirk GD. HIV-associated obstructive lung diseases: insights and implications for the clinician. The Lancet Respiratory Medicine. 2014;2(7):583-92.
20. Leung JM, Liu JC, Mtambo A, Ngan D, Nashta N, Guillemi S, et al. The determinants of poor respiratory health status in adults living with human immunodeficiency virus infection. AIDS Patient Care and STDs. 2014;28 (5):240-7.
21. Crothers K, Thompson BW, Burkhardt K, Morris A, Flores SC, Diaz PT, et al. HIV-associated lung infections and complications in the era of combination antiretroviral therapy. Proceedings of the American Thoracic Society. 2011;8(3):275-81.
22. Crothers K, Morris A. HIV infection and lung function decline: challenges, clinical implications, and new questions. AIDS (London, England). 2013;27(8):1345-7.
23. Abdool Karim SS, Churchyard GJ, Karim QA, Lawn SD. HIV infection and tuberculosis in South Africa: an urgent need to escalate the public health response. Lancet (London, England). 2009;374 (9693):921-33.

24. Scott L, da Silva P, Boehme CC, Stevens W, Gilpin CM. Diagnosis of opportunistic infections: HIV co-infections - tuberculosis. Current Opinion in HIV and AIDS. 2017;12 (2):129-38.
25. Sayers A, Ben-Shlomo Y, Blom AW, Steele F. Probabilistic record linkage. International Journal of Epidemiology. 2016;45 (3):954-64.
26. Carreras G, Simonetti M, Cricelli C, Lapi F. Deterministic and Probabilistic Record Linkage: an Application to Primary Care Data. Journal of medical systems. 2018;42 (5):82.
27. Cook LJ, Olson LM, Dean JM. Probabilistic record linkage: relationships between file sizes, identifiers and match weights. Methods of Information in Medicine. 2001;40 (3):196-203.
28. Alves DN, Bresani-Salvi CC, Batista JDL, Ximenes RAA, Miranda-Filho DB, Melo HRL, et al. Use of the Coding Causes of Death in HIV in the classification of deaths in Northeastern Brazil. Rev Saude Publica. 2017;51:88.
29. Copenhagen HIV Progamme. Coding Causes of Death in HIV Protocol. 2015. accessed 23 February 2019 https://www.chip.dk/Portals/0/files/Code%20Protocol%202.3.pdf
30. Kolwalska JD. The Coding Causes of death in HIV (CoDe) Project: initial results and evaluation of methodology. Epidemiology. 2011;22 (4):516-23.
31. Kavi Bhalla, James E Harrison, Saeid Shahraz, Lois A Fingerhut. Availability and quality of cause-of-death data for estimating the global burden of injuries. Bulletin of the World Health Organization. 2010 (88):831-8.
32. Nojilana B, Groenewald P, Bradshaw D, Reagon G. Quality of cause of death certification at an academic hospital in Cape Town, South Africa. South African Medical Journal. 2009;99 (9):648-52.
33. Johnson LF, May MT, Dorrington RE, Cornell M, Boulle A, Egger M, et al. Estimating the impact of antiretroviral treatment on adult mortality trends in South Africa: A mathematical modelling study. PLoS Medicine. 2017;14 (12):e1002468.
34. Statistics South Africa. Census 2011: Provincial profiles. 2010. accessed 25 May 2019 https://www.statssa.gov.za/publications/P03014/P030142011.pdf.
35. Fox MP, Maskew M, MacPhail AP, Long L, Brennan AT, Westreich D, et al. Cohort profile: the Themba Lethu Clinical Cohort, Johannesburg, South Africa. International Journal of Epidemiology. 2013;42 (2):430-9.
36. Fox MP, Maskew M, Brennan AT, Evans D, Onoya D, Malete G, et al. Cohort profile: the Right to Care Clinical HIV Cohort, South Africa. BMJ Open. 2017;7 (6):e015620.

37. National Department of Health Republic of South Africa. Clinical Guidelines for the Management of HIV & AIDS in Adults and Adolescents. 2010. accessed 14 March 2018 https://www.who.int/hiv/pub/guidelines/south_africa_art.pdf

38. National Department of Health Republic of South Africa. Circular on New Criteria for Initiating Adults on ART at CD4 Count of 350 cells/mL and Below. 2011. accessed 14 March 2018 https://foundation.co.za/document/publications/Published%20Articles/2013/Special%20Programmes%20Technical%20Assistance/Kinkel_BMCHSR_PALC_28_02_2014_final.pdf

39. National Department of Health Republic of South Africa. The South African Antiretroviral Treatment Guidelines. 2013. accessed 14 March 2018 https://sahivsoc.org/Files/2013%20ART%20Treatment%20Guidelines%20Final%2025%20March%202013%20corrected.pdf

40. National Department of Health, Republic of South Africa. National Consolidated Guidelines – For the Prevention of Mother-to-Child Transmission of HIV (PMTCT) and the Management of HIV in Children, Adolescents and Adults. 2015. accessed 15 July 2018 https://aidsfree.usaid.gov/sites/default/files/tx_south-africa_pmtct_2015.pdf.

