Long Term Health Outcomes of
Indigenous and Non-Indigenous Australians in the Northern Territory
After Serious Illness: A Survival Analysis Approach

by
Yaofeng Vincent He
BSc(Hons) in Mathematical Sciences

A thesis submitted for the degree of
Doctor of Philosophy

Menzies School of Health Research
Charles Darwin University
Darwin, Northern Territory, Australia

December 2015
DECLARATION

I hereby declare that the work herein, now submitted as a thesis for the degree of Doctor of Philosophy of the Charles Darwin University, is the result of my own investigations, and all references to the ideas and work of other researchers have been specifically acknowledged. I hereby certify that the work embodied in this thesis has not already been accepted in substance for any degree, and is not currently being submitted in candidature for any other degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, and online via the University’s Open Access repository eSpace.

This thesis is presented in the format of a series of journal papers, which have been or are in the process of being published. Other than the sepsis cohort study conducted in 2007-2008 by Dr Joshua Davis and the stroke project ethics approval obtained by Jiqiong You, the work was performed entirely during the period of doctoral degree candidature, between July 2012 and December 2015. The candidate performed the data management and statistical analysis. The candidate obtained ethics and data-custodian approval for the cancer and rheumatic heart disease project. Other than the sepsis project, the candidate was involved in conceiving and designing the experiments for all of the projects.

PhD Supervisors, Professor John Condon and Dr Yuejen Zhao, provided overall guidance and detailed advice throughout the project, and reviewed each draft paper. Several other people also provided advice and assistance with individual parts of the project; their contributions are described in the preface to each chapter. Papers that have already been
published, or accepted for publication, were also reviewed by anonymous expert reviewers, whose comments have contributed to the final work. This thesis has been professionally edited by Melinda Barlow, who offered input on only Standards D and E (language, illustrations, completeness and consistency) from the Australian Standards for Editing Practice.

Yaofeng Vincent He
ABSTRACT

This PhD thesis used survival analysis methods in a novel way to examine differences in long-term outcomes of patients diagnosed with serious, acute conditions in the Northern Territory of Australia. In particular, the thesis compared long-term outcomes for Indigenous patients against those for non-Indigenous patients. It is intended to function at two levels: at the theoretical level, this thesis investigates the potential of using survival analysis to produce useful information from existing data to improve our knowledge about long-term outcomes after acute care. At the practical level, it undertakes four illustrative studies to examine the long-term health outcomes for Indigenous patients compared with non-Indigenous patients after diagnosis and treatment for four severe and acute health conditions: sepsis, stroke, cancer and rheumatic heart disease (RHD).

All four studies found that long-term health outcomes after serious illness were poorer for Indigenous people than they were for non-Indigenous people. Indigenous sepsis patients had higher excess mortality. Indigenous stroke patients had higher case-fatality, stroke recurrence and long-term mortality; comorbidities explained part of the Indigenous disparity in long-term survival after a stroke, but did not explain the disparity in stroke recurrence. Indigenous RHD patients had higher mortality; of which 28% of the excess mortality was explained by comorbid renal failure and hazardous alcohol use. Indigenous cancer patients had a lower net survival rate for breast, colorectal, lung, head and neck cancers. In the presence of competing risk, Indigenous cancer patients with all types of cancer had higher probability of cancer death and higher probability of non-cancer death (except for those with lung cancer and head and neck cancers - both smoking-related cancers).

This thesis has demonstrated the great potential of survival analysis to produce useful information from linked administrative data to improve our knowledge about long-term health outcomes.
ACKNOWLEDGEMENTS

Foremost, I express my deepest gratitude to Mother Nature; the successful completion of this PhD thesis is made possible through a combination of luck and inspiration. In addition, the completion is also due to the generous support of the people below.

I first express my respectful gratitude to my deceased paternal grandfather and grandmother. I am greatly indebted to my father and mother for bringing me up to be the person I am today; without their support and encouragement, I would not have been able to further my studies in Australia, right after my undergraduate studies in my home country—Singapore. I am also greatly indebted to my Aunty, who is and has been a mentor in my life. She has guided me and helped me to find the right direction in life after my recovery from surgery due to an illness. Without her advice, I would not have studied statistics at university (I would have studied physics instead).

I pay my respect to and acknowledge the people and elders of the Aboriginal and Torres Strait Island Nations who are the Traditional Owners of the land and seas of Australia.

I am very grateful to Professor John Condon, who I have had the great fortune of having had as my main supervisor. This thesis would not have been possible without his advice, guidance and support. My PhD journey came with many surprises and unexpected events and it is through his good judgement and advice that I was able to convert them into the best possible outcomes. Other than the ‘outcome’, the ‘process’ is equally important as well; the end of this journey marks the start of another. I was inspired by his personal qualities and the way he dealt with things and interacted with other people, as well as allowing me to plan and take responsibility of my PhD project. He was always there, and regularly went the extra mile for
me to offer advice and help. I will remember his support throughout my life and take the
lessons learned from him with me into my future. I am thankful to my associate supervisors,
Dr Yuejen Zhao, Dr Joshua Davis and Dr Jim Burrow, who consistently provided the right
support at the right time. Without the help of my PhD supervisors, I would not have
successfully completed four different projects (on four different diseases) without the
relevant clinical experience.

I am grateful to Northern Territory Department of Health for providing me the workspace and
approval to conduct the research. I was attached as a student to the Health Gains Planning
Branch and have been able to access the various health administrative datasets since 2012
July. Special thanks to Dr Steve Guthridge (director of Health Gains Planning Branch) who
has given me the opportunity to develop and Jiqiong You who introduced me to the NT
hospital inpatient data and allowed me to be involved in her stroke project, which resulted in
the recurrent stroke and comorbidity study. I thank Karen Dempsey (former NT Cancer
Register data custodian) for extracting the NT Cancer Register data, Dr Xiaohua Zhang
(Health Gains Planning Branch cancer epidemiologist) for preparing life tables for the
relative survival analysis, Cath Milne (NT Rheumatic Heart Disease (RHD) control program
officer) for extracting the RHD Register data; Dr Shu Qin Li (Health Gains Planning Branch
senior epidemiologist) for linking the RHD Register to the hospital inpatient data and the
staff of the NT DoH data management and system reporting branch for extracting and linking
the hospital inpatient data and mortality data.

I thank Professor Peter Baade, who advised and promptly assisted in reviewing and revising
the manuscript for the cancer project. For the RHD complication project, special thanks to
Professor Jonathan Carapetis who provided the research question to start with and Dr Anna
Ralph who kindly provided the relevant clinical knowledge and assistance despite her busy schedule. I am also grateful to other co-authors of the RHD complication manuscript: Marea Fittock, Jessica de Dassel, Dr Kathryn Roberts, Dr Keith Edwards and Professor Bart Currie.

I would also like to thank Associate Professor Gail Garvey, a senior Indigenous research, for her guidance and advice during my PhD and SCNAP-IP (Supportive Care Needs Survey for Indigenous patient) project. During my PhD study, she has employed me to develop and evaluate a culturally appropriate tool (SCNAP-IP) to identify the support needs of Indigenous cancer patients and test the feasibility of this tool in the usual cancer care setting using psychometrics analysis, which is now being used across Australia in cancer care services.

Many thanks to Catherine Richardson and Caroline Sheridan for their support in the Education and Research support division.

I am grateful to Charles Darwin University in supporting me through the University Postgraduate Research Scholarship from Jan 2013 to July 2015.

I express my sincere thanks to my honours thesis supervisors, Associate Professor Xiang Liming and Assistant Professor Lian Heng at Singapore Nanyang Technological University; it was them that inspired my interest in survival analysis and data mining, therefore opening the door to health research. I also thank Professor Chong Siow-Ann, Dr Mythily Subramaniam, Janhavi Vaingankar and Dr Edimansyah Abdin from the research division of Singapore Institute of Mental Health that have selflessly shared their research experience and given me the opportunity to have first-hand experience in the research work environment from 2010-2012.
I would also like to acknowledge Melinda Barlow, who helped to promptly proofread this thesis based on Standards D and E (language, illustrations, completeness and consistency) from the Australian Standards for Editing Practice.

In my PhD journey, I feel privileged to have the companionship of the staff at the NT DoH Health Gains Planning Branch, the staff and students (Sharon Chirgwi, Lisa Whop, Abbey Diaz, Brooke Barnett, Jess De Dassel, Anna Szava, Jodi Phillips and Paul Lawton) at Menzies School of Health Research and Khailiang Khor, Huang-Yue Cheow and Ryan Liu.

I dedicated this thesis to my grandparents, parents and my Aunty.
# CONTENTS

1. Introduction ....................................................................................................... 17  
   1.1 Background ........................................................................................ 17  
   1.2 Thesis aim .......................................................................................... 21  
   1.3 Importance of health data, information and knowledge in public health policy ........................................................................ 22  
   1.4 Use of administrative data and data linkage in Australia .......... 23  
   1.5 Cohort study (subjects, outcomes/exposures and time) ............ 26  
   1.6 Outcome research and long term outcomes .................................. 27  
   1.7 Time (dates) in administrative data.................................................... 29  
   1.8 Survival analysis ................................................................................ 30  
   1.9 Outline of the thesis ........................................................................... 31  
   1.10 Reference ........................................................................................... 35  

2. Long term outcomes following hospital admission for sepsis using relative survival analysis: A prospective cohort study of 1,092 patients with 5 year follow up. 41  
   2.1 Preface .................................................................................................. 41  
   2.2 Article .................................................................................................. 42  
      2.2.1 Abstract .......................................................................................... 43  
      2.2.2 Introduction .................................................................................... 44  
      2.2.3 Methods.......................................................................................... 44  
      2.2.4 Results............................................................................................ 46  
      2.2.5 Discussion ...................................................................................... 49  
      2.2.6 Reference ....................................................................................... 54  

3. Adverse outcome after incident stroke hospitalization for Indigenous and non-Indigenous Australians in the Northern Territory......................... 57  
   3.1 Preface.................................................................................................. 57  
   3.2 Article .................................................................................................. 58  
      3.2.1 Abstract .......................................................................................... 59  
      3.2.2 Introduction .................................................................................... 59  
      3.2.3 Methods.......................................................................................... 59  
      3.2.4 Results............................................................................................ 60  
      3.2.5 Discussion ...................................................................................... 60  
      3.2.6 Reference ....................................................................................... 64  

4. Different survival analysis methods for measuring long term outcomes of Indigenous and non-Indigenous Australian cancer patients in the presence and absence of competing risks. 67  
   4.1 Preface.................................................................................................. 67  
   4.2 Article .................................................................................................. 68  
      4.2.1 Abstract .......................................................................................... 70  
      4.2.2 Background .................................................................................... 72  
      4.2.3 Methods.......................................................................................... 74  
      4.2.4 Results............................................................................................ 76  
      4.2.5 Discussion ...................................................................................... 81  
      4.2.6 Reference ....................................................................................... 85
TABLES

Table 1.1 Summary overview of the outcomes of interest, data sources and the specific survival analysis methods for the four PhD projects..................................................................................................................34

Table 2.1: Demographics, comorbidities, characteristics of infection and disease severity in those who died compared with those who were alive at the end of follow up.............47

Table 2.2: Interval-specific relative survival by age category (all sepsis patients)..................52

Table 2.3: Interval specific relative survival by age category (severe sepsis patients).............52

Table 2.4: Five-year relative survival (with 95% confidence interval) by age group, sex and Indigenous status..........................................................................................................................53

Table 2.5: Multivariable analysis of predictors of excess mortality using Poisson regression......53

Table 3.1: Demographic characteristics of people hospitalized for their first stroke, Northern Territory, 1996–2011, by Indigenous status.........................................................61

Table 3.2: Regression analysis for case fatality, recurrent stroke, and overall deaths.............62

Table 4.1: Demographic characteristics of people diagnosed with cancer, Australia NT, 1991-2009.........................................................................................................................77

Table 4.2: One-year and five-year cumulative relative survival (%) and 95% confidence interval by Indigenous status and cancer site, Australia NT, 2001-2009.................................77

Table 4.3: Regression analysis of cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk (Fine-Gray regression), all cancers combined, Australia NT, 1991-2009.........................................................79

Table 4.4: Regression analysis of time trend after diagnosis using cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk analysis (Fine-Gray regression), all cancers combined, Australia NT, 1991-2009......79

Table 4.5: Five-year cumulative probabilities of cancer and non-cancer death (%), age and sex adjusted..............................................................................................................80

Table S4.3: Regression analysis of cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk (Fine-Gray regression), all cancers combined, Australia NT, 1991-2009 (full model) .........................................................87

Table S4.4: Regression analysis of time trend after diagnosis using cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk analysis (Fine-Gray regression), all cancers combined, Australia NT, 1991-2009 (full model).................................................................88

Table 5.1: Diagnosis codes for adverse outcomes and health behaviour-related hospitalisation.......93
Table 5.2: Demographic Characteristics of ARF and RHD patients, NT, 1997 to 2013.................94

Table 5.3: Incidence rate (per 100 person-years) and cumulative incidence rate (%) (95% confidence interval) of adverse outcomes for ARF patients (n=572) at different years after the first ARF diagnosis.................................................................95

Table 5.4: Multivariable Cox regression for ARF adverse outcomes, NT, 1997 to 2013............95

Table 5.5: Hospitalization for ARF/RHD Treatment of ARF Patients (n=572) and Hospitalization (Within 1 Year of RHD diagnosis) for RHD Treatment of RHD Patients (n=1248), NT, 1997 to 2013..................................................................................................................96

Table 5.6: Incidence Rate and Cumulative Incidence Rate (95% Confidence Interval) of Adverse Outcomes for RHD Patients (n=1248) at Different Years After the First RHD Diagnosis.........................................................................................................................96

Table 5.7: Multivariable Cox Regression for RHD Adverse Outcomes, NT, 1997 to 2013............97

Table 5.8: Prevalence of Selected Comorbidities and Health Behaviour–Related Hospitalizations Among Indigenous (Urban and Remote) and Non-Indigenous RHD Patients (≥18 Years Old).................................................................................................................................98

Table 5.9: Multivariable Cox Regression for Mortality of RHD Patients (≥18 Years Old) With Adjustment for Different Comorbidities.................................................................98

Table 7.1: A summary of different survival analysis methods used in Chapter 4......................121
# FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1a: Kaplan-Meier crude survival estimates by age group for all sepsis patients (n=1,028)</td>
<td>48</td>
</tr>
<tr>
<td>Figure 2.1b: Kaplan-Meier crude survival estimates by age group for severe sepsis patients(n=228)</td>
<td>48</td>
</tr>
<tr>
<td>Figure 2.2a: Cumulative relative survival for patients with sepsis and severe sepsis, compared with a reference population</td>
<td>50</td>
</tr>
<tr>
<td>Figure 2.2b: Cumulative relative survival for only patients with severe sepsis (n=228), compared with a reference population</td>
<td>50</td>
</tr>
<tr>
<td>Figure 2.3a: Cumulative relative survival by age group for all sepsis patients (n=1,028)</td>
<td>51</td>
</tr>
<tr>
<td>Figure 2.3b: Cumulative relative survival by age group for severe sepsis patients (n=228)</td>
<td>51</td>
</tr>
<tr>
<td>Figure 3.1: Age-adjusted cumulative incidence function and hazard function for recurrent stroke after incident stroke hospitalization</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3.2: Stack cumulative incidence function plots for recurrent stroke, death, and survival after incident stroke hospitalization by age group and Indigenous status</td>
<td>63</td>
</tr>
<tr>
<td>Figure 6.1: Possible disease progression of ARF and RHD patients</td>
<td>112</td>
</tr>
</tbody>
</table>
CHAPTER ONE: INTRODUCTION

1.1 Background

Indigenous (Aboriginal people and Torres Strait Islanders) Australians are the original inhabitants of Australia, estimated to have arrived in Australia more than 40,000 years ago (compared to the first Europeans who colonised Australia in 1788). Indigenous Australians are a much younger population than other Australians (median age is 21.8 years compared with 37.6 years for the non-Indigenous Australian population (1)), are more likely to live in remote areas (21% compared to 2% for the non-Indigenous Australians (1)) and have suffered severe social and economic exclusion over the past two centuries since European colonisation (2-4).

Worldwide, Indigenous populations suffer higher mortality than their non-Indigenous counterparts (5). In Australia, the Indigenous to non-Indigenous differentials in life expectancy are greater than other developed countries (5); the life expectancy of Indigenous males born in 2010–2012 was estimated to be 10.6 years lower than that of non-Indigenous males (69.1 years compared with 79.7) and 9.5 years lower for females (73.7 compared with 83.1) (6). The epidemiological transition (7,8) (to the “Age of Degenerative and Man-Made Diseases” (7,8)) has seen chronic diseases replace infectious diseases as the leading causes of morbidity and mortality for both Indigenous and non-Indigenous Australians; cardiovascular disease and cancer are now the two leading causes of death for both populations (1) (25.5% and 19.8% of Indigenous deaths respectively; 32.3% and 29.7% of non-Indigenous deaths respectively). Chronic diseases now account for 81% of the gap between Indigenous and other Australians in mortality (3) and 70% of the gap in burden of disease (9). The greatest contributor to the Indigenous mortality gap was cardiovascular disease (including coronary heart disease, stroke, heart failure and rheumatic heart disease) (24.1%), followed by diabetes (19.1%), respiratory diseases (12.0%), cancers (11.3%), digestive diseases (6.9%) and kidney diseases (4.7%) (1,3).
Indigenous Australians comprise only a small proportion of the total Australian population, ranging from 2% to 4% in each state and territory, with one exception; Indigenous people comprise 29.8% of the population of the Northern Territory (NT) (10), which is a large, sparsely settled jurisdiction in central and northern Australia. Like Indigenous Australians elsewhere, the NT Indigenous population has a relatively young age distribution; their median age was 23.8 years (compared to 37.6 for the non-Indigenous Australian population in 2011), with 32.7% aged less than 15 years (10). NT Indigenous people were more likely to be socioeconomically disadvantaged and remote dwelling, with a higher proportion living in an area of relatively highest disadvantage (56.7%) and very remote area (58.3%) compared to that of the NT non-Indigenous population (1.7% and 8.2% respectively) in 2011(10). Compared to other Indigenous Australians nationally, Indigenous Australians in the NT were also more likely to live in an area of high disadvantage (73.9% vs 51.6%(3)) and remote/very remote areas (79.9% vs 21.4%(11)). Smoking prevalence was more than twice as high for NT Indigenous adults (56.0%) compared to non-Indigenous adults aged 18 and over (24.1%) in 2012-2013 (12). NT has the highest proportion of current smokers (54%) being Indigenous Australians that are aged 15 and over (national average for Indigenous current smokers was 44%) (3).

The mortality rates of the NT Indigenous population are much greater than those of the NT non-Indigenous population (13-17); although the NT Indigenous mortality rate has declined over recent decades (1966-2001), this reduction has been less than that of non-Indigenous Australians, so that the gap between them has increased (18). The life expectancy gap between Indigenous and non-Indigenous Australians is wider in the NT than elsewhere in Australia; the life expectancy of Indigenous males born in 2010–2012 was estimated to be 14.4 years lower than for non-Indigenous males (63.4 years compared with 77.8) and 14.4 years lower for females (68.7 compared with 83.1) (6). Other than their non-Indigenous counterparts, Indigenous people in the NT also have lower life expectancy than the Indigenous Australians nationally (male: 67.4 years and female: 72.3) (6).
Indigenous Australians have higher prevalence of and mortality from chronic diseases and multimorbidities than the Australian general population. Chronic disease is the main contributor to the Indigenous life expectancy gap in the NT and its contribution to the life expectancy gap has increased markedly over time (15). In NT remote Aboriginal communities, 40% of people had at least two chronic diseases before the age of 50 years, 60% had at least two and 30% had at least three diseases after the age of 50 years (19). In another study, it was found that the age adjusted mortality rate ratios of NT Aboriginal deaths to all Australian deaths and to NT non-Aboriginal deaths were significantly higher for all five chronic diseases examined (ischaemic heart disease, diabetes, stroke, chronic obstructive pulmonary disease and renal failure) (16).

To address the high prevalence of chronic disease of both the Indigenous and non-Indigenous population in the Northern Territory, the NT Preventable Chronic Disease Strategy (PCDS) was developed in 1997. It was the first integrated non-communicable diseases strategy in Australia. The strategy (20,21) highlighted the importance of a long-term and integrated approach and proposed a three point framework(20). It states that:

“...[C]hronic diseases affect the whole NT population, as unhealthy lifestyles are an Australia-wide problem, but Aboriginal and Torres Strait Islander Territorians are particularly affected, and are more likely to have multiple chronic diseases. The strategy outlined here is relevant to the whole population, Indigenous and non-Indigenous, urban and rural. The cost of not intervening early is too great - an inexorable rise in deaths, hospitalisations, disease complications and financial costs incurred in relation to events such as renal dialysis...

This strategy proposes a three-point framework to guide THS’s [Territory Health Services] activity in this area - prevention, early detection and best practice management.

“... Implementation of this strategy will lead to a delay in onset and a reduced number of adverse health outcomes in the short to medium term, as well as a reduction in long-term financial costs, but the full impact of all the interventions will not be felt for some years. So this is a staged long-term strategy to overcome problems that themselves have arisen over many years…”

“The PCDS approach is innovative in its unremitting commitment to integration - an integrated theoretical framework that encompasses social and medical determinants of health; an integration of client, clinical (individual-level) and public health (population-level) perspectives; an integrated approach to the underlying risk factors for chronic disease…”

According to the PCDS, it was essential to adopt an integrated approach and life course approach to the prevention, detection and management of chronic diseases, rather than the single-disease approaches dominating existing health systems, that are increasingly inappropriate for a population with multiple diseases (comorbidities) or risk factors (22).
Several studies have reported that Indigenous people in the NT suffer poorer outcomes from hospital treatment of serious illness (23,24). The reasons for adverse health outcomes are multi-factorial and complex (25), including greater burden of disease and poorer treatment received, which could be caused by environmental, social and economic factors, and in some cases genetic differences.

Although Indigenous Australians have higher prevalence of chronic diseases, it is still unclear to what extent this leads to their adverse health outcomes after treatment of serious illness (26). A study of people with acute myocardial infarction (AMI) in the NT (26) found that Indigenous patients had higher long-term death rates than non-Indigenous patients, which was only partially explained by higher levels of chronic disease co-morbidity. However, for other diseases, there is limited evidence available about the role of chronic disease co-morbidities in the worsening of outcomes for Indigenous people with serious, acute health conditions.

The NT Indigenous population is facing a double burden of chronic disease and infectious disease, which is often associated with socioeconomic disadvantage. Although the contribution of infectious diseases to the life expectancy gap has decreased over time, NT Indigenous Australians are still suffering from a much higher burden of infectious diseases than other Australians (27). They have the highest incidence rates of sepsis (28) and acute rheumatic fever (29) reported in the world, exacerbated by the complex interaction between infectious diseases, chronic diseases and poverty.

Most health care, and almost all acute care, for NT Indigenous people is provided by community health centres, public hospitals and other services (outpatients, pathology and pharmacy services as well emergency department presentations and a range of non-hospital services including urban community services and a network of 53 health centres located in remote Aboriginal communities) operated by the NT Department of Health. All departmental health services use the same unique client identifier and a central Client Master Index (that includes Indigenous status) for their client information systems. The unique client identifier was introduced into NT public hospitals in 1991-1992 (there is only one private hospital in the NT which treats very few Indigenous inpatients). The
NT has very high accuracy of Indigenous status data in health records. A data quality audit in 2008 found 96.9% agreement between Indigenous status reported at interview by hospital inpatients and what was recorded in the Client Master Index (30). The accuracy of Indigenous status data has improved in recent years for some (but not all) other states, but only the NT has high-quality data spanning several decades. The higher accuracy of Indigenous status data in health records in the NT, together with the presence of the unique client identifier, provides an opportunity to investigate long-term trends that is not possible elsewhere.

1.2 Thesis aim

This PhD thesis examines the long-term outcomes of four severe, acute conditions (and the role of comorbidities) of Indigenous and non-Indigenous Australians in the NT using survival analysis methods in a novel way. It is intended to function at two levels: at the theoretical level, it investigates the potential of using survival analysis to produce useful information, from existing data, to improve our knowledge about medium and long-term outcomes after acute care. At the practical level, this thesis undertakes four illustrative studies to examine the long-term health outcomes (recurrence and survival) for NT Indigenous Australians (compared with non-Indigenous) and the effects of comorbidities after treatment for severe and acute health conditions. It investigates the development of survival and/or recurrence/complications for people suffering four severe, acute conditions:

1. Sepsis: long-term survival after successful treatment of an acute sepsis episode
2. Stroke: case-fatality, stroke recurrence and long-term survival in those who survive their first stroke
3. Cancer: survival after diagnosis of cancer, and
4. Rheumatic heart disease: incidence of and mortality from serious complications (stroke, endocarditis, atrial fibrillation, heart failure) and long-term survival.

After introducing the Indigenous and non-Indigenous health conditions in Australia and specifically the Northern Territory in the ‘background’ section, this thesis would now introduce the importance of health information in improving health by informing public health policy and the use of existing data and data linkage in Australia.
1.3 Importance of health data, information and knowledge in public health policy

The importance of health information in improving health outcomes is widely recognised. For example the Health Metrics Network (HMN), a global health partnership focused on strengthening health information systems in low and middle income countries, was founded on the premise that “better health information means better decision-making, leading to better health” (31). Although there are increased efforts in public health data-collection in many countries, there is under-utilisation of such data due to the lack of understanding about how “such data can be assessed, analysed and interpreted to provide evidence for policy-makers”, resulting in “rich data” but “poor information” (31). Data is not equivalent to information. However these two terms are often used interchangeably in the health literature and it is important to distinguish the two.

The data–information–knowledge–wisdom hierarchy (DIKW)(32-35) - a term used in information science for many years (36) - illustrates the relationship between data, information and knowledge. Despite the different definitions in different fields (37,38), it has been widely accepted and agreed that “data, information, and knowledge are part of a sequential order” and “data are the raw material for information, and information is the raw material for knowledge” (37); generally speaking, “information is meaningful data, and knowledge is information that is true, justified, and believed” (39,40). According to Awad and Ghaziri (33), information is data “in formation”—data that is “classified, sorted, displayed, communicated verbally or graphically, or in the form of tables”; knowledge is “actionable information”; wisdom is the “highest level of abstraction with vision, foresight, and the ability to see beyond the horizon”. It has been claimed that an informed decision is better than an uninformed one (41); however this is only true if the information provided is correct and policy-makers have the wisdom to understand, accept and utilise the information in the decision making or public policy formation process.
Many researchers and policymakers have recognised the importance of the use of knowledge and evidence from research in decision making, public policy formation (42-46) and “the collaborative problem-solving between researchers and decision-makers that happens through linkage and exchange” (knowledge exchange) (47). The Australian Bureau of Statistics (ABS) published a guide for using statistics for evidence based policy in 2010, which states the importance of using statistical information “for making evidence based decisions that guide the implementation of new policy, monitor existing policy, and evaluate the effectiveness of policy decisions” (48). To provide information to aid decision making or public policy formation, quantitative data such as survey data, census data or existing data (e.g. administrative data) could be utilised (48). This thesis focuses on using survival analysis to produce information from existing data to improve our knowledge about long-term health outcomes of NT Indigenous and non-Indigenous Australians after serious illnesses.

1.4 Use of administrative data and data linkage in Australia

This thesis used several types of data sources to investigate health outcomes for Indigenous patients: health service administrative data (hospital inpatients); disease registers (cancer register); statutory data (deaths register); clinical information systems (RHD Register, a treatment recall and reminder system); and a dataset from a previous research project (sepsis). The use of existing data for clinical, epidemiological and health research is well-established (49); observational outcomes research frequently relies on existing data (50). Administrative data (data that were originally collected for reasons other than research (51)) has considerable but underutilised potential to be used for a wide range of purposes: research in evaluating policy; assessing interventions and treatment; assessing cost-effectiveness of healthcare; examining the influence of external events on health; investigating health inequalities; comparing geographical variation; identifying trends in health care; assessing data quality; and predictive research(52). Administrative data provide a relatively cheap, potentially less intrusive resource for research (52), with a large sample size, wide population coverage and long observation periods (53), potentially higher quality data (54) and shorter research timeframe (55). On the other hand, caution is required when using administrative data due to various possible
Limitations including: (i) lack of sociodemographic information and detailed clinical data (e.g. stroke severity), (ii) possible misclassification, (iii) possible selection bias, (iv) possible under-reporting and (v) the possibility of having missing items or missing records. Therefore, a well-designed study with the appropriate statistical techniques (e.g. propensity score methods) is imperative, which could overcome some of the limitations of using administrative datasets.

Tapping into the potential of using administrative data, many European, particularly Scandinavian, countries are much more advanced in the use of administrative data for research. Other than the potential for research, administrative data also have the potential to be a source of evidence for effective policy making and evaluation of existing policies. Compared to other countries, Australia has not fully tapped this potential, despite being well-positioned to utilise administrative data: government departments have many high-quality administrative datasets; computing technology and infrastructure make access to and analysis of very large datasets feasible; and legislation that enables the safe use of administrative data for secondary purposes (the Freedom of Information Act 1982 declared that “information held by the Government is to be managed for public purposes, and is a national resource”) has not been fully utilised in Australia due to the lack of “a culture of information sharing and proactive data release”.

Fortunately (or unfortunately), while “progress [in utilising health administrative data for research] in other Australian jurisdictions has been patchy”, Western Australia appears to lead the field in untapping the potential of existing health data for research. Since the late 1970s, Western Australia has utilised administrative data using data linkage techniques for health research. “Instigated in 1995 to link up to 40 years of data from over 30 collections for an historical population of 3.7 million”, the Western Australia Data Linkage System (WADLS) has...

---

1 It is worth noting that in Australia, only Western Australia has “no regulation covering use and disclosure of health information” (Lovett et al. 2008).
2 Data linkage is defined as “the bringing together from two or more different sources, data that relate to the same individual, family, place or event” (62). Commonly linked health data includes hospital data, birth and death registrations and residential care records. The development history of Western Australia’s data linkage can be found in Holman et al. (2008) and Holman et al. (1999). The PhD thesis “Monitoring health care using national administrative data collections” by Kelman CW offers a comprehensive overview of the development history of data-linkage in Australia.
become one of the leaders internationally (54), “noted for highly productive research utilising linked administrative data while maintaining privacy and confidentiality” (64). Since 1995, WADLS has supported over 800 projects (65), including health services utilisation and outcomes studies, aetiologic research, disease surveillance, needs analysis and methodological research (62). As well as providing commercial benefits\(^3\), WADLS also improves data quality of existing datasets, producing cost-efficient research that contributes to medical and scientific knowledge while conserving patient privacy (62). The first systematic compilation of projects using WADLS showed that these projects had led to significant policy and clinical practices reforms in Western Australia, achieving public good through research made possible by data linkage (66). Compared to the conventional experimental studies that are restrictive in their patient selection criteria, observational studies based on population-based administrative data are more representative (54,67); therefore the outcomes can be generalised to the wider population, which is important for policy-making for the whole population.

Inspired by WADLS’s success, the New South Wales (NSW) and Australian Capital Territory (ACT) governments established the Centre for Health Record Linkage (CheReL) in 2006 (with the help of WADLS (68)) to create and maintain a record linkage system for health and human services in NSW and the ACT. Australia’s first national data linkage network, the “…Population Health Research Network (PHRN), was established in 2009 with the support of the Australian Government (as part of the National Collaborative Research Infrastructure Strategy), State and Territory Governments and academic institutions” (69,70). Its principal purpose is to “build a nationwide data linkage infrastructure capable of securely and safely linking and integrating data collections from a wide range of sources [across different jurisdictions in Australia]” (71). The PHRN is a network of state-based data linkage units; WADLS and CheREL have been joined by the Queensland Research Linkage Group, SA-NT Datalink, the Tasmanian Data Linkage Unit and Victorian Data Linkages, with a national data linkage unit for Australian Government data (72) (the Centre for Data Linkage) and the Australian Institute of Health and Welfare as an Integrating Authority (71,72). Other than

\(^3\) Commercial return on research infrastructure investment has exceeded 1000% (62).
Australian Government data, most data linkage is conducted at the state level by each state’s linkage unit, so there is currently no established mechanism for linkage of data from different states (70).

Despite having national data linkage infrastructure in place, major challenges to data linkage research remain, most of which are not technical, but rather due to the legislative, organisational, social and political barriers (73). The complexity, duplication and lack of cohesion in the approval processes (74) are costly and produce lengthy delays (75) (e.g. a national study of cervical screening for Indigenous women “…took 18 months to acquire the approvals necessary to link the three data sources for Queensland and an additional 5 months for the data to be linked by the relevant data linkage agency; the process is still underway in other jurisdictions”). Due to these practical issues, there have been calls for the streamlining of data linkage applications and HREC approval processes across all jurisdictions (74) and an Integrated Health Record and Information System (IHRIS) (60,76) with a unique/universal patient identifier (UPI) for all persons in Australia in the future. The feasibility of national data linkage in Australia was demonstrated by a recent project that linked over 44 million morbidity and mortality records from four Australian states (77).

1.5 Cohort study (subjects, outcomes/exposures and time)

The four studies in this thesis are all cohort studies, investigating “long-term” outcomes. What is a cohort study? A cohort study investigates “the association between the identified exposure(s) and various outcome(s) over time” by “follow[ing] up two or more groups from exposure to outcome”. If one group exposed to some factors “has a higher or lower frequency of an outcome than the unexposed, then an association between exposure and outcome is evident” (78) The article “cohort studies: marching towards outcomes” by Grimes and Schulz (2002) (78) provided a good introduction

---

4 A summary of the application process, requirements and documents and Human Research Ethics Committee approval timeframe for data-linkage projects in different states can be found in the article Mitchell et al (2015) (74)

5 Dr Merran Smith, CEO of Population Health Research Network, has proposed “standardisation of application forms (Online Application System)”, “mutual recognition by ethics committees” and “indicative timeline targets for data custodian approval” across all jurisdictions at the SA-NT DataLink Research Conversations at 2014. (www.santdatalink.org.au/files/publication/M_Smith_Conversations_Dec2014.ppt)
of cohort study. In any cohort study, there are three key elements (79, 80): (a) subjects, (b) outcomes and exposures, and (c) time. In the background section of this introduction chapter, I have given an introduction of the ‘subjects’ and ‘exposures’ (NT Indigenous and non-Indigenous Australians and their related risk factors). Now I will introduce ‘outcome’.

1.6 Outcome research and long-term outcomes

According to Donabedian (1966) (81), “outcomes, by and large, remain the ultimate validators of the effectiveness and quality of medical care”, representing the results of the structure and process of medical care (82). The Donabedian (1966) framework (81) in health care quality evaluation research has provided the theoretical background for outcome research (51-53), which “provide scientific evidence relating to decisions made by all who participate in health care” (44). Due to the “growing realisation that evidence base is lacking for much of daily clinical practice” and the rapid growth in health care costs that “forced clinicians, payers and society to consider the value of medical care” (50), there is growing interest in outcome research. Outcome research complements rather than competes with traditional clinical research, thus improving the knowledge and evidence base of medicine (83). Traditional clinical research is a disease-centred approach that investigates the mechanisms of diseases and the effect of biochemical and physiologic factors on biophysiological changes. Outcomes research is a patient and community-centred approach that investigates the impact of disease on the patients and the effect of socioeconomic factors on patient-centred outcomes (50).

As patients’ short-term survival after serious illness improves, there is growing interest in long-term outcomes (84) rather than focusing only on the traditional immediate outcomes of critical care such as survival to Intensive Care Unit (ICU) and hospital discharge. It has been argued that “long-term

---

6 In the Donabedian (1966) framework, information about quality of care can be drawn from three categories: “structure,” “process,” and “outcomes.” ‘Structure’ refers to the conditions under which care is provided such as material resources (e.g. facilities and equipment), human resources (e.g. number, variety, and qualifications of professional and support personnel) and organisation characteristics (e.g. organisation of medical and nursing staff). ‘Process’ refers to the activities that constitute health care: diagnosis, treatment, rehabilitation, prevention and patient education. ‘Outcomes’ refers to changes in individuals and populations that can be attributed to health care (i.e. complication, mortality, recurrence, discomfort, dissatisfaction).
outcomes are much more important than immediate outcomes since relatively little is achieved if patients die soon after hospital discharge” (85). Good short-term outcomes do not necessarily imply good long-term outcomes; there have been many cases in which an intervention resulted in good short-term outcomes but was subsequently shown to have no effects or even harmful effects on long-term outcomes (84) (e.g. In the Cardiac Arrhythmia Suppression Trial (CAST), it was found that, after an acute myocardial infarction, Flecanide resulted in early positive effect (decreased arrhythmias) but negative late effect (higher mortality) (86)). Policy makers, health professionals and researchers need to keep this in mind.

Selecting the time point to measure the outcome is important as well. Although death is an unequivocal endpoint, there are different time points at which to measure it (e.g. 30 or 90 days, 6 months, 1 or 5 years); the selection of the most informative mortality time point depends on the research question, study design and the disease to be studied (50). For example, studies have suggested that 90-day mortality might be an inappropriate time point for a critical illness that required prolonged hospitalisations (87); long-term outcome would be a better outcome measurement.

However, it is difficult and expensive for clinical and experimental studies to measure long-term outcomes (84). Using administrative data for long-term follow-up of participants in a prospective cohort study could be an effective and efficient alternative. In the first paper of this thesis, death information (vital status and death date) was obtained in 2012 for a sepsis prospective study conducted in 2008-09. Administrative data have also been widely used for retrospective cohort studies in identifying the subjects and the outcomes. In the second paper of this thesis, the hospital inpatient data was used to identify the cohort of stroke patients and the adverse stroke outcomes (case-fatality, recurrent stroke hospitalisation and long-term death). In the third paper of this thesis, the cancer register data was used to identify the cohort of cancer patients and the National Death Index was used to obtain the death information (vital status, cause of death data and death date). In the fourth paper of this thesis, the RHD register data was used to identify the cohort of ARF/RHD patients and the information about the ARF recurrence and progression to RHD and severe RHD; information about
RHD complications (heart failure, endocarditis, stroke and atrial fibrillation) and death information (vital status and death date) were obtained from the hospital inpatients data and mortality data respectively.