41. Fox MP, Bor J, Brennan AT, MacLeod WB, Maskew M, Stevens WS, et al. Estimating retention in HIV care accounting for patient transfers: A national laboratory cohort study in South Africa. PLoS Medicine. 2018;15(6):e1002589.

42. Jacob Bor, William MacLeod, Katia Oleinik, James Potter, Alana T. Brennan, Sue Candy, et al. Building a National HIV Cohort from Routine Laboratory Data: Probabilistic Record-Linkage with Graphs. 2018. BioXirv. DOI:10.1101/450304.

43. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research Electronic Data Capture (REDCap) - A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009;42 (2):377-81.

44. Rizzi S, Halekoh U, Thinggaard M, Engholm G, Christensen N, Johannesen TB, et al. How to estimate mortality trends from grouped vital statistics. International Journal of Epidemiology. 2018;48 (2) 571-582.

45. Schoeman WJ. South African religious demography: The 2013 General Household Survey. HTS Teologiese Studies/Theological Studies. 2017;73 (2).

46. Alves D. Use of the Coding Causes of Death in HIV in the classification of deaths in North Eastern Brazil. Rev Saude Publica. 2015;51 (88).

47. Schneider E, Whitmore S, Glynn K, Dominguez K, Mitsch A, McKenna M. Revised surveillance case definitions for HIV infection among adults, adolescents, and children aged <18 months and for HIV infection and AIDS among children aged 18 months to <13 years. Morbidity and Mortality World Report Recommended Report. 2008;57:1-12.

48. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.

49. Statistics South Africa. Mid-year population estimates. 2019. accessed 23 August 2019 http://www.statssa.gov.za/publications/P0302/P03022019.pdf

50. Black A, Sitas F, Chibrawara T, Gill Z, Kubanje M, Williams B. HIV-attributable causes of death in the medical ward at the Chris Hani Baragwanath Hospital, South Africa. Plos One. 2019;14 (5).

51. Burger EH, Groenewald P, Rossouw A, Bradshaw D. Medical certification of death in South Africa--moving forward. South African Medical Journal 2015;105 (1):27-30.

52. Groenewald P, Nannan N, Bourne D, Laubscher R, Bradshaw D. Identifying deaths from AIDS in South Africa. AIDS (London, England). 2005;19 (2):193-201.

53. Meel BL. Certification of deaths at Umtata General Hospital, South Africa. Journal of Clinical Forensic Medicine. 2003;10 (1):13-5.

54. Ramirez-Avila L, Nixon K, Noubary F, Giddy J, Losina E, Walensky RP, et al. Routine HIV testing in adolescents and young adults presenting to an outpatient clinic in Durban, South Africa. PLoS One. 2012;7 (9):e45507.

55. Bassett IV, Huang M, Cloete C, Candy S, Giddy J, Frank SC, et al. Assessing the completeness and accuracy of South African National Laboratory CD4 and viral load data: a cross-sectional study. BMJ Open. 2018;8 (8):e021506.

56. Maskew M, Bor J, Hendrickson C, MacLeod W, Barnighausen T, Pillay D, et al. Imputing HIV treatment start dates from routine laboratory data in South Africa: a validation study. BMC Health Services Research. 2017;17 (1):41.

57. Brennan AT, Long L, Maskew M, Sanne I, Jaffray I, MacPhail P, et al. Outcomes of stable HIV-positive patients down-referred from a doctor-managed antiretroviral therapy clinic to a nurse-managed primary health clinic for monitoring and treatment. AIDS (London, England). 2011;25 (16):2027-36.

58. Long L, Brennan A, Fox MP, Ndibongo B, Jaffray I, Sanne I, et al. Treatment outcomes and cost-effectiveness of shifting management of stable ART patients to nurses in South Africa: an observational cohort. PLoS Medicine. 2011;8 (7):e1001055.
59. Suthar AB, Lawn SD, del Amo J, Getahun H, Dye C, Sculier D, et al. Antiretroviral therapy for prevention of tuberculosis in adults with HIV: a systematic review and meta-analysis. PLoS medicine. 2012;9 (7):e1001270.
60. Churchyard GJ, Mametja LD, Mvusi L, Ndjeka N, Hesseling AC, Reid A, et al. Tuberculosis control in South Africa: successes, challenges and recommendations. South African Medical Journal. 2014;104 (3 Suppl 1):244-8.
61. Kufa T, Mngomezulu V, Charalambous S, Hanifa Y, Fielding K, Grant AD, et al. Undiagnosed tuberculosis among HIV clinic attendees: association with antiretroviral therapy and implications for intensified case finding, isoniazid preventive therapy, and infection control. Journal of Acquired Immune Deficiency Syndromes. 2012;60 (2):e22-8.
62. Mukaka M, White SA, Terlouw DJ, Mwapasa V, Kalilani-Phiri L, Faragher EB. Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing? Trials. 2016;17:341.
63. Maphumulo, T & Bhengu, R. Challenges of quality improvement in healthcare of South Africa post-apartheid. A critical review. Curationis 2019; 42 (1) a1901
64. Osman M, Welte A, Dunbar R, Brown R, Hoddinott G, Hesseling A, et al. Morbidity and mortality up to 5 years post tuberculosis treatment in South Africa: A pilot study. International Journal of Infectious Diseases. 2019; 85:57-63

APPENDICES

This section is a collection of all the documents pertaining to the administrative processes concerned with this project. Approvals from various bodies were required and the documentation is presented here.

Appendix 1: Plagiarism declaration form



PLAGIARISM DECLARATION TO BE SIGNED BY ALL HIGHER DEGREE STUDENTS

SENATE PLAGIARISM POLICY: APPENDIX ONE

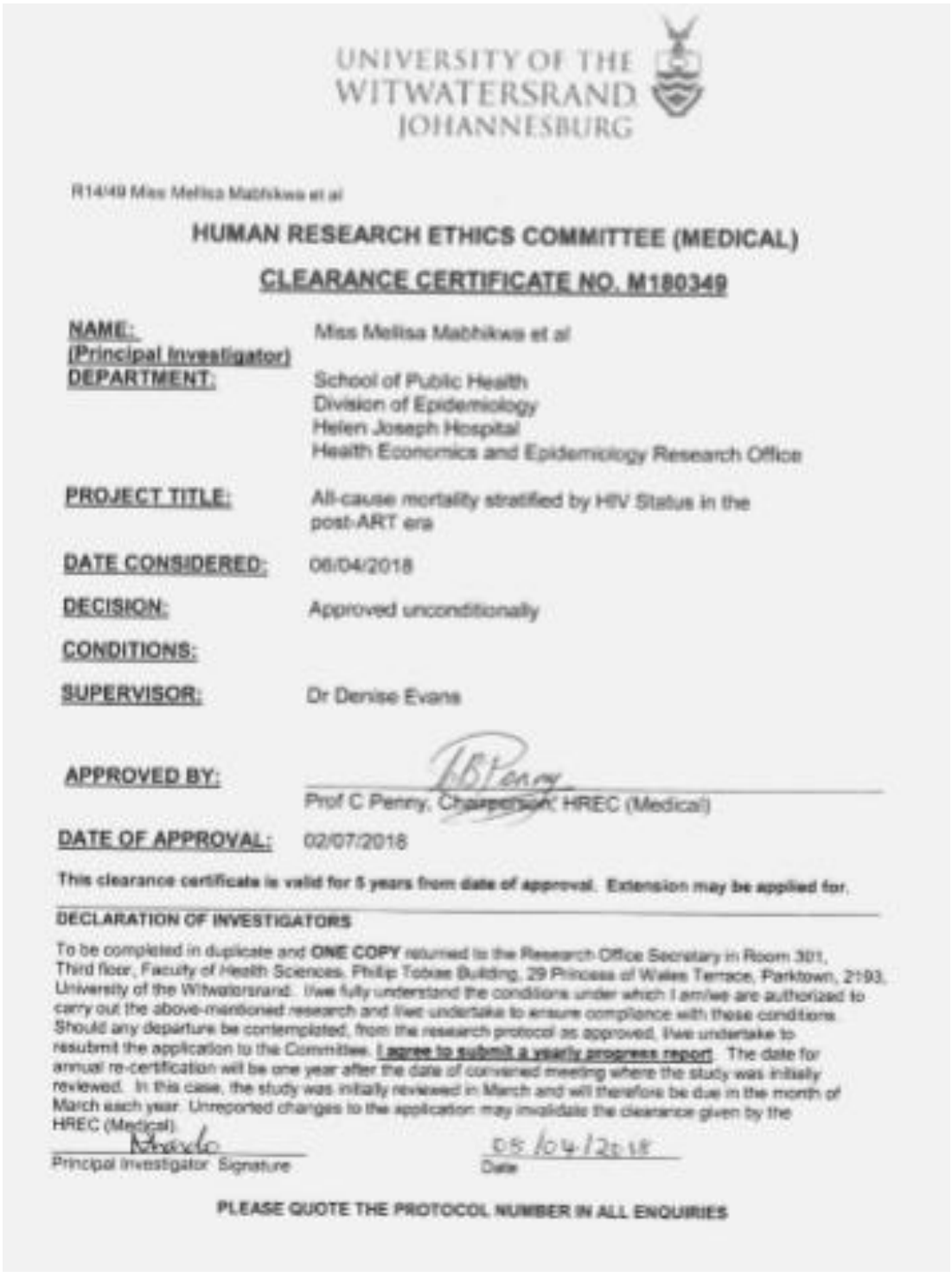
I Mellisa Nothando Mabhikwa (Student number: 708821) am a student registered for the degree of MSc Epidemiology in the academic year 2020.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signature: Nthando Date: 25 February 2020

Appendix 2: Ethics clearance certificate



UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

R14/49 Miss Mellisa Mabhikwa et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M180349

NAME: (Principal Investigator)	Miss Mellisa Mabhikwa et al
DEPARTMENT:	School of Public Health Division of Epidemiology Helen Joseph Hospital Health Economics and Epidemiology Research Office
PROJECT TITLE:	All-cause mortality stratified by HIV Status in the post-ART era
DATE CONSIDERED:	06/04/2018
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Dr Denise Evans
APPROVED BY:	Prof C Penny, Chairperson HREC (Medical)
DATE OF APPROVAL:	02/07/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. I agree to submit a yearly progress report. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in March and will therefore be due in the month of March each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