1.7 Time (dates) in administrative data

After subjects and outcome/exposure, time is the third key element of a cohort study. Since recorded history, human beings have developed ways to keep track of time and document records of past events which translated into history. This is for practical needs as well as for accountability; this record keeping helped to share the knowledge of the past with other people, informing them of the event and the time that it occurred, which smoothed the work process while providing them with a platform to improve the future by reflecting on the past. With the advancement of technology and invention of computers, information about time has been digitalised and stored as administrative data. ‘Date’ is an essential component of any administrative data. For this reason, administrative data is a good source to study the natural history of a disease over a long period of time, enabling a researcher to select a cohort of people who had a particular disease (e.g. stroke hospitalisation) in a period of time (e.g. year 1996-2013) based on the dates recorded in the administrative data. It also enables identification of the outcome event (e.g. death) that occurred for members of the cohort in the period following the onset of this disease (the study ‘follow-up’ period). However, not all events occurring in the cohort members might apparent from administrative data sources; some cohort members might leave the study early (e.g. die before experiencing a recurrent stroke, which is the outcome of interest) or be lost to follow-up (e.g. move interstate or overseas where their hospital data is not available to the study), resulting in incomplete observation of the outcome within the study period. Subjects might enter the study at different times and there might be varying follow-up times. Survival analysis can overcome such problems and is the most relevant statistical method in such situations.
1.8 Survival analysis

Survival analysis, also known as time to event analysis, is the study of the length of time until an event of interest occurs. It has different names in different disciplines—‘event history analysis’ in sociology, ‘duration analysis’ or ‘duration modelling’ in economics and ‘reliability theory’ or ‘reliability analysis’ in engineering. Other than scientific research, it has also been widely applied in social research fields such as demographic analyses, labour market studies, organisational behaviour and political science research.

The main advantage of using survival analysis in administrative data is its ability to deal with incomplete observations of survival time, referred as ‘censored’ observations. Subjects with shorter follow-up time can contribute that time to the measurement of time to an event, up to the end of their follow-up time, and so can still be included in the study. For example, in the stroke study, those patients that have not experienced the event (death) in the study period were considered as ‘censored observations’ and were still included in the survival analysis. Survival analysis accounts for varying length of follow-up time due to censoring, when subjects enter or leave the study at different times, which “enable[s] a very efficient use of data where there are varying durations of follow-up” (88).

The second advantage of survival analysis is that it examines the relationship of both the timing and occurrence of an event of interest rather than focusing only on the occurrence of the event. In this way, survival analysis provides more insights by incorporating time into the research designs; knowing when an event occurs provides more information than simply knowing that the event occurred (89). For example in the study of one year mortality of stroke patients, more information is obtained from examining “how long does the patient survive after initial hospitalisation within the first year?” using survival analysis methods, than examining “what proportion of patients survived one year after initial hospitalisation?” using logistic regression. In answering the second question using logistic regression, the researcher is assuming that a patient who died after one week will have an equal weighting in the analysis as those who died after 51 weeks.
Virnig and colleagues (89) provided a good introduction of the application of survival analysis in analysing administrative data. They stated “the key to well-conducted survival analysis is to give careful thought to definitions of outcomes, start times, end times, censoring events, and covariates”. All these have been carefully considered in all the four studies of this thesis. After reading the four studies and the discussion/conclusion chapter of this thesis, the readers will have a deeper understanding and greater appreciation of survival analysis.

1.9 Outline of the thesis

Using survival analysis in a novel way, the four studies in this PhD research expanded on previous studies, investigating the long-term health outcomes of Indigenous and non-Indigenous Australians suffering from four serious conditions: sepsis, stroke, cancer and rheumatic heart disease.

Severe sepsis is the most common cause of death in ICU. It is an acute and critical illness that causes several organ dysfunctions. Patients who survive the acute sepsis episode have excess mortality that persists beyond their ICU and hospital stay, particularly for patients with existing comorbidities. A previous study has found that although Indigenous sepsis patients have higher comorbidities, there was no significant difference in 28-day mortality rates between Indigenous and non-Indigenous sepsis patients (28). However, it was not clear if the Indigenous sepsis patients have poorer outcomes in the long-term. Chapter two of the thesis –the sepsis project –investigated long-term outcomes for sepsis patients using relative survival, which is a statistical technique commonly used in oncology, but rarely applied to sepsis outcomes. Relative survival is a statistical technique that does not use cause of death data, but rather uses death from any cause as the end-point and is calculated by dividing the observed survival of the study cohort by the ‘expected’ survival of a comparable group in the general population, estimated using life tables.
Stroke is a leading cause of adult disability and the third leading cause of death in Australia. Previous studies have found that NT Indigenous Australians have higher stroke incidence (90) and lower stroke survival (91), but there is limited information on how comorbidities contribute to their worse stroke outcome. Expanding on previous studies (86,87), chapter three—the stroke project—investigated the incidence of recurrent stroke, the prevalence of comorbidities, and the effect of comorbidities on the Indigenous to non-Indigenous differentials in stroke case fatality, stroke recurrence, and long-term survival. To our knowledge, the stroke study is the first study to use competing risk analysis to investigate the relationship between recurrent stroke and comorbidities of Indigenous Australians. Competing risk provides a useful alternative to Cox regression in the presence of competing risks events; it is relevant to the study of recurrent stroke in which death (a competing risk event) might occur before recurrent stroke (event of interest). The competing risk analysis approach is particularly suitable for the analysis of stroke recurrence in Indigenous stroke patients because they have higher mortality from other causes that may bias estimation of their incidence of stroke recurrence. In this study, the Fine-Gray regression is used for the competing risk survival analysis. Analogous to the Cox model (92), the Fine-Gray (93) model is a popular method to estimate the effects of covariates on different competing events, based on regression of the cumulative incidence function.

Cancer is the second leading cause of Indigenous deaths (19.8%) after cardiovascular disease (25.5%). Traditionally in cancer surveillance, the most common measured outcome is net survival, which ignores death from other causes (‘competing risks’). Net survival can be calculated using either cause-specific survival or relative survival methods. A national study of cancer survival for Indigenous patients (94) compared cause-specific with relative survival analysis methods and found that they produced similar results for analysis of time trends and regional variation in cancer survival for all cancers combined and some specific cancer sites. However, there are no studies that compare the different survival analysis methods when considering competing risks of death. This is particularly important for Indigenous cancer patients because they have a higher chance of death from competing causes due to the Indigenous population’s high chronic disease prevalence and mortality from other diseases. Expanding on a previous study (94), chapter four—the cancer project—compared survival
analysis methods that accounted for or ignored competing risks, used life tables and cause of death 
data (for more details, please refer to Table 7.1 in the appendix that summarises the four different 
survival analysis methods used in chapter four). One novel aspect of this study was the use of Cronin-
Feuer method to calculate cancer and non-cancer death probabilities using life tables, which is 
analogous to relative survival.

ARF is an infectious disease that is more common in children (9, 10), particularly Indigenous 
Australian children in the 5-14 age group (10, 11). If left untreated, recurring episodes of ARF cause 
damage to the heart valves, resulting in RHD that can be fatal in childhood or early adulthood (13). 
Between 1997 and 2010, Indigenous people accounted for 92.8% of new RHD cases in the NT (29). 
Indigenous Australians living in NT were 54.8 times more likely to die of RHD than non-Indigenous 
people, and had higher RHD mortality rate than Indigenous populations elsewhere in Australia (95). 
A previous study (29) used the NT RHD Register to investigate the long-term outcomes of ARF and 
RHD patients (ARF recurrence, progression from ARF to RHD, or development of cardiac failure). 
However, the effect of comorbidities and geographical variation on long-term outcomes of ARF and 
RHD patients was not clear, and there have been no reports of studies on serious RHD complications 
(heart failure, stroke, atrial fibrillation and endocarditis) among Indigenous Australians. Expanding on 
a previous study (29), chapter five of this thesis –the RHD project– used a data linkage approach 
(RHD register, hospital inpatient and death register data) to investigate the ARF and RHD patients’ 
adverse outcomes (e.g. RHD complications and mortality) and the effect of comorbidities and 
geographical variation on these adverse outcomes. Presenting the occurrence of these adverse 
outcomes in terms of cumulative incidence and incidence rate over time since diagnosis is a novel 
way in which survival analysis is used to identify the conditions that were more serious different time 
periods.
A summary overview of the outcomes of interest, data sources and the specific survival analysis methods for the four PhD projects is listed in Table 1.1.

Table 1.1 Summary overview of the outcomes of interest, data sources and the specific survival analysis methods for the four PhD projects

<table>
<thead>
<tr>
<th>Study</th>
<th>Outcome</th>
<th>Data source</th>
<th>Survival Analysis methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepsis</td>
<td>1. Death</td>
<td>1. Prospective study data 2. Death Register</td>
<td>1. Relative Survival</td>
</tr>
</tbody>
</table>

1. Deaths of stroke patients were identified from hospital inpatient data (see page 59).
52. Garratt E, Barnes, H. and Dibben, C. Health administrative data: Exploring the potential for academic research., St Andrews: Administrative Data Liaison Service 2010.
75. Whop L, Cunningham J, Condon J. How well is the National Cervical Screening Program performing for Indigenous Australian women? Why we don't really know, and what we can and should do about it. European journal of cancer care 2014;23(6):716-20.


CHAPTER TWO: LONG TERM OUTCOMES FOLLOWING HOSPITAL ADMISSION FOR SEPSIS USING RELATIVE SURVIVAL ANALYSIS: A PROSPECTIVE COHORT STUDY OF 1,092 PATIENTS WITH 5 YEAR FOLLOW UP

2.1 Preface

This chapter contains the first of the four articles that contribute to this thesis. It investigated the long-term outcomes of a prospectively recruited cohort of patients treated for sepsis at Royal Darwin Hospital in 2007-2008, including those admitted to ICU and non-ICU wards. The primary aim was to estimate the duration of the excess mortality risk in patients with sepsis over five years, following their episode of sepsis, using relative survival analysis.

The overall survival was calculated using the Kaplan-Meier method. The relative survival was calculated by dividing the observed survival of the study cohort by the ‘expected’ survival of a comparable group in the general population and estimated using life tables. The results of relative survival were expressed in several ways: cumulative relative survival, interval-specific relative survival and the excess mortality rate. A generalized linear model with Poisson error structure was used for the multivariable analysis. Relative survival was used to investigate the long-term survival of a cohort of sepsis patients because data for a comparison group of people without sepsis was not available, therefore relative survival analysis using general population probability of death was a novel approach in this situation.

Another novel aspect of this study was the use of a cohort of sepsis patients from a previous study (conducted in 2007-2008) and data linkage to administrative data in 2012 to obtain information about deaths - an efficient utilisation of data from a previous study than that enabled examination of long-term outcomes based on clinical data.
2.2 Article

This article has been published in *Plos One*, with Dr Joshua Davis and Vincent He as first co-authors. Davis JS, He V, Anstey NM, Condon JR (2014). Long Term Outcomes Following Hospital Admission for Sepsis Using Relative Survival Analysis: A Prospective Cohort Study of 1,092 Patients with 5 Year Follow Up. *PloS one*, 9(12), e112224.

Dr Joshua Davis (an infectious diseases specialist at Menzies School of Health Research (MSHR) and John Hunter Hospital) and Professor Nick Anstey (an infectious diseases specialist at Royal Darwin Hospital (RDH) and head of the Global Health Division at the MSHR) conceived and designed the study. The data management and analysis were performed by Dr Davis (table 1 and figure 1) and Vincent He (all the other tables and figures). All four co-authors contributed to the manuscript.
RESEARCH ARTICLE

Long Term Outcomes Following Hospital Admission for Sepsis Using Relative Survival Analysis: A Prospective Cohort Study of 1,092 Patients with 5 Year Follow Up

Joshua S. Davis¹,²*, Vincent He¹*, Nicholas M. Anstey¹,², John R. Condon¹

¹. Menzies School of Health Research, Charles Darwin University, Darwin, Australia, ². Department of Infectious Diseases, Royal Darwin Hospital, Darwin, Australia

*Joshua.Davis@menzies.edu.au

These authors contributed equally to this work.

Abstract

Background: Sepsis is a leading cause of death in intensive care units and is increasing in incidence. Current trials of novel therapeutic approaches for sepsis focus on 28-day mortality as the primary outcome measure, but excess mortality may extend well beyond this time period.

Methods: We used relative survival analysis to examine excess mortality in a cohort of 1,028 patients admitted to a tertiary referral hospital with sepsis during 2007–2008, over the first 5 years of follow up. Expected survival was estimated using the Ederer II method, using Australian life tables as the reference population. Cumulative and interval specific relative survival were estimated by age group, sex, sepsis severity and Indigenous status.

Results: Patients were followed for a median of 4.5 years (range 0–5.2). Of the 1028 patients, the mean age was 46.9 years, 52% were male, 228 (22.2%) had severe sepsis and 218 (21%) died during the follow up period. Mortality based on cumulative relative survival exceeded that of the reference population for the first 2 years post admission in the whole cohort and for the first 3 years in the subgroup with severe sepsis. Independent predictors of mortality over the whole follow up period were male sex, Indigenous Australian ethnicity, older age, higher Charlson Comorbidity Index, and sepsis-related organ dysfunction at presentation.

Conclusions: The mortality rate of patients hospitalised with sepsis exceeds that of the general population until 2 years post admission. Efforts to improve outcomes...
from sepsis should examine longer term outcomes than the traditional primary endpoints of 28-day and 90-day mortality.

Introduction

Severe sepsis is the most common cause of death in intensive care units [1], and its incidence has progressively increased over the past 20 years [2, 3]. Until recently, most clinical trials of new therapeutic approaches for severe sepsis have used 28-day mortality as their primary endpoint [4–6]. However, it is now increasingly recognised that the sequelae of sepsis extend well beyond the index hospitalisation and that longer-term outcomes should be used in order to better understand the effect of a given intervention [7, 8]. Despite this increased recent interest, it remains unclear how long the excess mortality risk persists after an episode of severe sepsis, with estimates ranging from 90 days to 5 years [8–13].

Most existing studies investigating longer-term outcomes of sepsis are difficult to extrapolate for several reasons. The majority of sepsis outcome studies include only patients admitted to an intensive care unit (ICU), but only 50–70% of patients hospitalised with severe sepsis ever enter an ICU [14, 15]. Most of these studies use retrospective analysis of existing large datasets, potentially underestimating the true incidence of sepsis [16]. Most importantly, with rare exceptions [13], these studies generally do not compare sepsis outcomes with those of an appropriately matched general population.

Relative survival analysis is a statistical technique commonly used in oncology [17–19], but it has very rarely been applied to sepsis outcomes [20]. It compares the survival of a cohort of patients over time with that of a background reference population.

We aimed to describe the long term outcomes of a prospectively recruited cohort of patients with sepsis, including those admitted to ICU and non-ICU wards. Our primary aim was to estimate the duration of the excess mortality risk in patients with sepsis over the first 5 years of follow up, using relative survival analysis.

Methods

Patients and Setting

The patients included in this cohort have been previously described in detail [21]. In brief, we prospectively enrolled every patient admitted over a 365 day period in 2007–2008 who met the 1992 ACCP-SCCM criteria for sepsis [22], in a tertiary referral hospital in tropical Australia.

Where patients were admitted with more than one episode of sepsis over the 12-month course of the original study, only the first episode was included in the
current analysis. Furthermore, all patients who were resident outside the Northern Territory in the year of initial admission (n=62) were excluded from the current analysis, because their vital status was difficult to determine accurately in the longer term.

Ethics approval
This study was approved by the Human Research Ethics Committee of the Menzies School of Health Research and Northern Territory Department of Health, who waived the requirement for individual informed consent.

Definitions
Severe sepsis was defined as sepsis plus at least one attributable organ dysfunction within the previous 24 hours, as per the definitions used in the PROWESS study [4]. Comorbidities were as defined by Charlson et al and quantified using the Charlson Comorbidity Index [23].

Outcomes
Information about all deaths that occur in the Northern Territory is provided to the Northern Territory Department of Health by the Registrar of Births, Deaths and Marriages and recorded in the public hospitals’ client administration system. We accessed this deaths information to determine vital status for all patients included in this study at 12 months, 3 years and 5 years after the commencement of the study.

Survival analysis
Patients were included in the survival model until death from any cause, censoring due to loss to follow up, or the end of the follow-up period (30/06/2012). Mortality of sepsis patients was measured as both overall survival and relative survival. Overall survival (also referred to as crude survival) is the proportion of sepsis patients still alive at a certain point in time after their sepsis episode; deaths include those unrelated to the sepsis episode, so overall survival does not measure excess mortality related specifically to the sepsis episode. Overall survival analysis was calculated using the Kaplan-Meier method.

By comparing deaths among the sepsis patients with the expected number of deaths based on general population mortality rates, relative survival is an estimate of excess mortality in the sepsis patients. To estimate relative survival, the background mortality rate for the general population was derived from life tables. Expected survival was estimated using the Ederer II method [24] from Australian population life tables for non-Indigenous people and life tables for the total NT Indigenous population for Indigenous people, stratified by age, sex and calendar period. At the time of analysis, life table data were only available up until 2006, therefore, expected survival for the study period were based on 2006 data. The
results of relative survival analysis were expressed in several ways: one-year and five-year relative survival; interval-specific relative survival; and the excess mortality rate. Cumulative relative survival is the observed survival among the subjects under study at a point in time after their first sepsis episode divided by the expected survival of people of the same age, sex and Indigenous status in the general population; for example, five-year relative survival is the cumulative relative survival at five years after first sepsis episode [24]. In general, cumulative relative survival progressively decreases over time since the first episode until there is no more attributable excess mortality of the condition of interest (relative to the background population), after which it plateaus at a constant level. Interval-specific relative survival is the ratio of the survival in the population of interest relative to the general population during specific time intervals. Interval-specific relative survival is lower than 1.0 when the survival of the study cohort is lower than expected (based on the mortality rates of the general population) and plateaus at 1.0 when the survival of the cohort becomes the same as the background population. The excess mortality rate is the difference between the observed mortality rate in the subjects and the expected rate based on that of the general population (matched for age, sex and Indigenous status) [25].

Statistical analysis
All analyses were conducted using Stata/SE 12 (StataCorp, College Station, TX). Continuous variables were compared using t-test for normally distributed data and Mann-Whitney U test for non-normal data. Categorical variables were compared using χ² test. Crude survival rates were derived using Kaplan-Meier estimates. Interval-specific relative survival ratio (RSR) and cumulative RSR were derived using the Stata user-defined program “strs” [26] and excess mortality rates were derived using flexible parametric models, fitted using the Stata user-defined program “stpm2” [25].

A generalized linear model with Poisson error structure was used to model the excess mortality based on collapsed data, in which the outcome was the observed number of deaths and was assumed to be Poisson distributed. The predictors of excess mortality were chosen using forward selection, resulting in the final model with the following covariates: follow-up time since diagnosis, gender, Indigenous status, age group, Charlson index group, severe sepsis and bacteraemic.

Results
Patient demographics and outcomes
Patients were followed for a median of 4.5 years (range 0–5.2 years), giving a cumulative time at risk of 3,997 patient-years. Of the 1,028 patients with sepsis, 218 had died by the end of the follow-up period. Severe sepsis was present in 228 patients (22% - Table 1). Those who died during the follow up period were older, more likely to be male and Indigenous, had greater sepsis severity at baseline and a
larger number of co-morbidities. Pneumonia was over-represented amongst those who died, and skin and soft tissue infection was under-represented (Table 1).