05/04/2018
Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix 3: NHLS data access approval

NATIONAL HEALTH LABORATORY SERVICE

Modderfontein Road, Sandringham, 2031
Tel: +27 (0)11 386 6142
Fax: +27 (0)11 386 6296
Email: babatyi.kgokong@nhls.ac.za
Web: www.nhls.ac.za

19 June 2018

Applicant: Ms Melissa Mabhikwa
Institution: University of the Witwatersrand
Division: Epidemiology
Department: Public Health
Email: 708821@students.wits.ac.za
Cell: 061 203 9014

Re: Provisional Approval to access National Health Laboratory Service (NHLS) Data

Your application to undertake a research project '**All-cause Mortality Stratified by HIV Status in the Post-ART Era**' using data from the NHLS database has been reviewed. This letter serves to advise that the application has been provisionally approved.

Please note that final approval will be granted on your compliance with the NHLS conditions of service and that the study can only be undertaken provided that the following conditions have been met.

- Ethics approval is obtained from a recognised SA Health Research Ethics Committee.
- Processes are discussed with the relevant NHLS departments (i.e. Information Management Unit and Operations Office) and are agreed upon.
- Confidentiality is maintained at participant and institutional level and there is no disclosure of personal information or confidential information as described by the NHLS policy.
- A final report of the research study and any published paper resulting from this study are submitted and addressed to the NHLS Academic Affairs and Research office and the NHLS has been acknowledged appropriately.
- NHLS Data cannot be used to track patients as no pre-approval/consent is obtained from Patients.

Please note that this letter constitutes provisional approval by the NHLS Academic Affairs and Research Office. Once Ethics approval has been granted please send it to sipokazi.thompson@nhls.ac.za. Any data related queries may be directed to NHLS Corporate Data Warehouse, contact number: 011 386 6074 email: zarina.sabat@nhls.ac.za

pp VUSI SIHOA
Dr Babatyi Malope-Kgokong
National Manager: Academic Affairs and Research

Appendix 4: HJH ethics approval

GAUTENG PROVINCE
HEALTH
REPUBLIC OF SOUTH AFRICA

Gauteng Department of Health
Helen Joseph Hospital
Enquiries: Dr. M. Mukansi
Research Committee: Chairperson
Tel :(011) 489-0306/1087
Fax :(011) 489 1038
E mail:Murimisi.mukansi@wits.ac.za

24 May 2018

To whom it may concern

Subject: HELEN JOSEPH HOSPITAL RESEARCH COMMITTEE APPLICATION

PROTOCOL TITLE: All – Cause mortality stratified by HIV status in the post –Art Era
Protocol Ref No: Mellissa Mabhikwa

Ethic Clearance: Pending
Principal investigator: Mellissa Mabhikwa

Department: Helen Joseph hospital

Committee Recommendations
The Committee is giving you Conditional access that while awaiting the final ethical clearance certificate from the university of Witwatersrand HREC.
It is the duty of the researcher to collect the data to the relevant department after the Research Committee approved the study.

Dr. M. Mukansi
Chairperson of HJH Ethic and Research Committee

Appendix 5: Data sharing agreement for TE data access through HE^{2}RO

M. N. M

Health Economics and Epidemiology Research Office

Unit 2, 39 Empire Road
Parktown, Johannesburg, 2193, South Africa
Tel +27 (0)10 0017930

A division of the Wits Health Consortium (Pty) Ltd, a wholly owned subsidiary of the University of the Witwatersrand.

Data Sharing Agreement:
Process to be followed for granting access to analytic datasets created and maintained by the Health Economics and Epidemiology Research Office (HE^{2}RO)

Requests from individuals wanting to make use of data forming part of the HE^{2}RO analytic datasets or data collected as part of any HE^{2}RO research project or study for research purposes must follow the following procedure and sign this document:

1. A written request to use any data set or part thereof must be sent to HE^{2}RO's Data Manager Mr Given Malete (gmalete@heroza.org).
   - 1.1. The request should include:
     - 1.1.1. Names of individuals who will be using the data and their affiliations.
     - 1.1.2. Information on whether the research is for degree purposes and details of the institution that will grant the degree.
     - 1.1.3. A study protocol detailing the study background and objectives of using the data, planned analysis, variables required, planned public distribution of results and project timelines including date of finalization.
     - 1.1.4. Details of funding mechanism for the study and study budget detailing allocation for HE^{2}RO data management activities.
     - 1.1.5. Local and/or international IRB clearance certificates where available or date of submission for review if not yet obtained.
   - 1.2. The Data Manager will communicate the request to the directors and senior researchers in HE^{2}RO.
   - 1.3. For requests for data collected as part of any HE^{2}RO research project or study, the Data Manager will communicate the request to the project or study's Principle Investigator.

2. A consensus decision of the research team at HE^{2}RO whether to grant permission for data use will be required in all cases.
   - 2.1. Where the analysis involves site-specific data, the Data Manager will ensure that the relevant documentation required by the sites is supplied to the applicant. It will be the

M.N.M

responsibility of the applicant to submit any such documentation to the appropriate person/s at the sites and obtain site approval if required.