**Crude survival analysis**

One-year and five-year crude survival for all sepsis patients (with 95% confidence intervals) were 87.5% (85.3%–89.3%), and 77.6% (74.5%–80.4%) respectively. The corresponding figures for the severe sepsis subgroup were 74.1% (67.9%–79.3%) and 66.2% (59.7%–72.0%).

### Table 1. Demographics, comorbidities, characteristics of infection and disease severity in those who died compared with those who were alive at the end of follow up.

<table>
<thead>
<tr>
<th>Predisposition</th>
<th>Total (n = 1,028)</th>
<th>Died (n = 218)</th>
<th>Alive (n = 810)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td>46.9 (17.3)</td>
<td>58.9 (15.7)</td>
<td>43.6 (16.2)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Male</td>
<td>532 (51.8%)</td>
<td>128 (58.7%)</td>
<td>404 (49.9%)</td>
<td>0.02</td>
</tr>
<tr>
<td>Indigenous</td>
<td>510 (49.6%)</td>
<td>128 (58.7%)</td>
<td>382 (47.2%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hazardous alcohol use&lt;sup&gt;c&lt;/sup&gt;</td>
<td>284/636 (44.7%)</td>
<td>70/133 (52.6%)</td>
<td>214/503 (42.5%)</td>
<td>0.04</td>
</tr>
<tr>
<td>Current smoking&lt;sup&gt;c&lt;/sup&gt;</td>
<td>329/694 (47.4%)</td>
<td>69/139 (49.6%)</td>
<td>296/555 (53.3%)</td>
<td>0.44</td>
</tr>
<tr>
<td>Chronic renal disease</td>
<td>122 (11.9%)</td>
<td>54 (24.8%)</td>
<td>68 (8.4%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Chronic liver disease</td>
<td>79 (7.7%)</td>
<td>32 (14.7%)</td>
<td>47 (5.8%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes</td>
<td>250 (24.3%)</td>
<td>81 (37.2%)</td>
<td>169 (20.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>COPD</td>
<td>135 (13.1%)</td>
<td>54 (24.8%)</td>
<td>81 (10.0%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Malignancy</td>
<td>44 (4.3%)</td>
<td>22 (10.1%)</td>
<td>22 (2.7%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Charlson comorbidity index&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1 (0–2)</td>
<td>3 (1–4)</td>
<td>0 (0–1)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Infection</th>
<th>Total (n = 1,028)</th>
<th>Died (n = 218)</th>
<th>Alive (n = 810)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteraemic</td>
<td>163 (15.8%)</td>
<td>61 (28.0%)</td>
<td>102 (12.6%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>317 (30.8%)</td>
<td>96 (44.0%)</td>
<td>221 (27.3%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Urosepsis</td>
<td>122 (11.9%)</td>
<td>34 (15.6%)</td>
<td>88 (10.9%)</td>
<td>0.06</td>
</tr>
<tr>
<td>Skin and soft tissue infection</td>
<td>349 (34.0%)</td>
<td>43 (19.7%)</td>
<td>306 (37.8%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Intra-abdominal infection</td>
<td>109 (10.6%)</td>
<td>17 (7.8%)</td>
<td>92 (11.4%)</td>
<td>0.13</td>
</tr>
<tr>
<td>Gram positive bacterium</td>
<td>255 (24.8%)</td>
<td>50 (22.9%)</td>
<td>205 (25.3%)</td>
<td>0.47</td>
</tr>
<tr>
<td>Gram negative bacterium</td>
<td>205 (19.9%)</td>
<td>50 (22.9%)</td>
<td>155 (19.1%)</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response and Organ Dysfunction</th>
<th>Total (n = 1,028)</th>
<th>Died (n = 218)</th>
<th>Alive (n = 810)</th>
<th>P value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Required ICU admission</td>
<td>161 (15.7%)</td>
<td>62 (28.4%)</td>
<td>99 (12.2%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Severe sepsis</td>
<td>228 (22.2%)</td>
<td>77 (35.3%)</td>
<td>151 (18.6%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>APACHE II score&lt;sup&gt;d&lt;/sup&gt;</td>
<td>8 (4–13)</td>
<td>14 (9–20)</td>
<td>6 (3–10)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>SOFA score&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1 (0–3)</td>
<td>3 (1–5)</td>
<td>1 (0–2)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Data are given as n(%) unless stated otherwise.

Parameters are grouped according to PIRO classification [39].

<sup>a</sup>P-value comparing those alive at end of follow up period compared with those who were not.

<sup>b</sup>Mean (SD).

<sup>c</sup>Data were not available for all patients regarding alcohol use and smoking.

<sup>d</sup>Median (IQR).

doi:10.1371/journal.pone.0112224.t001
Fig. 1. a. Kaplan-Meier crude survival estimates by age group for all sepsis patients (n=1,028). b. Kaplan-Meier crude survival estimates by age group for severe sepsis patients (n=228).

doi:10.1371/journal.pone.0112224.g001
Kaplan-Meier estimates by age group are shown in Fig. 1, in which people aged over 64 years had the lowest crude survival and people under 45 years old had the highest.

Relative survival analysis

Figs. 2a and 2b show cumulative relative survival by age group for the overall cohort and those with severe sepsis. For the whole cohort, the curve plateaus after approximately 2 years (Fig. 2a), and for those with severe sepsis, it plateaus at 3 years (Fig. 2b). This indicates that the excess mortality persists for 2 years in sepsis patients, and 3 years in those with severe sepsis, after which sepsis patients have a mortality rate similar to that of the general population. The relative survival of sepsis patients was worse in older age groups, and this difference was most marked in the first 6 months after presentation (Figs. 3a and 3b).

Interval-specific relative survival (Tables 2 and 3) lead to a similar conclusion – survival of the sepsis cohort is lower than that of the general population in each period until approximately 2 years after diagnosis, and that of the severe sepsis subgroup until 3 years.

Across all subgroups except males greater than 64 years old, Indigenous patients had lower five-year relative survival than non-Indigenous patients (Table 4), with this difference being most evident in young female patients. Across all age groups, females had higher five-year relative survival than males.

Multivariable analysis of predictors of mortality

In multivariable regression analysis, (Table 5) excess mortality was highest in the first year and decreased over time after the sepsis episode, dropping to almost zero (relative to the first year) in the fourth and fifth years of follow-up. Other independent predictors of excess mortality were male sex, Indigenous ethnicity, older age, more comorbidities and sepsis-related organ dysfunction at presentation.

Discussion

In this cohort of patients hospitalised with sepsis, and including approximately 4,000 patient years of follow up, the excess mortality persisted for 2 years overall and for 3 years in the subgroup with severe sepsis.

Although there is increasing interest in longer term outcomes following sepsis [8–11], the use of relative survival analysis is rare in the sepsis literature. Our study differs from the only previous publication to use relative survival analysis in a sepsis cohort [20] in several ways. In contrast to the prospective design of our study, the study by Ghelani was retrospective, with sepsis defined using ICD-9 coding, which decreases both the sensitivity and specificity of the diagnosis of sepsis compared with prospective methods. Ghelani et al found that the cumulative relative survival in the ICU sepsis cohort continued to decline over the
Fig. 2. a. Cumulative relative survival for patients with sepsis and severe sepsis, compared with a reference population. Dotted lines represent 95% confidence intervals. b. Cumulative relative survival for only patients with severe sepsis (n=228), compared with a reference population. Dotted lines represent 95% confidence intervals.

doi:10.1371/journal.pone.0112224.g002
entire follow up period (range 4.2–9.6 years), never reaching a plateau. This is in contrast to our finding of a plateau occurring after 2 years, and is difficult to explain. Ghelani’s finding implies that the excess risk of death following a sepsis episode never abates, suggesting that the characteristics of the population are responsible rather than the insults of the sepsis episode.

Fig. 3. a. Cumulative relative survival by age group for all sepsis patients (n=1,028). b. Cumulative relative survival by age group for severe sepsis patients (n=228).

doi:10.1371/journal.pone.0112224.g003
Although we have demonstrated excess mortality in the sepsis cohort compared to the general population, we cannot determine the mechanism of mortality, as we did not have access to cause of death data. Acute sepsis-related organ failure generally resolves by the time of hospital discharge in those who survive. However, systemic inflammatory activation persists for weeks-months afterwards [27], as does functional immunosuppression [28, 29]. Systemic inflammation is associated with endothelial cell activation and dysfunction and an increase in endogenous nitric oxide inhibitors [30–33, 34, 35] which are turn associated with acute vascular events [36, 37] and increased early mortality in sepsis [38]. Hence delayed secondary infections and vascular events are plausible potential explanations for this observed excess mortality. An alternative explanation is that patients hospitalised with sepsis may differ from the background population in important comorbidities such as diabetes mellitus, excess alcohol use and chronic renal disease. Our data do not provide evidence to distinguish between these two

Table 2. Interval-specific relative survival by age category (all sepsis patients).

<table>
<thead>
<tr>
<th>Time period (years of follow up)</th>
<th>Overall Age 15–44</th>
<th>Age 45–64</th>
<th>Age &gt;=65</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–0.5</td>
<td>0.90 (0.88–0.92)</td>
<td>0.95 (0.93–0.97)</td>
<td>0.90 (0.87–0.93)</td>
</tr>
<tr>
<td>0.5–1.0</td>
<td>0.98 (0.97–0.99)</td>
<td>0.99 (0.98–1.00)</td>
<td>0.97 (0.97–0.99)</td>
</tr>
<tr>
<td>1.0–1.5</td>
<td>0.97 (0.96–0.98)</td>
<td>0.99 (0.97–1.00)</td>
<td>0.97 (0.95–0.99)</td>
</tr>
<tr>
<td>1.5–2.0</td>
<td>0.99 (0.98–1.00)</td>
<td>1.00 (0.98–1.00)</td>
<td>0.99 (0.97–1.00)</td>
</tr>
<tr>
<td>2.0–2.5</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00 (1.00–1.00)</td>
<td>0.99 (0.97–1.00)</td>
</tr>
<tr>
<td>2.5–3.0</td>
<td>0.99 (0.98–1.00)</td>
<td>1.00 (1.00–1.00)</td>
<td>0.99 (0.96–1.00)</td>
</tr>
<tr>
<td>3.0–3.5</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00 (1.00–1.00)</td>
<td>1.00 (0.98–1.00)</td>
</tr>
<tr>
<td>3.5–4.0</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00 (0.98–1.00)</td>
<td>1.00 (0.98–1.00)</td>
</tr>
<tr>
<td>4.0–4.5</td>
<td>1.00 (0.99–1.00)</td>
<td>1.00 (1.00–1.00)</td>
<td>0.99 (0.97–1.00)</td>
</tr>
<tr>
<td>4.5–5.0</td>
<td>1.00 (0.98–1.00)</td>
<td>0.99 (0.96–1.00)</td>
<td>1.00 (0.95–1.01)</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0112224.t002

Table 3. Interval specific relative survival by age category (severe sepsis patients).

<table>
<thead>
<tr>
<th>Time period (years of follow up)</th>
<th>Overall Age 15–44</th>
<th>Age 45–64</th>
<th>Age &gt;=65</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–0.5</td>
<td>0.77 (0.70–0.82)</td>
<td>0.87 (0.78–0.93)</td>
<td>0.76 (0.66–0.84)</td>
</tr>
<tr>
<td>0.5–1.0</td>
<td>0.98 (0.95–1.00)</td>
<td>0.99 (0.91–1.00)</td>
<td>0.98 (0.90–1.00)</td>
</tr>
<tr>
<td>1.0–1.5</td>
<td>0.97 (0.92–0.99)</td>
<td>0.99 (0.91–1.00)</td>
<td>0.95 (0.86–0.98)</td>
</tr>
<tr>
<td>1.5–2.0</td>
<td>0.98 (0.94–1.00)</td>
<td>0.99 (0.91–1.00)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
<tr>
<td>2.0–2.5</td>
<td>0.99 (0.96–1.00)</td>
<td>1.00 (1.00–1.00)</td>
<td>0.99 (0.90–1.00)</td>
</tr>
<tr>
<td>2.5–3.0</td>
<td>0.99 (0.95–1.00)</td>
<td>1.00 (1.00–1.00)</td>
<td>0.96 (0.87–0.99)</td>
</tr>
<tr>
<td>3.0–3.5</td>
<td>1.00 (0.96–1.01)</td>
<td>1.00 (1.00–1.00)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
<tr>
<td>3.5–4.0</td>
<td>1.00 (0.96–1.01)</td>
<td>0.99 (0.91–1.00)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
<tr>
<td>4.0–4.5</td>
<td>1.01 (1.01–1.01)</td>
<td>1.00 (1.00–1.00)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
<tr>
<td>4.5–5.0</td>
<td>1.01 (1.01–1.01)</td>
<td>1.00 (1.00–1.00)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0112224.t003
explanations because the background population life tables used in the relative survival analysis were not stratified by co-morbidity. However, a previous large study by Quartin et al. comparing sepsis patients with a control hospitalised population found that the excess mortality in the sepsis group lasted for at least 5 years, and persisted even after adjustment for comorbidities [13]. This suggests either that the mortality excess which persists beyond the acute period is related to the sepsis episode itself rather than to underlying comorbidities, or that unmeasured comorbidities or characteristics are responsible for excess mortality, as suggested by Ghelani’s data [20]. Neither our study nor previously published work can definitively determine whether excess mortality in sepsis patients is due to sepsis itself, underlying comorbidities, or a combination of the two.

Our finding that excess mortality persists for at least 2–3 years supports calls to extend the traditional end-points of interventional sepsis studies beyond the most commonly used end-point 28-day mortality [7]. Further work is needed to identify the aetiology of the excess mortality following hospital admission for sepsis, particularly that which occurs more than 1 year post discharge.

### Table 4. Five-year relative survival (with 95% confidence interval) by age group, sex and Indigenous status.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age-group</th>
<th>Non-Indigenous</th>
<th>Indigenous</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>15–44 years</td>
<td>0.98 (0.92–1.00)</td>
<td>0.89 (0.78–0.96)</td>
</tr>
<tr>
<td></td>
<td>45–64 years</td>
<td>0.90 (0.82–0.95)</td>
<td>0.75 (0.63–0.85)</td>
</tr>
<tr>
<td></td>
<td>&gt;=65 years</td>
<td>0.55 (0.42–0.67)</td>
<td>0.57 (0.26–0.89)</td>
</tr>
<tr>
<td>Female</td>
<td>15–44 years</td>
<td>1.00 (1.00–1.00)</td>
<td>0.86 (0.77–0.92)</td>
</tr>
<tr>
<td></td>
<td>45–64 years</td>
<td>0.93 (0.83–0.98)</td>
<td>0.74 (0.63–0.83)</td>
</tr>
<tr>
<td></td>
<td>&gt;=65 years</td>
<td>0.77 (0.54–0.94)</td>
<td>0.72 (0.48–0.92)</td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0112224.t004

### Table 5. Multivariable analysis of predictors of excess mortality using Poisson regression.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Adjusted hazard ratio</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of follow-up: 2</td>
<td>0.35</td>
<td>0.22, 0.54</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Year of follow-up: 3</td>
<td>0.12</td>
<td>0.05, 0.29</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Year of follow-up: 4</td>
<td>0.04</td>
<td>0.00, 0.25</td>
<td>0.001</td>
</tr>
<tr>
<td>Year of follow-up: 5&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.04</td>
<td>0.00, 0.65</td>
<td>0.02</td>
</tr>
<tr>
<td>Male&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.58</td>
<td>1.10, 2.27</td>
<td>0.01</td>
</tr>
<tr>
<td>Indigenous&lt;sup&gt;c&lt;/sup&gt;</td>
<td>1.56</td>
<td>1.05, 2.31</td>
<td>0.03</td>
</tr>
<tr>
<td>Age group: 45–64&lt;sup&gt;d&lt;/sup&gt;</td>
<td>1.68</td>
<td>1.04, 2.73</td>
<td>0.03</td>
</tr>
<tr>
<td>Age group: &gt;=65&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3.40</td>
<td>2.02, 5.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Charlson index 1–2&lt;sup&gt;e&lt;/sup&gt;</td>
<td>3.82</td>
<td>1.90, 7.69</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Charlson index &gt;=3&lt;sup&gt;e&lt;/sup&gt;</td>
<td>11.98</td>
<td>6.05, 23.74</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Severe Sepsis&lt;sup&gt;f&lt;/sup&gt;</td>
<td>2.30</td>
<td>1.61, 3.27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Bacteraemic&lt;sup&gt;g&lt;/sup&gt;</td>
<td>1.67</td>
<td>1.15, 2.42</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Reference categories – a. First year of follow up; b. Female. c. Non-Indigenous d. Age group 15-44 years. e. Charlson comorbidity index=0; f. Non-severe sepsis patients; g. Non-bacteraemic patients.

doi:10.1371/journal.pone.0112224.t005
Our study has some limitations. We studied a cohort of patients from tropical Australia of whom a substantial proportion were Indigenous Australians, and with a relatively young mean age, and thus the applicability of our findings to other populations is unclear. However, the fact that the post-sepsis mortality excess has been shown to persist for well beyond 90 days in several other cohorts support our findings. We used life tables from 2006 (the most recent available robust data), but the period under study was 2007–2012. If the mortality rate in the background population had decreased over this time, our models may have overestimated the excess mortality in the sepsis population. However, given the short time between these two periods, this is unlikely to have been significant. We did not have access to data on cause of death, nor on comorbidities in the reference population. However, a strength of our study is that, unlike the majority of sepsis studies investigating longer term outcomes post-sepsis [9–12], we compared the survival of a sepsis cohort with that of an appropriate reference population.

Conclusions
In conclusion, the risk of mortality from an acute episode of sepsis extends well beyond hospital discharge. Relative survival analysis is a useful tool for examining the excess mortality in patients with sepsis. Future efforts to improve outcomes from sepsis should focus on endpoints as long as 3 years post discharge, and such endpoints should no longer be limited to the first 90 days following hospital admission.

Acknowledgments
We would like to thank Mark McMillan and Alex Humphrey for help with data collection and data entry; and from the Health Gains Planning Branch of the NT Government: Yuejen Zhao for providing mortality data and Xiaohua Zhang for providing the reference population life tables.

Author Contributions
Conceived and designed the experiments: JSD NMA. Performed the experiments: JSD VH. Analyzed the data: JSD VH. Contributed reagents/materials/analysis tools: JRC. Wrote the paper: JSD VH NMA JRC.

References


CHAPTER THREE: ADVERSE OUTCOME AFTER INCIDENT STROKE
HOSPITALIZATION FOR INDIGENOUS AND NON-INDIGENOUS AUSTRALIANS IN THE NORTHERN TERRITORY

3.1 Preface
The previous chapter demonstrated the importance of measuring survival for up to three years after acute sepsis, rather than only for the short periods (28 or 90 days) commonly used in clinical studies of sepsis treatment. Survival is not the only important outcome after serious illness; non-fatal but debilitating consequences can also occur after severe, acute illness.

This chapter used the competing risks approach to analyse recurrent stroke and long-term deaths of NT Indigenous and non-Indigenous stroke patients. The conventional method in estimating recurrent stroke probabilities is to ignore the competing risks (death) and treat death as a censoring event. However, this might not be appropriate for serious illnesses such as stroke where the probability of death is much higher than the primary outcome of interest (i.e. recurrent stroke); people might die before experiencing the recurrent event. The competing risk approach deals with this issue. The competing risks analysis was used because people who have suffered a stroke are at higher risk of both recurrent stroke and death. Competing risk analysis enables the factors associated with risk of recurrent stroke to be distinguished from those associated with higher risk of death. Presenting the probability of recurrent stroke, death and survival as stacked cumulative incidence plots is a novel way to show how the probabilities change over time for different age groups, which differs between Indigenous and non-Indigenous patients, that may be a more understandable way to communicate risk to patients and clinicians.
3.2 Article

This chapter has been published in the *International Journal of Stroke*.


Vincent He conceived and designed the experiment, performed the data management and statistical analysis, conducted the literature review and wrote the manuscript draft. Jiqiong You (health economist at NT DoH Health Gains Planning Branch) obtained the ethics approval and extracted the hospital inpatient data. Dr James N. Burrow (neurologist at Royal Darwin Hospital) provided clinical insights. All the authors (Vincent He, Professor John Condon, Jiqiong You, Dr Yuejen Zhao and Dr James N. Burrow) contributed to the review and revision of the related journal paper.
Adverse outcome after incident stroke hospitalization for Indigenous and non-Indigenous Australians in the Northern Territory

Vincent YF. He1*, John R. Condon1, Jiqiong You2, Yuejen Zhao2, and James N. Burrow3

Background Survival after a stroke is lower for Indigenous than other stroke patients in Australia. It is not known whether recurrence is more common for Indigenous patients, or whether their higher prevalence of comorbidity affects their lower survival.

Aims This study aimed to investigate the stroke recurrence and role of comorbidities in adverse stroke outcomes (recurrence and death) for Indigenous compared with other Australians.

Methods A retrospective cohort study of first hospitalization for stroke (n = 2105) recorded in Northern Territory hospital inpatient data between 1996 and 2011 was conducted. For the multivariable analyses of adverse outcomes, logistic regression was used for case fatality and competing risk analysis for recurrent stroke and long-term death. Comorbidities (identified from inpatient diagnosis data) were analyzed using the Charlson Comorbidity Index (modified for stroke outcomes).

Results Prevalence of comorbidities, case fatality, incidence of re-hospitalization for recurrent stroke, and long-term death rate were higher for Indigenous than non-Indigenous stroke patients. Adjustment for comorbidity in multivariable analyses considerably reduced Indigenous patients’ excess risk for case fatality (odds ratio: 1·25, 0·88–1·78) and long-term death (standard hazard ratio: 1·27, 1·01–1·61) (but not recurrence), implying that their excess risk of death was in part due to higher comorbidity prevalence.

Conclusion Indigenous stroke patients have higher prevalence of comorbidities than non-Indigenous stroke patients, which explained part of the disparity in both case fatality and long-term survival but did not explain the disparity in stroke recurrence at all.

Key words: comorbidity, competing risks, hospital data, Indigenous Australian, recurrent stroke, survival analysis

Introduction

In Australia, stroke is a leading cause of adult disability (1–3) and the third leading cause of death (4). The length of stay in hospital for treatment of stroke is one of the longest, and the costs involved rank as one of the highest, of all reasons for hospitalization (2). Stroke has a greater impact on Indigenous Australians (Aboriginal and Torres Strait Islander peoples) who have higher stroke incidence (5,6), higher hospitalization rate (2), lower survival rate (3), and higher burden of disease due to stroke than non-Indigenous Australians (1,7,8).

However, there is limited information about stroke recurrence for Indigenous stroke patients (9,10). Indigenous stroke patients have higher comorbidities than non-Indigenous stroke patients (1,6,11), but there is limited information on how comorbidities contribute to their worse stroke outcome. Therefore in this study, we aim to investigate the incidence of recurrent stroke, the prevalence of comorbidities, and the effect of comorbidities on the Indigenous to non-Indigenous differentials in stroke case fatality, recurrent stroke, and long-term survival in the Northern Territory (NT), Australia.

Methods

First stroke episodes occurring among NT residents between January 1, 1996 and December 31, 2011 were identified from hospital inpatient data for all five NT public hospitals; the inpatient data contains a summary of every inpatient episode, with a unique client identifier used by all five hospitals. The only private hospital does not treat acute stroke, and no hospitals have a specialized stroke unit. For each episode, the main cause for admission (principle diagnosis) and other related illnesses (additional diagnoses) are coded using the International Classification of Diseases, 9th revision (ICD9) (Jan 1991 to June 1998) and ICD10 (July 1998 onwards). Stroke episodes were identified using ICD9 and ICD10 (in italics) codes: 430–432, I60–I62 (hemorrhagic); 434, I63 (ischemic); and 433, 436, I64 (other strokes). The first stroke (the ‘index episode’) was defined as a hospitalization with a principal diagnosis of stroke with no previous hospitalization (principal or secondary diagnosis) for stroke in the preceding five-years.

Case fatality (death during the index episode) and deaths in subsequent hospitalizations were identified from the hospital inpatient data. Patients who survived their first stroke were followed for recurrent stroke and death until 30 June 2013. Long-term deaths were identified from deaths occurring in subsequent hospitalization episodes. Recurrent stroke was defined as a second hospitalization with a principal diagnosis of stroke. Hospitalizations for stroke within 21 days of the index episode were assumed to be for further treatment or progression of the first stroke (12) and were not considered as recurrent stroke.

Comorbidities were identified from the hospital inpatient dataset using principal and secondary diagnoses for the index...
episode and episodes in the preceding five-years. The modified Charlson Comorbidity Index (13) adapted for stroke (i.e. excluding cerebrovascular disease and hemiplegia) was used in analysis. Other conditions that are risk factors for recurrent strokes also included atrial fibrillation, hypercholesterolemia, hypertension, rheumatic heart disease, and transient ischemic attack.

Indigenous patients were identified from hospital records, which have a high level of accuracy in the NT (96-9% agreement with self-reported Indigenous status) (14). Stroke patients living in the cities of Darwin and Alice Springs and their hinterlands were classified as urban residents; all others were classified as remote residents.

All statistical analyses were conducted using Stata, version 13 (StataCorp, College Station, TX, USA). The chi-squared test was used to compare differences between Indigenous and non-Indigenous patients in demographics characteristics. The Mann–Whitney test was used to compare length of stay of the initial episode with re-hospitalization due to stroke. Logistic regression was used to compare comorbidity prevalence, case fatality, and stroke recurrence between Indigenous and non-Indigenous patients. Multivariable logistic regression was used to identify factors associated with case fatality.

The outcomes of long-term death and recurrent stroke have important competing effects because death might occur before recurrent stroke. Therefore, the Fine–Gray competing risk model (15) was used to estimate the cumulative incident function for stroke recurrence or long-term death. In Fine–Gray multivariable analysis, the standard hazard ratio (SHR) for recurrent stroke is the ratio of recurrent stroke hazard rates with death as a competing risk. The censoring date for the survival analysis was 30 June 2013. The survival time for stroke patients who died on the day of admission was counted as half a day.

The final regression model included: Indigenous status (Indigenous compared with non-Indigenous); age (per year); gender (female compared with male); stroke subtype (ischemic and other, compared with hemorrhagic); year of diagnosis (per year); remoteness of residence (remote compared with urban); presence of atrial fibrillation; presence of hypercholesterolemia; modified CCI (1 and 2 or more compared with 0); and an interaction term for Indigenous status by age at diagnosis (because the effect of age at diagnosis was found to be different for Indigenous compared with non-Indigenous patients).

The study was approved by the Human Research Ethics Committee of the NT Department of Health and the Menzies School of Health Research (HREC-2011-1680).

Results

Between January 1, 1996 and December 31, 2011, 2105 NT residents were admitted to hospital for the first time with a principal diagnosis of stroke; 43-9% were Indigenous and 56-6% male. Indigenous patients were more likely than non-Indigenous to be younger (median age 53 vs. 64), female, reside in remote regions, and have more comorbidities (Table 1). They also had higher case fatality and stroke recurrence and a longer length of stay for their first stroke episode \( (P = 0.03) \) (Table 1). Hemorrhagic stroke was more common for Indigenous patients, while ischemic stroke was more common for non-Indigenous patients.

After adjusting for other demographic (Indigenous status, age, gender, stroke subtype, year of diagnosis, and remoteness of patients’ residence) and clinical variables (atrial fibrillation and hypercholesterolemia), but not comorbidities, Indigenous stroke patients had a higher case fatality \( [\text{odds ratio (OR)}: 1.41, 95\% \text{ CI}: 1.04–1.92] \). Adjustment for comorbidities (Table 2) reduced this disparity \( (\text{OR}: 1.25, 0.88–1.78) \), suggesting that comorbidity is a major contributor to the disparity.

The age-adjusted cumulative incidence of recurrent stroke was higher for Indigenous patients than non-Indigenous patients after one-year \( (7.70\% \text{ vs. } 5.73\%) \) and five-years \( (13.43\% \text{ vs. } 10.08\%) \) (Fig. 1). The risk of recurrent stroke hospitalization was greatest in the first few months following the incident hospitalization, especially for Indigenous patients. Figure 2 shows the stacked cumulative incidence plots of recurrent stroke, death, and survival by age group. For both Indigenous and non-Indigenous stroke patients, the probability of death was higher than stroke recurrent hospitalization, and both were highest in the 65+ age group (five-year CIF of death: non-Indigenous 29.4%, Indigenous 35.8%; five-year CIF of recurrent stroke hospitalization: non-Indigenous 11.6%, Indigenous 15.3%).

After adjusting for demographic and clinical variables, but not comorbidities, Indigenous patients had a higher risk of recurrent stroke \( (\text{SHR}: 1.61, 95\% \text{ CI}: 1.17–2.21) \) and long-term death \( (\text{SHR}: 1.55, 1.24–1.94) \) than non-Indigenous stroke patients. Also adjusted for comorbidities partially reduced Indigenous patients’ excess risk of long-term death \( (\text{SHR}: 1.27, 1.01–1.61) \) but did not reduce their excess risk of recurrence \( (\text{SHR}: 1.82, 1.32–2.51) \) (Table 2).

The effects of age at diagnosis on recurrent stroke were different for Indigenous and non-Indigenous patients. For non-Indigenous patients, recurrent stroke incidence was greater for older than younger patients, increased by 3% per year of age \( (\text{SHR}: 1.03, 95\% \text{ CI}: 1.02–1.04) \) (Table 2), but for Indigenous patients, incidence increased by only 1% per year or age \( (\text{SHR}: 1.01, 95\% \text{ CI}: 1.00–1.02) \).

Discussion

Indigenous stroke patients had higher case fatality, higher recurrence rate, and higher long-term death rate than non-Indigenous patients. The result of this study is consistent with similar studies reported for other populations. For NT Indigenous and non-Indigenous patients combined, 13.5% had a hospitalization for recurrent stroke within five-years, higher than in a Scotland study \( (9.8\%) \) (16) but lower than in a Singapore study \( (15.7\%) \) (17). Indigenous stroke patients had higher recurrent strokes than non-Indigenous stroke patients, consistent with a previous report from Western Australia (9). The higher prevalence of comorbidities in Indigenous stroke patients in the NT is also consistent with another report from a Western Australian study (6).

Comorbidity is a major contributor to the disparities in case fatality and long-term survival (but not stroke recurrence) between Indigenous and non-Indigenous stroke patients.
Table 1: Demographic characteristics of people hospitalized for their first stroke, Northern Territory, 1996–2011, by Indigenous status

<table>
<thead>
<tr>
<th></th>
<th>Non-Indigenous (n = 1180)</th>
<th>Indigenous (n = 925)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Men</td>
<td>729</td>
<td>61.8</td>
<td>462</td>
</tr>
<tr>
<td>Remote residents</td>
<td>205</td>
<td>17.4</td>
<td>597</td>
</tr>
<tr>
<td>Age at onset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤44</td>
<td>143</td>
<td>12.1</td>
<td>258</td>
</tr>
<tr>
<td>45–64</td>
<td>457</td>
<td>38.7</td>
<td>439</td>
</tr>
<tr>
<td>≥65</td>
<td>580</td>
<td>49.2</td>
<td>227</td>
</tr>
<tr>
<td>Stroke subtypes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemorrhagic</td>
<td>364</td>
<td>30.8</td>
<td>340</td>
</tr>
<tr>
<td>Ischemic strokes</td>
<td>468</td>
<td>39.7</td>
<td>308</td>
</tr>
<tr>
<td>Other undefined strokes</td>
<td>348</td>
<td>29.5</td>
<td>277</td>
</tr>
<tr>
<td>History of comorbidities*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute myocardial</td>
<td>66</td>
<td>5.6</td>
<td>62</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>107</td>
<td>9.1</td>
<td>130</td>
</tr>
<tr>
<td>Peripheral vascular disease</td>
<td>81</td>
<td>6.9</td>
<td>41</td>
</tr>
<tr>
<td>Dementia†</td>
<td>63</td>
<td>5.3</td>
<td>141</td>
</tr>
<tr>
<td>COPD</td>
<td>138</td>
<td>11.7</td>
<td>141</td>
</tr>
<tr>
<td>Rheumatoid disease</td>
<td>14</td>
<td>1.2</td>
<td>15</td>
</tr>
<tr>
<td>Peptic ulcer</td>
<td>11</td>
<td>0.9</td>
<td>8</td>
</tr>
<tr>
<td>Mild liver disease</td>
<td>19</td>
<td>1.6</td>
<td>57</td>
</tr>
<tr>
<td>Diabetes</td>
<td>135</td>
<td>11.4</td>
<td>225</td>
</tr>
<tr>
<td>Diabetes with end-organ damage</td>
<td>145</td>
<td>12.3</td>
<td>281</td>
</tr>
<tr>
<td>Renal failure</td>
<td>68</td>
<td>5.8</td>
<td>168</td>
</tr>
<tr>
<td>Cancer‡</td>
<td>61</td>
<td>5.2</td>
<td>21</td>
</tr>
<tr>
<td>Moderate/severe liver disease</td>
<td>5</td>
<td>0.4</td>
<td>22</td>
</tr>
<tr>
<td>Metastatic cancer</td>
<td>21</td>
<td>1.8</td>
<td>8</td>
</tr>
<tr>
<td>Other comorbidities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td>206</td>
<td>17.5</td>
<td>118</td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td>187</td>
<td>15.8</td>
<td>120</td>
</tr>
<tr>
<td>Hypertensive diseases</td>
<td>618</td>
<td>52.4</td>
<td>572</td>
</tr>
<tr>
<td>Rheumatic heart disease</td>
<td>43</td>
<td>3.6</td>
<td>94</td>
</tr>
<tr>
<td>Transient ischemic attack</td>
<td>52</td>
<td>4.4</td>
<td>30</td>
</tr>
<tr>
<td>Modified CCI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>654</td>
<td>55.4</td>
<td>321</td>
</tr>
<tr>
<td>1</td>
<td>247</td>
<td>20.9</td>
<td>239</td>
</tr>
<tr>
<td>≥2</td>
<td>279</td>
<td>23.6</td>
<td>365</td>
</tr>
<tr>
<td>Case fatality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>163</td>
<td>13.8</td>
<td>146</td>
</tr>
<tr>
<td>Hemorrhage§</td>
<td>81</td>
<td>22.3</td>
<td>96</td>
</tr>
<tr>
<td>Ischemic strokes§</td>
<td>49</td>
<td>10.5</td>
<td>26</td>
</tr>
<tr>
<td>Other undefined strokes§</td>
<td>33</td>
<td>9.5</td>
<td>24</td>
</tr>
<tr>
<td>Recurrent stroke hospitalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>147</td>
<td>12.5</td>
<td>137</td>
</tr>
<tr>
<td>Hemorrhage§</td>
<td>32</td>
<td>8.8</td>
<td>35</td>
</tr>
<tr>
<td>Ischemic strokes§</td>
<td>70</td>
<td>15.0</td>
<td>54</td>
</tr>
<tr>
<td>Other undefined strokes§</td>
<td>45</td>
<td>12.9</td>
<td>48</td>
</tr>
<tr>
<td>LOS§ at index stroke</td>
<td>4</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>LOS§ at recurrent stroke</td>
<td>8</td>
<td>9</td>
<td>9</td>
</tr>
</tbody>
</table>

*AIDS is not included in the full list of Modified CCI in this table because there were no cases of AIDS for both Indigenous and non-Indigenous stroke patients in this study. The other five comorbidities were selected based on comprehensive literature review as risk factors for recurrent strokes.

†After age adjustment, Indigenous stroke patients were more likely to have dementia comorbidity [OR: 2.23, 95% CI: (1.41, 3.53)] than non-Indigenous stroke patients.

‡Cancer includes leukemia and lymphoma.

§Denominator for percentage is the number of cases in each stroke type, respectively.

¶Length of stay (median).

COPD, chronic obstructive pulmonary disease.
### Table 2  Regression analysis* for case fatality, recurrent stroke, and overall deaths

<table>
<thead>
<tr>
<th></th>
<th>Case fatality</th>
<th></th>
<th>Recurrent stroke</th>
<th></th>
<th>Long-term death</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OR 95% CI</td>
<td>SHR 95% CI</td>
<td>OR 95% CI</td>
<td>SHR 95% CI</td>
<td>OR 95% CI</td>
<td>SHR 95% CI</td>
</tr>
<tr>
<td>Indigenous status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Indigenous*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>1·25 0·88 1·78</td>
<td>1·82 1·32 2·51</td>
<td>1·27 1·01 1·61</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at diagnosis (per year)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>1·02 1·00 1·03</td>
<td>1·00 0·99 1·01</td>
<td>1·01 1·00 1·02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Indigenous</td>
<td>1·03 1·02 1·05</td>
<td>1·03 1·02 1·04</td>
<td>1·03 1·02 1·04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1·03 0·79 1·33</td>
<td>0·97 0·76 1·23</td>
<td>0·98 0·83 1·15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemorrhage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ischemic</td>
<td>0·24 0·18 0·33</td>
<td>1·60 1·17 2·19</td>
<td>0·45 0·37 0·54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other undefined</td>
<td>0·19 0·13 0·27</td>
<td>1·41 1·01 1·95</td>
<td>0·44 0·36 0·55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year of diagnosis (per year)</td>
<td>0·93 0·91 0·96</td>
<td>0·99 0·96 1·02</td>
<td>0·97 0·95 0·99</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remoteness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remote</td>
<td>1·08 0·79 1·47</td>
<td>1·15 0·86 1·54</td>
<td>1·03 0·85 1·24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrial fibrillation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1·47 1·04 2·08</td>
<td>1·44 1·07 1·95</td>
<td>1·25 1·01 1·55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypercholesterolemia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>0·68 0·45 1·04</td>
<td>1·58 1·15 2·16</td>
<td>0·89 0·70 1·12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modified CCI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥2</td>
<td>2·78 2·02 3·82</td>
<td>0·66 0·49 0·88</td>
<td>2·18 1·77 2·68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Base age 53 and base year 2004.
†Reference categories.
‡The OR (for case fatality) and SHRs for the interaction term for age by Indigenous status in each model was: case fatality 0·985 (95% CI: 0·968–1·003); recurrent stroke 0·974 (95% CI: 0·960–0·988); long-term death 0·986 (95% CI: 0·974–0·998).