2.2. Once the above groups have reached consensus, the Data Manager will communicate a decision to the applicant.

2.3. The dataset from the relevant cohort(s) containing information required in terms of the applicant's protocol (1.1.3) would be provided by staff at HE²RO in a format suitable for analysis to be agreed upon with the applicant upon signing of this documentation. In addition, timelines for creation and transfer of the datasets will be set upon signing of this document.

2.4. The Data and Site Logistics Unit of HE²RO agrees to provide the required datasets in accordance with point 2.3 within all reasonable conditions. The Unit cannot be held responsible for delays in providing data caused by failure of the applicant to supply HE²RO staff with all necessary documentation. This includes but it not limited to:

- ✓ A signed copy of this data sharing agreement
- ✓ The protocol detailed in point 1.1.3.
- ✓ Copies of local and international IRB clearance certificates where applicable
- ○ Signed confidentiality agreement
- ○ Study budget allocation and approval thereof

3. By signing this SOP the applicant agrees to the following statements of assurance:

3.1. All references to the data either in public verbal presentation or in print must credit the Health Economics and Epidemiology Research Office, Department of Internal Medicine, School of Clinical Medicine, University of the Witwatersrand, the specific cohort as the source.

3.2. In addition, the applicant will also acknowledge the HE²RO's funding in any work that emanates from using this data accordingly (currently INROADS USAID-674-A-12-00029, check with the data manager for the appropriate mechanism).

3.3. Should the applicant wish to use any part of the supplied dataset for analysis beyond the originally submitted proposal, a further written request will be required and the procedures outlined above followed.

3.4. A final draft of the results of the research will be submitted to the HE²RO for information and comment before it is submitted for publication or public presentation.

M.N.M

3.5. The authorship of the paper will be decided using internationally recognized criteria, and must recognize HE²RO and the Department of Medicine researchers appropriately, as well as any international researchers involved in the project.

3.6. The applicant will also be required to submit a copy of the final product resulting from use of the data set to the Project Manager of HE²RO (currently Mr Jovan Dutt, jdutt@heroza.org).

3.7. The applicant will agree not the share the data set to anyone other than the person/s outlined in the data management plan of the study protocol. If this is not specified in the study protocol, HE²RO data management, storage, transfer and sharing policies will apply. (Copies of these policies are available from the Project Manager of HE²RO (currently Mr Jovan Dutt, jdutt@heroza.org).

4. The Human Research Ethics Committee (Medical) of the University of the Witwatersrand has approved the use of these datasets subject to conditions (particularly with regard to the identity of participants) outlined in protocol (M140201).

4.1. The applicant must comply with the conditions laid out in this protocol.

4.2. It will be the responsibility of the applicant to ensure that approval to do their study is covered by the protocol.

4.3. The applicant will be responsible for obtaining approval from any other authorities or Internal Review Boards as may be necessary.

4.4. Note that applicants requesting datasets for higher degree purposes must submit a separate application to the Human Research Ethics Committee of the University of the Witwatersrand.

5. The applicant must make an oral presentation to the Health Economics and Epidemiology Office at the start of the investigation and again once it has been concluded. This will be part of the regular academic programme of HE²RO and can be done by teleconference or Skype if necessary.

6. A file will be maintained in the offices of HE²RO for all correspondence in this regard. In particular:

6.1. Correspondence documenting the approval process as outlined in Point 2 above

6.2. The signed agreement of the applicant (this Data Sharing Agreement).

6.3. A copy of the final product resulting from use of the data.

M.N.M

Health Economics and Epidemiology Research Office

Unit 2, 39 Empire Road
Parktown, Johannesburg, 2193, South Africa
Tel +27 (0)10 0017930

A division of the Wits Health Consortium (Pty) Ltd, a wholly owned subsidiary of the University of the Witwatersrand.

I have read and accept these conditions.

Applicant: MISS MELLISA MABHIKWA
Mlrado

Date: 05 June 2018

Signed on behalf of HE²RO:

Date:

Appendix 6: Turnitin Report



708821:revised_draft_11dec_DE_06Feb2020MEL11FEB.docx

ORIGINALITY REPORT

12%	11%	5%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

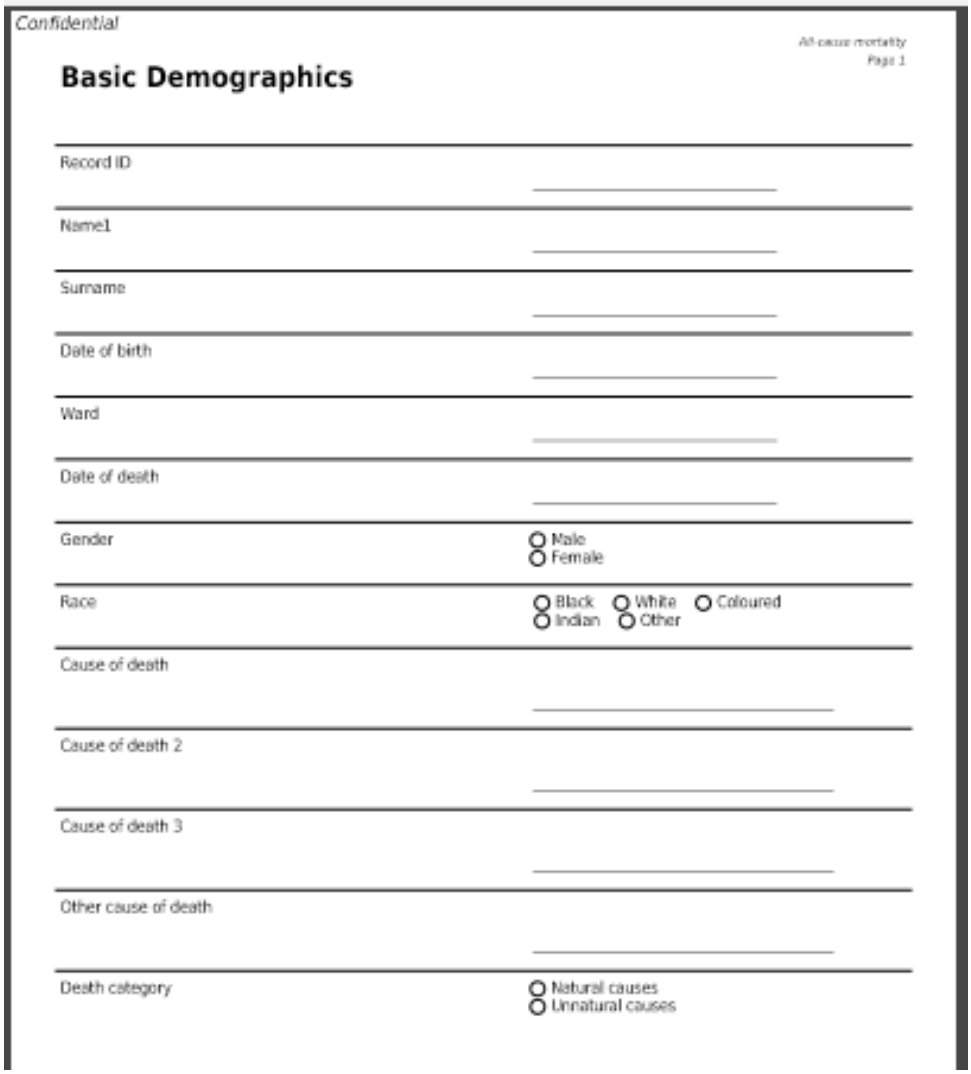
1	www.ncbi.nlm.nih.gov Internet Source	2%
2	www.sahealthinfo.org Internet Source	2%
3	www.auc.ab.ca Internet Source	1%
4	scielosp.org Internet Source	1%
5	espace.library.uq.edu.au Internet Source	1%
6	www.vtb.uscourts.gov Internet Source	1%
7	L C Long, D Evans, S Rosen, C Sauls, A T Brennan, I M Sanne, M P Fox. "Can routine inpatient mortality data improve HIV mortality estimates? Inpatient mortality at an urban hospital in South Africa", South African Medical Journal, 2018 Publication	1%

8	discovery.ucl.ac.uk Internet Source	1%
9	wiredspace.wits.ac.za Internet Source	1%
10	thorax.bmj.com Internet Source	1%
11	www.dovepress.com Internet Source	1%
12	Ingrid Valerie Bassett, Mingshu Huang, Christie Cloete, Sue Candy, Janet Giddy, Simone Claire Frank, Robert A Parker. "Assessing the completeness and accuracy of South African National Laboratory CD4 and viral load data: a cross-sectional study", BMJ Open, 2018 Publication	1%

SUPPLEMENTARY MATERIAL

This section presents information relevant to this study, but due to space constraints could not be presented in the main body of the research report.

Data was collected using a REDCap eCRF on a mobile tablet. The fields collected are shown in the Fig. S1 below. The name and surname fields were deleted after matching was completed and kept in a separate password protected linking file.


Confidential

All-cause mortality
Page 1

Basic Demographics

Record ID

Name1

Surname

Date of birth

Ward

Date of death

Gender — Male, Female

Race — Black, White, Coloured, Indian, Other

Cause of death

Cause of death 2

Cause of death 3

Other cause of death

Death category — Natural causes, Unnatural causes

Figure S1: REDCap data collection electronic case report form

The Coding Causes of Death in HIV protocol was used to classify causes of death. The cause of death categories are shown in Fig. S2 below.