**Fig. 1** Age-adjusted cumulative incidence function and hazard function for recurrent stroke after incident stroke hospitalization.
Adjustment for comorbidity in multivariable analyses considerably reduced Indigenous patients’ excess risk for both case fatality and long-term death implying that their excess risk was at least partly due to their higher prevalence of comorbidities, which in turn is related to their poorer social, economic, and environmental circumstances (18), data on which was not available for this study. A similar finding has been previously reported after acute myocardial infarction in the NT (19).

Other than higher comorbidity prevalence, the worse long-term outcomes for Indigenous stroke patients might be attributable to the higher prevalence of stroke risk factors. Smoking prevalence is more than twice as high for NT Indigenous (56·0% in 2012–13) than non-Indigenous adults (24·1%) (20). In 2012–2013, 30·5% of NT Indigenous adults consumed alcohol at ‘risky’ and ‘high risk’ levels (21). In addition, Indigenous stroke patients’ worse long-term outcomes might also be attributable to treatment deficiencies. There is evidence of deficiencies in treatment for Indigenous compared with non-Indigenous stroke patients, such as barriers to accessing treatment (7) and lower adherence to recommended stroke management (1). Indigenous patients have been found to be less likely to receive evidence-based medications for ischemic stroke (11), less likely to receive aspirin within 48 hours of admission and antithrombotic medication after hospital discharge (1).

Recurrent stroke was more common for older than younger non-Indigenous stroke patients, but there was much less difference between age groups for Indigenous patients. This might be due to the high prevalence of stroke risk factors in young Indigenous stroke patients. Among NT Indigenous people, the highest prevalence of smoking is in the age group 25–34 for men (79·3%) and age group 18–24 for women (56·6%) (20). In Western Australia, almost a quarter of Indigenous stroke survivors aged 25–34 years have rheumatic heart disease (6). In addition, young Indigenous stroke patients are more likely to survive and experience a recurrent stroke, compared with older patients who might experience death from stroke or other diseases before a recurrent stroke. Another possibility is that older Indigenous stroke patients are less likely to be hospitalized after a recurrent stroke for a range of clinical, cultural, and logistical reasons, which need to be considered in the management of Indigenous stroke patients (6).

We used a competing risks model for the analysis of stroke recurrence and long-term death because competing risks could affect differentials between Indigenous to non-Indigenous patients for these outcomes; to our knowledge, this is the first study to use competing risk analysis to investigate the role of comorbidities in these differentials. All-cause mortality is higher for the Indigenous than non-Indigenous populations, so lower incidence of recurrence might be observed for Indigenous stroke patients because they are more likely to die from other causes before suffering a recurrent stroke. In the presence of competing risk events, the cumulative incidence function is more appropriate than the nonparametric Kaplan–Meier cumulative hazard function, and the competing risks regression model outperforms the Cox model, especially when the hazard of the competing events is large (22,23).

Limitations of our study relate to the use of hospital records, retrospective data, and coding reliability. The number of deaths could be underestimated because prehospital deaths are not recorded due to the lack of data for non-hospitalized stroke patients. Thus, association between remoteness, pre-hospital deaths, and death for all stroke patients could not be assessed. Secondly, no information on risk factors for stroke recurrence.
such as smoking, diet, excessive alcohol use, and stroke severity, or on patient’s social, economic, or environmental circumstances, was available. The lack of information about the stroke severity and treatment is a major limitation. Higher stroke severity increases the risk of stroke recurrence, but we were unable to determine whether Indigenous stroke patients are more likely to suffer more severe strokes. Indigenous stroke patients might also have less access to or lower compliance with treatment, particularly after discharge from hospital; their higher levels of comorbidity might reduce their treatment options or the effectiveness of treatment. Thirdly, the reliability of comorbidity data relies on the accuracy of diagnosis coding in hospital records.

Cause of death was not available for this study. It has been previously reported that the most common causes of death in stroke patients who survive their initial stroke are cerebrovascular disease and heart disease (24,25). Patients that survived a first ischemic stroke were more than twice as likely to experience recurrent stroke as myocardial infarct but were more than twice as likely to die from a cardiac event as to die from a recurrent stroke (26). Knowing the cumulative incidence function of death and recurrent strokes could guide clinicians in devising treatment plans whether to focus on preventing recurrent stroke or cardiac death, knowing the similarities and differences between the treatment strategies for the two conditions (26–29). This is particularly important for Indigenous patients who have higher prevalence of and mortality from cardiac disease (6,19,30).

Conclusion

This study demonstrates that clinical risk factors for stroke such as comorbidity history can be identified from hospital data and analyzed in relation to stroke. In this study, it was found that Indigenous stroke patients have higher prevalence of comorbidities than non-Indigenous stroke patients, which explained part of the disparity in both case fatality and long-term survival but did not explain the disparity in stroke recurrence at all.

Acknowledgements

We acknowledge the acute care staff and data warehouse team of the NT Department of Health for their assistance with data access.

References

CHAPTER FOUR: DIFFERENT SURVIVAL ANALYSIS METHODS FOR MEASURING LONG TERM OUTCOMES OF INDIGENOUS AND NON-INDIGENOUS AUSTRALIAN CANCER PATIENTS IN THE PRESENCE AND ABSENCE OF COMPETING RISKS.

4.1 Preface

The previous chapter used the competing risks approach to investigate stroke recurrence and the role of comorbidities in adverse stroke outcomes (recurrent stroke and long-term deaths). It was found that comorbidities explained part of the disparity between Indigenous and other Australians in long-term survival but did not explain the disparity in stroke recurrence.

Competing risk is also an important issue in cancer survival. Net survival - the standard measure of cancer survival calculated using relative survival analysis - measures only the probability that cancer patients will die from their cancer; it ignores the chance of dying from other causes. Indigenous Australians have much higher all-cause mortality than other Australians; for them, net survival is an incomplete and potentially misleading measure of their cancer prognosis. This study compares two different survival analysis methods that take into account competing risks of death, to estimate the cancer and non-cancer death probability, of both Indigenous and non-Indigenous cancer patients. It also compared the multivariable results (i.e. Indigenous differentials) produced by relative survival (Poisson regression), cause specific survival (Cox regression) and competing risk (Fine-Gray regression) survival analysis methods. To estimate the cancer and non-cancer death probability, this study compared the Cronin-Feurer method (that uses life tables) with the Fine-Grey method (that uses cause of death data), which are the two ‘actual prognosis’ measures that consider competing risks.

For more details, please refer to Table 7.1 in the appendix that summarises the four different survival analysis methods used in Chapter Four.
4.2 Article

This chapter been resubmitted to the Population Health Metrics.

Vincent He conceived and designed the experiment, applied for and obtained the ethics and data-custodian approval (cancer register data), performed the data management and statistical analysis, conducted the literature review and wrote the manuscript draft. Professor John Condon, Dr Yuejen Zhao, Professor Peter Baade (senior research fellow and biostatistician at Cancer Council Queensland) and Dr Xiaohua Zhang (cancer epidemiologist at Health Gains Planning Branch) contributed to the review and revision of the related journal paper. Karen Dempsey (the former NT Cancer Register data-custodian at NT DoH Health Gains Planning Branch) extracted the cancer register data.
Different survival analysis methods for measuring long-term outcomes of Indigenous and non-Indigenous Australian cancer patients in the presence and absence of competing risks

(Running title: Cancer survival analysis methods)

Vincent YF He\textsuperscript{1,*}(BSc Hons), John R Condon\textsuperscript{1} (FAFPHM PhD), Peter D Baade\textsuperscript{1,2} (Astat PhD), Xiaohua Zhang\textsuperscript{3} (MPH PhD), Yuejen Zhao\textsuperscript{3} (BM MBiostats PhD)

\textsuperscript{1}Menzies School of Health Research, Charles Darwin University, Darwin, Australia
\textsuperscript{2}Cancer Council Queensland, Brisbane, Australia.
\textsuperscript{3}Health Gains Planning Branch, Northern Territory Government Department of Health, Darwin, Australia

Authors Details
Vincent Yaofeng He (Correspondence): vincent.he@menzies.edu.au, PO Box 41096, Casuarina NT 0811.
John Redmond Condon: john.condon@menzies.edu.au, PO Box 41096, Casuarina NT 0811.
Peter D Baade: PeterBaade@cancerqld.org.au, PO Box 201, Spring Hill Qld 4004.
Xiaohua Zhang: xiaohua.zhang@nt.gov.au, PO Box 40596, Casuarina, NT 0811, Australia.
Yuejen Zhao: yuejen.zhao@nt.gov.au, PO Box 40596, Casuarina, NT 0811, Australia.
Abstract

Background

Net survival is the most common measure of cancer prognosis and has been used to study differentials in cancer survival between ethnic or racial population sub-groups. However, net survival ignores competing risks of deaths and so provides incomplete prognostic information for cancer patients, and when comparing survival between populations with different all-cause mortality. Another prognosis measure, ‘crude probability of death’, which takes competing risk of death into account, overcomes this limitation. Similar to net survival, it can be calculated using either life tables (using Cronin-Feuer method) or cause of death data (using Fine-Gray method). The aim of this study is two-fold: (1) to compare the multivariable results produced by different survival analysis methods; and (2) to compare the Cronin-Feuer with the Fine-Gray methods, in estimating the cancer and non-cancer death probability of both Indigenous and non-Indigenous cancer patients and the Indigenous cancer disparities.

Methods

Cancer survival was investigated for 9595 people (18.5% Indigenous) diagnosed with cancer in the Northern Territory of Australia between 1991 and 2009. Cox proportional hazard model, Poisson and Fine-Gray regression were used in the multivariable analysis. The crude probabilities of cancer and non-cancer deaths were estimated using cause of death data with the Fine-Gray method and life tables with the Cronin-Feuer method.

Results

Multivariable regression using the relative survival, cause-specific survival and competing risk analysis produced similar results. In the presence of competing risks, the Cronin-Feuer method produced similar results to Fine-Gray in the estimation of cancer death probability (higher Indigenous cancer death probabilities for all cancers) and non-cancer death probabilities (higher Indigenous non-cancer death probabilities for some cancers).
probabilities for all cancers except lung cancer and head and neck cancers). Cronin-Feuer estimated much lower non-cancer death probabilities than Fine-Gray for non-Indigenous patients with head and neck cancers and lung cancers (both smoking-related cancers).

**Conclusion**

Despite its limitations, when valid and reliable subgroup-specific life tables are available and cause of death data is unavailable or unreliable, the Cronin-Feuer method is a reasonable alternative to the Fine-Gray method for assessing the Indigenous survival differential in the presence of competing risks.

Keywords: survival analysis, Indigenous Australians, competing risks, Fine-Gray model, net survival, crude probability of death, cancer prognosis, life tables, smoking, cause of death data
Background

The selection of the appropriate method and data to measure health disparities is a critical component of monitoring and evaluating government policy in eliminating disparities in health care as the choices will affect the disparities measured [1]. Studies comparing cancer survival between ethnic or racial population sub-groups have used different survival analysis methods, each of which has specific challenges to the validity of estimates of survival for such sub-groups and disparities between them [2]. Therefore, it is imperative to understand the assumptions, merits and limitations of prognosis measures [3] before using them.

The most common measure of cancer prognosis is net survival, which is the proportion of people who have not died from their cancer at the specified time after diagnosis [3-5]. Net survival is useful for national and international comparisons of cancer survival, where cancer patients’ risk of death may vary considerably between populations because of differences in general population mortality rates[6]. In particular, net survival has been used to study differentials in survival between ethnic or racial population sub-groups [2, 7-10].

In Australia, it is known that Indigenous cancer patients have lower net survival than non-Indigenous patients (five-year relative survival: 46.7% vs 83.4%) [8], but there is limited information on whether Indigenous cancer patients also have higher non-cancer death rates, given the higher prevalence of chronic diseases and much higher all-cause mortality rates in the Indigenous Australian population [11]. One of net survival’s limitations is that it ignores competing risks of deaths, relating only to the hypothetical world in which patients are only able to die of their cancer [3-6, 12], while a patient is more interested in the real world scenario when cancer and non-cancer causes of death are both possibilities[5, 6, 10]. When considering the available treatment options and weighing up the benefits, drawbacks and toxicities of cancer therapy, cancer patients and clinicians generally want to know about actual prognosis: what is the patient’s chance of dying from either cancer or another cause compared with the chance of
surviving [5]? This is particularly important for the Indigenous population or other populations who have higher chance of death from competing causes [3, 5, 6].

One of net survival’s limitations is that it ignores competing risks of deaths, relating only to the hypothetical world in which patients are only able to die of their cancer [3-6, 12], while a patient is more interested in the real world scenario when cancer and non-cancer causes of death are both possibilities[5, 6, 10]. Crude probability of cancer death (hereafter referred to as ‘cancer death probability’) is the probability of death from cancer in the presence of other causes; while crude probability of non-cancer death (hereafter referred to as ‘non-cancer death probability’) is the probability of death from other causes in the presence of cancer. In the statistical literature [3, 5, 6], net survival is known as a ‘cancer prognosis’ measure (ignores competing risks) while crude probability of death is known as an ‘actual prognosis’ measure (that considers competing risk).

Net survival can be estimated using cause-specific survival and relative survival methods [3, 5, 6, 13-16]. Cause-specific survival is calculated using cancer-specific deaths obtained from the cause of death data in the death certificates and/or clinical sources, treating cancer deaths as the end-point and deaths from other causes as censored observations; relative survival does not use cause of death data, but rather uses death from any cause as the end-point and is calculated by dividing the observed survival of the study cohort by the ‘expected’ survival of a comparable group in the general population, estimated using life tables [4-6, 13-16]. Similar to net survival, the crude probability of cancer and non-cancer deaths can be calculated in two ways—one based on the cause of death using methods such as the Fine-Gray competing risk model [10, 17, 18] (analogous to the cause-specific survival Cox model[19]), the other based on life tables using methods such as the Cronin-Feuer method [20] (analogous to relative survival).

To our knowledge, there is currently no study that uses the Cronin-Feuer method to investigate cancer and non-cancer death probabilities of minority sub-populations; possibly because of limited availability of
lifetables for such sub-populations. This is the case in Australia; nationally the availability of life tables for Indigenous Australians is limited because the reliability of life tables is unduly compromised by both uncertainty about population estimates and misclassification of Indigenous status in death registrations [8]. These two limitations are less of a problem in the Northern Territory (NT), where Indigenous Australians comprise 30% of the population (compared to 2-4% in other states). Life tables for the NT Indigenous population are more complete and reliable than for other states or the total Australian Indigenous population due to the more reliable and consistent Indigenous demographic data [21]. Indigenous status is also recorded with a high degree of accuracy in the NT Cancer Register (NTCR) [22]. The availability of high-quality data for the Indigenous population and Indigenous cancer patients, puts the NT in the best position to compare the implications of using the Cronin-Feuer and Fine-Gray methods to assess the Indigenous disparities in cancer and non-cancer deaths.

The aim of this study is two-fold: (1) to compare the multivariable results (i.e. Indigenous differentials) produced by relative survival, cause-specific survival and competing risk (Fine-Gray) survival analysis methods; and (2) to compare the Cronin-Feuer with the Fine-Gray survival analysis methods, in estimating the cancer and non-cancer death probability of both Indigenous and non-Indigenous cancer patients and the Indigenous cancer disparities in the presence of competing causes of death.

**Methods**

*Data*

Cancer registrations data for all NT residents diagnosed with cancer between 1st January 1991 and 31st December 2009 was obtained from the NTCR for the following data items: sex; date of birth; Indigenous status; remoteness of residence (urban or remote); date of diagnosis; cancer site, coded according to the International Classification of Diseases 10 (ICD-10); date of death; and underlying cause of death. Vital status was verified by matching the NTCR dataset to the National Death Index for deaths occurring up to
31 December 2011 (at the time of analysis, cause of death data were only available up until 31 December 2011). For the survival analysis of site-specific cancers, female breast (C50), colorectal (C18-20), lung (C33-34) and head and neck (C1-C14 & C30-32) cancers were chosen because there were sufficient numbers of both Indigenous and non-Indigenous cases for the analysis and they are Australia’s designated national “priority” cancers [23] (except head and neck cancers). Head and neck cancer was chosen as it is a major cause of morbidity and mortality associated with smoking.

Statistical analysis

The censoring date for the survival analysis was 31 December 2011. The survival time for cancer patients who died on the day of diagnosis was counted as half a day.

Relative survival was calculated using the ‘strs’ procedures [24] of Stata, in which the expected survival was estimated using the Ederer II method [25]. The NT non-Indigenous population has very high migration to and from other parts of Australia and similar health status and mortality to Australians generally, so life tables for the relevant years (1991-2009) for the total Australian population were used for non-Indigenous cancer patients. Life tables for the NT Indigenous population (for which deaths and population data are more reliable than for Indigenous people in other states and territories[21]) were used for the Indigenous cancer patients. Cancer and non-cancer death probabilities were calculated in two ways: the Fine-Gray model used cause of death information from the cancer cohort, and was derived using the “stcrreg” command in Stata, while the Cronin-Feuer method used life table data [20] derived using the “strs” command in Stata [24].

For multivariable analysis, the Cox proportional hazard regression, Poisson regression (generalised linear model with Poisson error structure) and Fine-Gray regression were used to investigate the Indigenous survival disparity; multivariable analysis was not available for the Cronin-Feuer method. The same independent variables were included in all final models: Indigenous status (Indigenous compared to non-Indigenous); age at diagnosis (per year); gender (female compared with male) and cancer site. Since the
an interaction term for Indigenous status by age at diagnosis (base age 55 years) was added to the models. For the Cox proportional hazard regression, scaled Schoenfeld residuals were used to check if the proportional hazards assumptions of each variable were satisfied. As the proportional hazards assumption was not met for Indigenous status, a step function of Indigenous status with follow-up time (as annual intervals) was also included in the model.

In the multivariable analysis for time trends, as all subjects had at least two years of potential follow up, the follow-up was limited to the first two years after diagnosis so that the shorter follow-up time for subjects diagnosed late in the study period did not bias time trends. Regression models included the same terms as above, plus year of diagnosis.

All analyses were conducted using Stata/SE 13 (StataCorp, College Station, TX). Ethics approval was obtained from the Human Research Ethics Committee of the Menzies School of Health Research and NT Department of Health. Approval to use the cancer registrations data was obtained from the NT Cancer Registry.

**Results**

There were 9595 invasive cancer cases between 1991 and 2009 that fulfilled the study criteria (Table 1). 18.6 % were identified as Indigenous, who were more likely to be female, younger and live in remote regions. A higher proportion of Indigenous (69.8%) than non-Indigenous (46.4%) patients died during the study period.

Adjusting for age and sex, net survival (estimated using relative survival) was lower for Indigenous than non-Indigenous patients at both one and five years after diagnosis for breast, colorectal, head and neck cancers (Table 2).
### Table 1 Demographic characteristics of people diagnosed with cancer, Australia NT, 1991-2009

<table>
<thead>
<tr>
<th></th>
<th>Indigenous (n=1789)</th>
<th>Non-Indigenous (n=7806)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>46.5</td>
<td>59.1</td>
</tr>
<tr>
<td>Female</td>
<td>53.5</td>
<td>40.9</td>
</tr>
<tr>
<td><strong>Age at diagnosis (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 to 49 years</td>
<td>36.3</td>
<td>30.3</td>
</tr>
<tr>
<td>50 to 59 years</td>
<td>26.6</td>
<td>26.0</td>
</tr>
<tr>
<td>60 to 69 years</td>
<td>22.0</td>
<td>23.7</td>
</tr>
<tr>
<td>70 years and over</td>
<td>15.1</td>
<td>20.0</td>
</tr>
<tr>
<td><strong>Median age(years)</strong></td>
<td>55</td>
<td>57</td>
</tr>
<tr>
<td><strong>Remoteness (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban</td>
<td>29.9</td>
<td>84.8</td>
</tr>
<tr>
<td>Remote</td>
<td>70.1</td>
<td>15.2</td>
</tr>
<tr>
<td><strong>Vital Status at 31/12/2011 (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alive</td>
<td>30.2</td>
<td>53.6</td>
</tr>
<tr>
<td>Dead</td>
<td>69.8</td>
<td>46.4</td>
</tr>
</tbody>
</table>

### Table 2 One-year and five-year cumulative relative survival (%) and 95% confidence interval by Indigenous status and cancer site (age and sex adjusted), Australia NT, 2001-2009

<table>
<thead>
<tr>
<th>Cancer Site</th>
<th>Cases</th>
<th>Indigenous One-year</th>
<th>Indigenous Five-year</th>
<th>Non-Indigenous One-year</th>
<th>Non-Indigenous Five-year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast</td>
<td>152</td>
<td>87.3(80.1-92.1)</td>
<td>66.4(55.7-75.1)</td>
<td>954</td>
<td>97.7(95.9-98.7)</td>
</tr>
<tr>
<td>Colorectal</td>
<td>96</td>
<td>64.3(53.6-73.1)</td>
<td>40.4(29.1-51.4)</td>
<td>784</td>
<td>84.2(81.3-86.8)</td>
</tr>
<tr>
<td>Head and Neck¹</td>
<td>194</td>
<td>54.0(46.5-60.9)</td>
<td>36.0(28.5-43.6)</td>
<td>514</td>
<td>85.7(82.3-88.5)</td>
</tr>
<tr>
<td>Lung¹</td>
<td>293</td>
<td>27.6(22.5-33.0)</td>
<td>8.9(5.7-12.9)</td>
<td>740</td>
<td>36.0(32.1-39.8)</td>
</tr>
</tbody>
</table>

¹Smoking-related cancers.
The multivariable results for all cancers combined (adjusted for cancer site) were similar for all terms included in the model for each of the three methods: relative survival analysis using Poisson regression; cause-specific analysis using Cox proportional hazards regression; and competing risk analysis using Fine-Gray regression (Table 3 and Table 4). The Poisson regression, Cox proportional hazards regression and the Fine-Gray regression (for cancer deaths) showed that the excess mortality of Indigenous patients was highest in the first year after diagnosis and decreased over time until the fifth year, with male and older patients experiencing higher risk of cancer death (Table 3). Age at diagnosis had different effects for Indigenous and non-Indigenous cases; the cancer death rate increased by 1% per year of age for Indigenous cases and 3% per year of age for non-Indigenous cases (Table 3). The Fine-Gray model (for non-cancer deaths) showed that Indigenous patients were more likely than non-Indigenous to die of non-cancer causes and that, in contrast to cancer deaths, this differential increased rather than decreased with time in the first five years after diagnosis (Table 3, Additional file 1: Table S3a). Remoteness of residence at time of diagnosis was not associated with risk of either cancer or non-cancer death for Indigenous or non-Indigenous patients and was therefore not included in all the regression models. The results for the analysis of time trend in two-year survival rate were similar for all three regression models (Table 4, Additional file 2: Table S4a); the death rate in the first two years after diagnosis decreased by 3% per diagnosis year after adjustment for Indigenous status, sex and age.
### Table 3 Regression analysis of cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk (Fine-Gray regression), all cancers combined\(^1\), Australia NT, 1991-2009

<table>
<thead>
<tr>
<th></th>
<th>Relative Survival HR(95% CI)(^2)</th>
<th>Cause-specific Competing (due to cancers) SHR(95% CI)</th>
<th>Competing (other death) SHR(95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous(^3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year after diagnosis</td>
<td>2.15(1.96-2.36)</td>
<td>2.17(1.98-2.38)</td>
<td>1.97(1.79-2.17)</td>
</tr>
<tr>
<td>2nd year after diagnosis</td>
<td>1.35(1.11-1.64)</td>
<td>1.47(1.22-1.76)</td>
<td>1.53(1.27-1.85)</td>
</tr>
<tr>
<td>3rd year after diagnosis</td>
<td>1.09(0.79-1.52)</td>
<td>1.32(1.01-1.72)</td>
<td>1.45(1.11-1.90)</td>
</tr>
<tr>
<td>4th year after diagnosis</td>
<td>1.29(0.83-1.99)</td>
<td>1.26(0.87-1.80)</td>
<td>1.44(1.00-2.08)</td>
</tr>
<tr>
<td>5th year after diagnosis</td>
<td>0.56(0.25-1.28)</td>
<td>0.78(0.46-1.32)</td>
<td>0.95(0.56-1.62)</td>
</tr>
<tr>
<td>Female vs Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indigenous(^3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year after diagnosis</td>
<td>2.15(1.96-2.36)</td>
<td>2.17(1.98-2.38)</td>
<td>1.97(1.79-2.17)</td>
</tr>
<tr>
<td>2nd year after diagnosis</td>
<td>1.35(1.11-1.64)</td>
<td>1.47(1.22-1.76)</td>
<td>1.53(1.27-1.85)</td>
</tr>
<tr>
<td>3rd year after diagnosis</td>
<td>1.09(0.79-1.52)</td>
<td>1.32(1.01-1.72)</td>
<td>1.45(1.11-1.90)</td>
</tr>
<tr>
<td>4th year after diagnosis</td>
<td>1.29(0.83-1.99)</td>
<td>1.26(0.87-1.80)</td>
<td>1.44(1.00-2.08)</td>
</tr>
<tr>
<td>5th year after diagnosis</td>
<td>0.56(0.25-1.28)</td>
<td>0.78(0.46-1.32)</td>
<td>0.95(0.56-1.62)</td>
</tr>
<tr>
<td>Age at diagnosis(^4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Indigenous</td>
<td>1.03(1.03-1.03)</td>
<td>1.03(1.03-1.03)</td>
<td>1.03(1.02-1.03)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>1.02(1.01-1.02)</td>
<td>1.01(1.01-1.02)</td>
<td>1.01(1.00-1.02)</td>
</tr>
</tbody>
</table>

\(^1\)Model adjusted for cancer site (with colorectal cancer as the reference category for cancer site).
\(^2\)HR=hazard ratio; SHR=standard hazard ratio.
\(^3\)Applies to the reference categories of the interaction terms (i.e. people of median age 55 years).
\(^4\)Per year of age.

### Table 4 Regression analysis of time trend after diagnosis using cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk analysis (Fine-Gray regression), all cancers combined\(^1\), Australia NT, 1991-2009

<table>
<thead>
<tr>
<th></th>
<th>Relative Survival HR(95% CI)(^2)</th>
<th>Cause-specific Competing (due to cancers) SHR(95% CI)</th>
<th>Competing (other death) SHR(95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous(^3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year of diagnosis</td>
<td>0.97(0.96-0.97)</td>
<td>0.97(0.96-0.97)</td>
<td>0.97(0.96-0.97)</td>
</tr>
</tbody>
</table>

\(^1\)Model adjusted for cancer site (with colorectal cancer as the reference category for cancer site).
\(^2\)HR=hazard ratio; SHR=standard hazard ratio.
\(^3\)Applies to the reference categories of the interaction terms (i.e. people of median age 55 years in 2009).
\(^4\)Per year of age.
The measures of crude probabilities of deaths are presented as proportion of cases that have died at five years after diagnosis (Table 5) from cancer and other causes. Cancer death probabilities (for all cancers combined and for the four individual cancer sites) were higher for Indigenous than non-Indigenous patients five years after cancer diagnosis, whether calculated by the Fine-Gray or Cronin-Feuer methods. Non-cancer death probabilities were also higher for Indigenous than non-Indigenous patients for all cancers combined and for breast and colorectal cancers, as estimated by both the Cronin-Feuer and Fine-Gray methods. Amongst the patients with head and neck cancers and lung cancer (both smoking related cancers) and colorectal cancer, the non-cancer death probabilities estimated by Cronin-Feuer were much lower than those estimated by the Fine-Gray methods for the non-Indigenous, but not the Indigenous patients. Amongst the patients with breast cancer, the cancer and non-cancer death probabilities estimated by Cronin-Feuer were higher than those estimated by the Fine-Gray methods. The Indigenous disparities (in cancer death) estimated by Cronin-Feuer method were lower than those estimated by the Fine-Gray method for all cancers except for breast cancer; the Indigenous disparities (in non-cancer death) estimated by Cronin-Feuer method were higher than those estimated by the Fine-Gray method for all cancers except for lung cancer and head and neck cancers.

Table 5: Five-year cumulative probabilities of cancer and non-cancer death (%), age and sex adjusted

<table>
<thead>
<tr>
<th>Cancer type</th>
<th>Cronin-Feuer Risk</th>
<th>Fine-Gray Risk</th>
<th>Ratio(CF/FG)</th>
<th>Cronin-Feuer Risk</th>
<th>Fine-Gray Risk</th>
<th>Ratio(CF/FG)</th>
<th>CF</th>
<th>FG</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cancers</td>
<td>61.0</td>
<td>65.7</td>
<td>0.93</td>
<td>33.2</td>
<td>33.2</td>
<td>1.00</td>
<td>1.84</td>
<td>1.98</td>
</tr>
<tr>
<td>Breast</td>
<td>31.6</td>
<td>28.9</td>
<td>1.09</td>
<td>13.2</td>
<td>12.2</td>
<td>1.08</td>
<td>2.39</td>
<td>2.37</td>
</tr>
<tr>
<td>Colorectal</td>
<td>57.7</td>
<td>65.1</td>
<td>0.89</td>
<td>40.6</td>
<td>40.1</td>
<td>1.01</td>
<td>1.42</td>
<td>1.62</td>
</tr>
<tr>
<td>Head &amp; neck</td>
<td>62.5</td>
<td>67.7</td>
<td>0.92</td>
<td>35.8</td>
<td>33.6</td>
<td>1.07</td>
<td>1.75</td>
<td>2.01</td>
</tr>
<tr>
<td>Lung</td>
<td>88.4</td>
<td>92.5</td>
<td>0.96</td>
<td>86.4</td>
<td>85.5</td>
<td>1.01</td>
<td>1.02</td>
<td>1.08</td>
</tr>
<tr>
<td>Other causes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All cancers</td>
<td>6.1</td>
<td>4.0</td>
<td>1.53</td>
<td>3.7</td>
<td>3.0</td>
<td>1.23</td>
<td>1.65</td>
<td>1.33</td>
</tr>
<tr>
<td>Breast</td>
<td>10.6</td>
<td>4.3</td>
<td>2.47</td>
<td>4.4</td>
<td>2.0</td>
<td>2.20</td>
<td>2.41</td>
<td>2.15</td>
</tr>
<tr>
<td>Colorectal</td>
<td>7.6</td>
<td>6.6</td>
<td>1.15</td>
<td>4.4</td>
<td>5.3</td>
<td>0.83</td>
<td>1.73</td>
<td>1.25</td>
</tr>
<tr>
<td>Head &amp; neck</td>
<td>6.2</td>
<td>5.9</td>
<td>1.05</td>
<td>2.8</td>
<td>8.2</td>
<td>0.34</td>
<td>2.21</td>
<td>0.72</td>
</tr>
<tr>
<td>Lung</td>
<td>4.0</td>
<td>3.0</td>
<td>1.33</td>
<td>1.6</td>
<td>3.9</td>
<td>0.41</td>
<td>2.50</td>
<td>0.77</td>
</tr>
</tbody>
</table>

1Disparity measured using Cronin-Feuer method (expressed as ratio of death probabilities among the Indigenous cancer cohort to the death probabilities among the non-Indigenous cohort diagnosed with the same cancer).
2Disparity measured using Fine-Gray method.
3Smoking-related cancers
Discussion

This study is the first to examine the implications of various survival methods to assess the survival inequalities faced by Indigenous cancer patients in the NT, in particular the impact of competing causes of death. Multivariable regression using relative survival (Poisson model), cause-specific survival (Cox regression) and competing risk analysis (Fine-Gray model) produced similar results. For all cancers combined, death rates were higher for Indigenous, male and older cancer patients for both cancer and non-cancer death. In the absence of competing risk, Indigenous people have lower five-year net survival than non-Indigenous people for all three cancers (breast, colorectal, head and neck cancers).

In the presence of competing risks, the Cronin-Feuer method produced similar results to the Fine-Gray method in the estimation of cancer death probability (higher Indigenous cancer death probabilities for all cancers) and non-cancer death probabilities (higher Indigenous non-cancer death probabilities for all cancers except lung cancer and head and neck cancers) five years after cancer diagnosis. The Cronin-Feuer method estimated much lower non-cancer death probabilities than the Fine-Gray method for non-Indigenous patients with head and neck cancers and lung cancer, which are both smoking-related cancers.

Comparison of death probabilities estimated using the Cronin-Feuer method and Fine-Gray models provides insights into ‘external factors’ affecting measures of survival for various cancer types such as smoking related cancers. The death probabilities estimated using the Cronin-Feuer method can be seen as the death probabilities that the cancer patients were ‘expected to have’, in the hypothetical scenario where cancer patients have the same death risks as the general population, unaffected by ‘external factors’ that cancer patients experienced such as unhealthy behaviours, access to health care or screening effects. The death probabilities estimated using cause of death data and the Fine-Gray method can be seen as the ‘observed death probabilities’ in which cancer patients might have different risks of death to the general population due to various ‘external factors’. Therefore, higher (or lower) non-cancer death probabilities
estimated by the Fine-Gray model imply higher (or lower) non-cancer death risks in cancer patients than the general population.

When analysing smoking-related cancers, caution is required for prognosis measures calculated using general population life tables such as relative survival and the Cronin-Feuer method [26-28], which assume that cancer patients have the same death risks as the general population. This assumption is probably violated for smoking-related cancers, because smoking increases mortality from other causes of death such as respiratory and cardiovascular diseases, resulting in higher non-cancer death risks in cancer patients than the general population. Therefore, the Cronin-Feuer method will underestimate the non-cancer death probabilities of the patients with smoking-related cancers, if it uses the general population life tables for the cancer patients with higher non-cancer death risks. The results from our study support the hypothesis that methods using life tables (relative survival and Cronin-Feuer) underestimate non-cancer deaths and overestimate cancer deaths for smoking-related cancers for non-Indigenous patients, suggesting that Cronin-Feuer method should be used with caution in estimating cancer and non-cancer death probability for non-Indigenous patients with smoking related cancers and quantifying the Indigenous disparity for smoking-related cancers.

Previous studies suggested that relative survival tends to underestimate social inequalities in cancer survival [29-31]. Our study suggests that the Cronin-Feuer method does underestimate Indigenous to non-Indigenous disparities in cancer death probabilities (except for breast cancer) while overestimating disparities in non-cancer death probabilities (except for lung, head and neck cancers, which are both smoking-related cancers) after considering competing risks. Underestimation of cancer death disparity by the Cronin-Feuer method was largest for colorectal cancer and head and neck cancers. For colorectal cancer, this underestimation was mainly due to underestimation of the probability of cancer death for Indigenous patients. For head and cancers, underestimation of disparity is due to a combination of underestimation of probability of cancer death for Indigenous patients and overestimation of cancer death
probabilities for non-Indigenous patients. However, for breast cancer patients the disparity in cancer
deaths was similar for the two methods; the Fine-Gray method produced lower estimates of both cancer
and non-cancer death probability than the Cronin-Feuer method for both Indigenous and non-Indigenous
patients.

The lower estimates of cancer death probabilities of the Fine-Gray method might indicate that cancer
survival for breast cancer patients is better than expected (calculated using life table). The lower estimates
of non-cancer death probabilities of the Fine-Gray method might indicate lower non-cancer five-year
death risk in breast cancer patients than the background population, supported by the findings of another
study[32] (not specifically Indigenous cancer patients) that breast cancer patients have reduced risks for
death (compared to the general population) from cardiovascular disease (SMR=83.9), which accounts for
the majority (55%) of non-cancer deaths among cancer patients generally in Australia; higher risks for
non-cancer deaths were observed for all cancer types except breast cancer and melanoma. The lower than
expected cancer and non-cancer death probability for breast cancer patients may be due to these patients
having higher socioeconomic status, which might be a protective factor of cancer and non-cancer deaths,
as indicated by the finding that breast cancer is more common in areas of higher socioeconomic status in
Australia[33] (breast cancer incidence rate was 122 per 100,000 in the highest socioeconomic status
group, compared to 103 per 100,000 in the lowest socioeconomic status group).

**Conclusion**

While net survival is useful for reporting trends in cancer survival, comparing different groups of cancer
patients and investigating the impact of various factors on cancer treatment, crude probability of death is
important when communicating risks to patients [6] during clinical decision making [10, 13]. Previous
studies have suggested that relative survival tends to underestimate social inequalities in cancer survival;
our study suggests that the Cronin-Feuer method does underestimate Indigenous to non-Indigenous
disparities in cancer death probabilities (except for breast cancer) in the presence of competing risks. Our
study also suggests that when analysing smoking-related cancers, caution is required when measuring cancer survival disparity using population life tables. Despite its limitations[20], when valid and reliable subgroup-specific life tables are available and cause of death data is unavailable or unreliable, the Cronin-Feuer method is a reasonable alternative to the Fine-Gray method for assessing the Indigenous survival differential in the presence of competing risks.

Declarations

Acknowledgements and Funding

VHe was supported by a University Postgraduate Research Scholarship from Charles Darwin University.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

VHe and JCo conceived and initiated the study. VHe undertook data management and/or analysis. VHe, JCo, PB, XZ and YZ draft the manuscript. All authors read and approved the final manuscript.
Reference

11. AIHW 2015. The health and welfare of Australia’s Aboriginal and Torres Strait Islander peoples: 2015. Cat. no. IHW 147. Canberra: AIHW.


## Appendix

Additional file 1: Table S3a: Regression analysis of cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk (Fine-Gray regression), all cancers combined\(^1\), Australia NT, 1991-2009 (full model). Description: Table S3a including hazard ratios for specific cancer sites.