The CoDe system: All of the causes should be coded by using the CoDe codes provided in the below listing. Note that the system is separated in three sections, where the first section including more specific causes takes priority over the second, which again takes priority over the third section:

Code		Description
01		AIDS (ongoing active disease)
	01.1	Infection
	01.2	Malignancy
02		Infection (other than 01.1)
	02.1	Bacterial
	02.1.1	Bacterial with sepsis
	02.2	Others
	02.2.1	Other with sepsis
	02.3	Unknown aetiology
	02.3.1	Unknown with sepsis
03		Chronic viral hepatitis (progression of / complication to)
	03.1	HCV
	03.1.1	HCV with cirrhosis
	03.1.2	HCV with liver failure
	03.2	HBV
	03.2.1	HBV with cirrhosis
	03.2.2	HBV with liver failure
04		Malignancy (other than 01.2 and 03, 03.1, 03.2)
05		Diabetes Mellitus (complication to)
06		Pancreatitis
07		Lactic acidosis
08		MI or other ischemic heart disease
	08.1	AMI
	08.1.1	Definitive AMI (Dundee 1)
	08.1.2	Possible AMI (Dundee 2/9)
	08.2	Other ischemic heart disease
09		Stroke
10		Gastro-intestinal haemorrhage (if chosen, specify underlying cause)
11		Primary pulmonary hypertension
12		Lung embolus

Code		Description
13		Chronic obstructive lung disease
14		Liver failure (other than 03, 03.1, 03.2)
15		Renal failure
16		Accident or other violent death (not suicide)
17		Suicide
18		Euthanasia
19		Substance abuse (active)
	19.1	Chronic Alcohol abuse
	19.2	Chronic intravenous drug-use
	19.3	Acute intoxication (indicate agent)
If the cause of death can't be specifically classified, general classification can be used:		
20		Haematological disease (other causes)
21		Endocrine disease (other causes)
22		Psychiatric disease (other causes)
23		CNS disease (other causes)
24		Heart or vascular (other causes)
25		Respiratory disease (other causes)
26		Digestive system disease (other causes)
27		Skin and motor system disease (other causes)
28		Urogenital disease (other causes)
29		Obstetric complications
30		Congenital disorders
If the cause of death is unclassifiable, use:		
90		Other causes (provide details in Section 1)
91		Unclassifiable causes
92		Unknown

- ***Immediate cause of death:*** The disease(s) or injury directly leading to death.
- ***Contributing cause of death:*** The disease(s) or injury, which contributed to a fatal outcome.
- ***Underlying cause of death:*** The disease or injury, which initiated the train of morbid events leading directly or indirectly to death, or the circumstance of the accident or violence, which produced the fatal injury.

Please note:

Hodgkin's Lymphoma is now classified as a Non-AIDS defining malignancy.

Death reason 8 (MI or other ischemic heart disease) has been subdivided based on use of the WHO Monica Dundee score, and a sudden cardiac death category has been included and an external cardiologist will be supervising the coding of such events.

Figure S2: Classification of deaths according to the Coding Causes of Death in HIV protocol

Source: Copenhagen HIV Progamme. Coding Causes of Death in HIV Protocol. 2015. Accessed 23 February 2019 https://www.chip.dk/Portals/0/files/Code%20Protocol%202.3.pdf

Table S1: Percentages of missing data per variable before and after matching mortuary register data to NHLS and TE.

Variable	Primary data source	Missing data before matching n %	Missing data after matching n %
Sex	Mortuary register	1 (0.1)	0
Race	Mortuary register	118 (6.7)	89 (5.0)
Ward	Mortuary register	61 (3.5)	61 (3.5)
Date of birth	Mortuary register	0	0
Date of death	Mortuary register	0	0
Age	Mortuary register	0	0
Cause of death	Mortuary register	17 (1.0)	17 (1.0)
HIV status	NHLS or TE	1766 (100)	762 (43.1)
CD4+ cell count	NHLS or TE	1766 (100)	1419 (80.4)
Viral load count	NHLS or TE	1766 (100)	1530 (86.6)
WHO stage	TE	1766 (100)	1587 (89.9)
ART regimens	TE	1766 (100)	1491 (84.4)
Completed TB treatment in the last 12 months	TE	1766 (100)	1652 (93.5)
Haemoglobin counts	TE	1766 (100)	1481 (83.8)
BMI	TE	1766 (100)	1485 (84.1)

Note: Missingness of data reduced after records had been matched between the mortuary register and TE and NHLS. Where mortuary register records had a match in both NHLS and TE, clinical data (viral loads, CD4 counts etc.) from TE took precedence over NHLS data because TE data is updated daily. N=1766 mortuary register sample. HIV – Human Immunodeficiency Virus; WHO – World Health Organisation; ART – antiretroviral treatment; TB – tuberculosis; BMI – body mass index; TE – TherapyEdge-HIV™

Table S2: Causes of death by sex.

	Sex		
Code	**Male**	**Female**	**Total**
1. AIDS			
1.1 Infection	130 (54.9)	107 (45.1)	237
1.2 Malignancy	12 (66.7)	6 (33.3)	18
2.Infection			
2.1 Bacterial	147 (48.7)	155 (51.3)	302
2.2 Others	26 (47.3)	29 (52.7)	55
2.3 Unknown aetiology	22 (37.3)	37 (62.7)	59
4.Malignancy (other than 1.2)	68 (53.5)	59 (46.4)	127
5.Diabetes mellitus	18 (51.4)	17 (48.6)	35
8. MI or other ischemic heart disease	17 (58.6)	12 (41.4)	29
9. Stroke	48 (47.1)	54 (52.9)	102
10. Gastro-intestinal haemorrhage	6 (28.6)	15 (71.4)	21
13. Chronic obstructive lung disease	31 (54.4)	26 (45.6)	57
14. Liver failure	21 (50.0)	21 (50.0)	42
15. Renal failure	50 (50.5)	49 (49.5)	99
16. Accident or other violent death	68 (85.0)	12 (15.0)	80
20. Haematological disease	14 (46.7)	16 (53.3)	30
24. Heart or vascular	68 (48.2)	73 (51.8)	141
25. Respiratory disease	8 (36.4)	14 (63.6)	22
26. Digestive system disease	33 (57.9)	24 (42.1)	57
90. Other causes	42 (46.2)	55 (53.8)	91
91. Unclassifiable causes	43 (47.8)	47 (52.2)	90
92. Unknown	38 (52.8)	34 (47.2)	72
Total	910 (51.5)	856 (48.5)	1766

Note: AIDS – Acquired Immunodeficiency Syndrome; MI – myocardial infarction

Table S3: Causes of death by age group.

	Age group			
Code	**18-29**	**30-49**	**50+**	**Total**
1. AIDS				
1.1 Infection	29 (12.2)	162 (8.4)	46 (19.4)	237
1.2 Malignancy	0	14 (77.8)	4 (22.2)	18
2.Infection				
2.1 Bacterial	18 (5.9)	132 (43.7)	152 (50.3)	302
2.2 Others	7 (12.7)	26 (47.3)	22 (40.0)	55
2.3 Unknown aetiology	6 (10.2)	24 (40.7)	29 (49.2)	59
4.Malignancy (other than 1.2)	2 (1.6)	22 (17.3)	103 (81.1)	127
5.Diabetes mellitus	1 (2.9)	6 (17.1)	28 (80.0)	35
8. MI or other ischemic heart disease	0	7 (24.1)	22 (75.9)	29
9. Stroke	3 (2.9)	24 (23.5)	75 (73.5)	102
10. Gastro-intestinal haemorrhage	2 (9.5)	4 (19.0)	15 (71.4)	21
13. Chronic obstructive lung disease	0	4 (7.0)	53 (93.0)	57
14. Liver failure	3 (7.1)	21 (50.0)	18 (42.9)	42
15. Renal failure	8 (8.1)	28 (28.3)	63 (63.6)	99
16. Accident or other violent death	29 (35.3)	26 (32.5)	25 (31.3)	80
20. Haematological disease	3 (10.0)	17 (56.7)	10 (33.3)	30
24. Heart or vascular	4 (2.8)	25 (17.7)	112 (79.4)	141
25. Respiratory disease	2 (9.1)	5 (22.7)	15 (68.2)	22
26. Digestive system disease	3 (5.3)	10 (17.5)	44 (77.2)	57
90. Other causes	5 (5.5)	20 (22.0)	46 (50.5)	91
91. Unclassifiable causes	4 (4.4)	35 (38.9)	51 (57.7)	90
92. Unknown	7 (9.7)	22 (30.6)	43 (59.7)	72
Total	124 (7.0)	635 (36.0)	1007 (57.0)	1766

Note: AIDS – Acquired Immunodeficiency Syndrome; MI – myocardial infarction

Table S4: The ten leading underlying natural causes of death in South Africa 2014-2016.

Causes of death (based on ICD-10)	2014			2015			2016		
	Rank	Number	%	Rank	Number	%	Rank	Number	%
Tuberculosis (A15-A19)	1	39 695	8,3	1	34 042	7,2	1	29 513	6,5
Diabetes mellitus (E10-E14)	3	24 092	5,1	2	25 774	5,4	2	25 255	5,5
Other forms of heart disease (I30-I52)	4	23 009	4,8	4	23 299	4,9	3	23 515	5,1
Cerebrovascular diseases (I60-I69)	2	24 258	5,1	3	23 505	5,0	4	23 137	5,1
Human immunodeficiency virus [HIV] disease (B20-B24)	6	22 866	4,8	5	22 557	4,8	5	21 830	4,8
Hypertensive diseases (I10-I15)	7	18 416	3,9	7	19 845	4,2	6	19 960	4,4
Influenza and pneumonia (J09-J18)	5	22 878	4,8	6	21 001	4,4	7	19 638	4,3
Other viral diseases (B25-B34)	9	14 574	3,1	8	16 475	3,5	8	16 577	3,6
Ischaemic heart diseases (I20-I25)	...	...	...	10	12 714	2,7	9	12 883	2,8
Chronic lower respiratory diseases (J40-J47)	10	12 793	2,7	9	13 006	2,7	10	12 659	2,8
Intestinal infectious diseases (A00-A09)	8	14 834	3,1	...	...	...	...	...	...
Other natural causes		208 537	43,7		207 820	43,9		200 403	43,9
Non-natural causes		50 939	10,7		53 228	11,2		51 242	11,2
All causes		**476 891**	**100,0**		**473 266**	**100,0**		**456 612**	**100,0**

** Including deaths due to *MDR-TB* and *XDR-TB*.

Source: Statistics South Africa. Mortality and causes of death in South Africa, 2016: Findings from death notification. http://www.statssa.gov.za/publications/P03093/P030932016.pdf accessed on 25 February 2020