<table>
<thead>
<tr>
<th>Relative survival</th>
<th>Cause-specific mortality</th>
<th>Competing (due to cancers)</th>
<th>Competing (other death)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR (95% CI)</td>
<td>HR (95% CI)</td>
<td>SHR (95% CI)</td>
<td>SHR (95% CI)</td>
</tr>
<tr>
<td>Indigenous(^3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st year after diagnosis</td>
<td>2.15(1.96-2.36)</td>
<td>2.17(1.98-2.38)</td>
<td>1.97(1.79-2.17)</td>
</tr>
<tr>
<td>2nd year after diagnosis</td>
<td>1.35(1.11-1.64)</td>
<td>1.47(1.22-1.76)</td>
<td>1.53(1.27-1.85)</td>
</tr>
<tr>
<td>3rd year after diagnosis</td>
<td>1.09(0.79-1.52)</td>
<td>1.32(1.01-1.72)</td>
<td>1.45(1.11-1.90)</td>
</tr>
<tr>
<td>4th year after diagnosis</td>
<td>1.29(0.83-1.99)</td>
<td>1.26(0.87-1.80)</td>
<td>1.44(1.00-2.08)</td>
</tr>
<tr>
<td>5th year after diagnosis</td>
<td>0.56(0.25-1.28)</td>
<td>0.78(0.46-1.32)</td>
<td>0.95(0.56-1.62)</td>
</tr>
<tr>
<td>Female vs Male</td>
<td>0.83(0.77-0.90)</td>
<td>0.84(0.78-0.90)</td>
<td>0.85(0.78-0.92)</td>
</tr>
<tr>
<td>Age at diagnosis(^4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-Indigenous</td>
<td>1.03(1.03-1.03)</td>
<td>1.03(1.03-1.03)</td>
<td>1.03(1.02-1.03)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>1.02(1.01-1.02)</td>
<td>1.01(1.01-1.02)</td>
<td>1.01(1.00-1.02)</td>
</tr>
<tr>
<td>Cancer Sites(^5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head and neck</td>
<td>1.17(1.00-1.37)</td>
<td>1.14(0.98-1.32)</td>
<td>1.14(0.98-1.32)</td>
</tr>
<tr>
<td>Stomach</td>
<td>2.97(2.49-3.54)</td>
<td>2.84(2.39-3.36)</td>
<td>2.79(2.37-3.28)</td>
</tr>
<tr>
<td>Liver</td>
<td>3.81(3.06-4.74)</td>
<td>3.34(2.69-4.16)</td>
<td>2.95(2.28-3.82)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>6.73(5.52-8.21)</td>
<td>5.94(4.88-7.23)</td>
<td>5.74(4.65-7.09)</td>
</tr>
<tr>
<td>Lung</td>
<td>3.94(3.47-4.48)</td>
<td>3.68(3.26-4.16)</td>
<td>3.56(3.15-4.03)</td>
</tr>
<tr>
<td>Bone</td>
<td>0.97(0.67-1.41)</td>
<td>0.93(0.65-1.34)</td>
<td>0.91(0.62-1.31)</td>
</tr>
<tr>
<td>Skin Cancers</td>
<td>0.18(0.14-0.23)</td>
<td>0.22(0.18-0.27)</td>
<td>0.22(0.18-0.27)</td>
</tr>
<tr>
<td>Breast</td>
<td>0.32(0.26-0.40)</td>
<td>0.38(0.31-0.46)</td>
<td>0.38(0.32-0.46)</td>
</tr>
<tr>
<td>Female genital cancers</td>
<td>0.98(0.80-1.19)</td>
<td>0.96(0.80-1.16)</td>
<td>0.94(0.78-1.14)</td>
</tr>
<tr>
<td>Male genital cancers</td>
<td>0.23(0.18-0.30)</td>
<td>0.30(0.25-0.36)</td>
<td>0.31(0.26-0.37)</td>
</tr>
<tr>
<td>Kidney</td>
<td>0.78(0.59-1.04)</td>
<td>0.72(0.55-0.96)</td>
<td>0.71(0.54-0.94)</td>
</tr>
<tr>
<td>Bladder</td>
<td>0.58(0.44-0.76)</td>
<td>0.57(0.45-0.74)</td>
<td>0.59(0.46-0.75)</td>
</tr>
<tr>
<td>Brain</td>
<td>3.31(2.65-4.14)</td>
<td>3.27(2.63-4.06)</td>
<td>3.22(2.60-3.99)</td>
</tr>
<tr>
<td>Thyroid</td>
<td>0.30(0.17-0.55)</td>
<td>0.32(0.19-0.53)</td>
<td>0.32(0.19-0.52)</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>0.92(0.74-1.15)</td>
<td>0.93(0.75-1.15)</td>
<td>0.92(0.75-1.13)</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>1.59(1.27-1.99)</td>
<td>1.48(1.19-1.85)</td>
<td>1.44(1.14-1.83)</td>
</tr>
<tr>
<td>Unknown primary</td>
<td>4.44(3.83-5.16)</td>
<td>4.09(3.54-4.73)</td>
<td>3.92(3.32-4.63)</td>
</tr>
<tr>
<td>Others</td>
<td>0.99(0.82-1.19)</td>
<td>0.87(0.73-1.05)</td>
<td>0.83(0.69-1.00)</td>
</tr>
</tbody>
</table>

\(^1\) Model adjusted for cancer site (with colorectal cancer as the reference category for cancer site).
\(^2\) HR=hazard ratio; SHR=standard hazard ratio.
\(^3\) Applies to the reference categories of the interaction terms (i.e. people of median age 55 years).
\(^4\) Per year of age.
\(^5\) Compared to colorectal cancer.
Additional file 2: Table S4a: Regression analysis of time trend after diagnosis using cause-specific mortality (Cox proportional hazard regression), relative survival (Poisson regression) and competing risk analysis (Fine-Gray regression), all cancers combined\(^1\), Australia NT, 1991-2009 (full model). Description: Table S4a including hazard ratios for specific cancer sites.

<table>
<thead>
<tr>
<th>Cancer Sites</th>
<th>Relative Survival HR(95% CI)(^1)</th>
<th>Cause-specific HR(95% CI)</th>
<th>Competing (due to cancers) SHR(95% CI)</th>
<th>Competing (other death) SHR(95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous(^3)</td>
<td>1.99(1.82-2.16)</td>
<td>2.01(1.85-2.19)</td>
<td>1.94(1.77-2.13)</td>
<td>2.49(1.70-3.63)</td>
</tr>
<tr>
<td>Female vs Male Age at diagnosis(^4)</td>
<td>0.86(0.79-0.93)</td>
<td>0.84(0.77-0.91)</td>
<td>0.87(0.79-0.94)</td>
<td>0.74(0.53-1.04)</td>
</tr>
<tr>
<td>Non-Indigenous</td>
<td>1.03(1.03-1.04)</td>
<td>1.03(1.03-1.04)</td>
<td>1.03(1.03-1.03)</td>
<td>1.07(1.06-1.08)</td>
</tr>
<tr>
<td>Indigenous</td>
<td>1.02(1.01-1.02)</td>
<td>1.02(1.01-1.02)</td>
<td>1.02(1.01-1.02)</td>
<td>1.02(1.00-1.04)</td>
</tr>
<tr>
<td>Year of diagnosis</td>
<td>0.97(0.96-0.97)</td>
<td>0.97(0.96-0.97)</td>
<td>0.97(0.96-0.97)</td>
<td>0.97(0.95-1.00)</td>
</tr>
<tr>
<td>Cancer Sites(^5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Head and neck</td>
<td>1.31(1.09-1.56)</td>
<td>1.27(1.07-1.51)</td>
<td>1.28(1.08-1.53)</td>
<td>1.07(0.58-1.98)</td>
</tr>
<tr>
<td>Stomach</td>
<td>3.47(2.86-4.21)</td>
<td>3.06(2.54-3.70)</td>
<td>3.05(2.54-3.68)</td>
<td>0.74(0.30-1.82)</td>
</tr>
<tr>
<td>Liver</td>
<td>4.82(3.81-6.09)</td>
<td>4.00(3.17-5.05)</td>
<td>3.69(2.83-4.82)</td>
<td>1.43(0.57-3.58)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>9.11(7.39-11.24)</td>
<td>6.86(5.57-8.45)</td>
<td>6.81(5.52-8.39)</td>
<td>0.59(0.18-1.98)</td>
</tr>
<tr>
<td>Lung</td>
<td>4.77(4.13-5.52)</td>
<td>4.05(3.52-4.67)</td>
<td>4.01(3.47-4.62)</td>
<td>0.68(0.38-1.21)</td>
</tr>
<tr>
<td>Bone</td>
<td>1.04(0.68-1.61)</td>
<td>0.94(0.61-1.46)</td>
<td>0.93(0.60-1.44)</td>
<td>2.48(0.83-7.39)</td>
</tr>
<tr>
<td>Skin Cancers</td>
<td>0.19(0.14-0.26)</td>
<td>0.21(0.16-0.28)</td>
<td>0.22(0.16-0.28)</td>
<td>0.74(0.38-1.43)</td>
</tr>
<tr>
<td>Breast</td>
<td>0.24(0.18-0.33)</td>
<td>0.28(0.21-0.36)</td>
<td>0.28(0.22-0.36)</td>
<td>0.60(0.28-1.30)</td>
</tr>
<tr>
<td>Female genital cancers</td>
<td>1.04(0.83-1.30)</td>
<td>0.99(0.79-1.23)</td>
<td>0.98(0.79-1.22)</td>
<td>1.47(0.73-2.98)</td>
</tr>
<tr>
<td>Male genital cancers</td>
<td>0.25(0.19-0.33)</td>
<td>0.29(0.23-0.36)</td>
<td>0.29(0.23-0.36)</td>
<td>1.15(0.68-1.94)</td>
</tr>
<tr>
<td>Kidney</td>
<td>0.86(0.62-1.20)</td>
<td>0.80(0.58-1.10)</td>
<td>0.78(0.56-1.09)</td>
<td>1.78(0.79-4.01)</td>
</tr>
<tr>
<td>Bladder</td>
<td>0.70(0.52-0.94)</td>
<td>0.74(0.56-0.97)</td>
<td>0.75(0.58-0.99)</td>
<td>0.82(0.35-1.90)</td>
</tr>
<tr>
<td>Brain</td>
<td>4.19(3.28-5.35)</td>
<td>3.75(2.95-4.77)</td>
<td>3.76(2.97-4.76)</td>
<td>0.74(0.18-3.08)</td>
</tr>
<tr>
<td>Thyroid</td>
<td>0.29(0.14-0.61)</td>
<td>0.34(0.18-0.62)</td>
<td>0.34(0.18-0.62)</td>
<td>1.03(0.24-4.42)</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>1.08(0.84-1.39)</td>
<td>1.05(0.83-1.35)</td>
<td>1.04(0.81-1.32)</td>
<td>1.17(0.53-2.60)</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>2.05(1.60-2.61)</td>
<td>1.91(1.50-2.43)</td>
<td>1.87(1.45-2.42)</td>
<td>1.62(0.73-3.62)</td>
</tr>
<tr>
<td>Unknown primary</td>
<td>5.70(4.83-6.73)</td>
<td>4.79(4.08-5.64)</td>
<td>4.68(3.92-5.59)</td>
<td>0.77(0.37-1.58)</td>
</tr>
<tr>
<td>Others</td>
<td>1.24(1.01-1.52)</td>
<td>1.12(0.91-1.37)</td>
<td>1.09(0.88-1.34)</td>
<td>2.09(1.21-3.64)</td>
</tr>
</tbody>
</table>

\(^1\) Model adjusted for cancer site (with colorectal cancer as the reference category for cancer site).
\(^2\) HR=hazard ratio; SHR=standard hazard ratio.
\(^3\) Applies to the reference categories of the interaction terms (i.e. people of median age 55 years in 2009).
\(^4\) Per year of age.
\(^5\) Compared to colorectal cancer.
CHAPTER FIVE: LONG TERM OUTCOMES FROM ACUTE RHEUMATIC FEVER (ARF) AND RHEUMATIC HEART DISEASE (RHD): A DATA-LINKAGE AND SURVIVAL ANALYSIS APPROACH.

5.1 Preface

This chapter demonstrated the usefulness of survival analysis using linked data (NT RHD Register, hospital inpatient and death registrations data) to analyse first ARF recurrence, progression from ARF to RHD, progression to severe RHD, development of RHD complications (heart failure, endocarditis, stroke and atrial fibrillation) and mortality. Because the patients have varying follow-up time, this study presented the occurrence of these adverse outcomes in terms of cumulative incidence and incidence rate over time since diagnosis to allow variable follow-up time for each patient; therefore this is a novel way in which survival analysis is used to identify the conditions that were more serious at different time periods. Linking hospital inpatient data to the RHD Register enabled us to more accurately identify the timing of the onset of RHD, to identify the occurrence of serious complications, and to document chronic disease comorbidities in ARF and RHD patients. Expanding on previous research, this study investigated the adverse for people with ARF and RHD and the effect of chronic disease comorbidities and other factors on these outcomes.
5.2 Article


Vincent He applied for and obtained the ethics and data-custodian approval (Rheumatic Heart Disease register, NT hospital inpatient data and mortality data), performed the data management and statistical analysis, conducted the literature review and wrote the manuscript draft. Vincent He, Professor John Condon, Professor Jonathan Carapetis (Director of Telethon Kids Institute and paediatrician at Princess Margaret Hospital for Children), and Dr Anna Ralph (infectious diseases specialist at RDH, Senior NHMRC Clinical Research Fellow at MSHR) and conceived and designed the experiment. Professor John Condon, Dr Anna Ralph, Dr Yuejen Zhao, Dr Kathryn Roberts (Paediatrician at RDH and PhD scholar at MSHR), Jessica de Dassel (RHD Secondary Prophylaxis project manager and PhD scholar at MSHR), Dr Keith Edwards (Community Paediatrician at NT DoH), Marea Fittock (NT Rheumatic Heart Disease Program Coordinator) and Professor Jonathan Carapetis contributed to the review and revision of the related journal paper. The staff of the DoH data management and system reporting branch extracted and linked the hospital inpatient data and mortality data. Cath Milne (NT RHD control program project manager and RHDAustralia Data Liaison Officer) extracted the NT RHD Register data. Shu Qin Li (Senior Epidemiologist at NT DoH Health Gains Planning Branch) linked the NT RHD Register to the hospital inpatient data.
LONG-TERM OUTCOMES FROM ACUTE RHEUMATIC FEVER AND RHEUMATIC HEART DISEASE
A DATA-LINKAGE AND SURVIVAL ANALYSIS APPROACH

BACKGROUND: We investigated adverse outcomes for people with acute rheumatic fever (ARF) and rheumatic heart disease (RHD) and the effect of comorbidities and demographic factors on these outcomes.

METHODS: Using linked data (RHD register, hospital, and mortality data) for residents of the Northern Territory of Australia, we calculated ARF recurrence rates, rates of progression from ARF to RHD to severe RHD, RHD complication rates (heart failure, endocarditis, stroke, and atrial fibrillation), and mortality rates for 572 individuals diagnosed with ARF and 1248 with RHD in 1997 to 2013 (94.9% Indigenous).

RESULTS: ARF recurrence was highest (incidence, 3.7 per 100 person-years) in the first year after the initial ARF episode, but low-level risk persisted for >10 years. Progression to RHD was also highest (incidence, 35.9) in the first year, almost 10 times higher than ARF recurrence. The median age at RHD diagnosis in Indigenous people was young, especially among males (17 years). The development of complications was highest in the first year after RHD diagnosis: heart failure incidence rate per 100 person-years, 9.09; atrial fibrillation, 4.70; endocarditis, 1.00; and stroke, 0.58. Mortality was higher among Indigenous than non-Indigenous RHD patients (hazard ratio, 6.55; 95% confidence interval, 2.45–17.51), of which 28% was explained by comorbid renal failure and hazardous alcohol use. RHD complications and mortality rates were higher for urban than for remote residents.

CONCLUSIONS: This study provides important new prognostic information for ARF/RHD. The residual Indigenous survival disparity in RHD patients, which persisted after accounting for comorbidities, suggests that other factors contribute to mortality, warranting further research.
Clinical Perspective

What Is New?

- Data linkage between an rheumatic heart disease (RHD) register, hospital data, and death register was used to investigate RHD disease progression, complication development, and survival and to examine the impact of comorbidities.
- In the first year after an acute rheumatic fever (ARF) episode, the incidence of progression to RHD was 10 times higher than ARF recurrence.
- Ten percent of RHD patients had severe disease at RHD diagnosis.
- The presence of comorbidities was associated with a higher incidence of RHD complications and mortality.
- Among the RHD patients, comorbid renal failure and hazardous alcohol use accounted for 28% of the excess Indigenous mortality.

What Are the Clinical Implications?

- To emphasize the need for integrated chronic disease management strategies for patients with ARF/RHD. This recommendation has relevance globally for settings with high ARF/RHD rates.
- To raise the question of what factors, other than higher comorbidity burden, account for remaining gaps in Indigenous mortality and complication rates.
- To show that although secondary penicillin prophylaxis is an important strategy to reduce ARF recurrence and RHD development, the persisting high progression rate from ARF to RHD reinforces the need for new ARF treatments and broader health policies focusing on primary and primordial prevention strategies.

Rheumatic heart disease (RHD) involves damage to the heart valves during episodes of acute rheumatic fever (ARF) after group A streptococcal infection. It is a disease overwhelmingly acquired in childhood among children living in poverty and overcrowded conditions. In the Australian general population and other developed countries, new cases of RHD have been rare for many decades; it is now almost entirely a disease of the elderly who acquired it as children ≥60 years ago.

However, RHD remains very common among Indigenous (Aboriginal and Torres Strait Islander) Australians,1 many of whom live in conditions of poverty and overcrowding. Such conditions are particularly common in the Northern Territory (NT), a large, sparsely populated area of northern Australia where Indigenous Australians make up 30% of the population and live mostly in remote, isolated communities.2 ARF and RHD incidence rates among the NT Indigenous population are among the highest rates reported worldwide. A recent study reported that the ARF incidence and RHD incidence were both >60 times higher for Indigenous compared with non-Indigenous NT residents, with significantly higher excess mortality.1 Another recent study reported that Indigenous Australians living in the NT are 54.8 times more likely to die of RHD than non-Indigenous Australians and have higher RHD mortality than Indigenous people elsewhere in Australia.3

In 1997, an RHD control program began in the NT. A register of all people diagnosed with ARF or RHD in the NT is a central component of this control program. The RHD Register is a recording and reminder system for regular penicillin prophylaxis (to prevent recurrent ARF episodes that cause cumulative damage to heart valves) and to coordinate specialist monitoring and management for those with heart valve damage. The RHD Register has also enabled comprehensive epidemiological analysis of ARF and RHD incidence, prevalence, and disease progression.1

Limited evidence is available on the development of complications of RHD4,5 or on the impact of chronic disease comorbidities on outcomes for RHD patients, despite the high prevalence of many chronic diseases among Indigenous Australians.6 Expanding on previous research,1 this study incorporated a data-linkage approach using the RHD Register, hospital inpatient data, and death register data to investigate adverse outcomes (eg, RHD complications and mortality) for people with ARF and RHD and the effect of chronic disease comorbidities and other factors (eg, hazardous alcohol use, exposure to violence, and trauma) on these outcomes.

METHODS

Data Sources

The NT RHD Register was used to identify NT residents with a first episode of ARF or diagnosed with RHD between January 1, 1997, and June 30, 2013. Data obtained from the register included demographic (sex, date of birth, Indigenous status, place of residence), diagnostic (date of first ARF episode, date of RHD diagnosis), and clinical information. The register also includes the NT’s unique health client identifier (the Hospital Registration Number) that is used by all public health services (hospitals, primary health care, etc) in the NT.

An extract of all NT residents on the RHD Register was linked to the NT public hospitals inpatient data set (matched on Hospital Registration Number) to obtain data on all hospital inpatient episodes for individuals on the register. Date of birth and sex were used to verify the match. Data were obtained from the inpatient data set for each inpatient episode: dates of admission and discharge and principal and additional diagnoses. Diagnosis data were used to identify inpatient episodes with ARF or RHD as a principal or secondary diagnosis and those with chronic disease comorbidities and, for RHD patients only, those who had been hospitalized for serious RHD complications (heart failure, stroke, atrial fibrillation, and endocarditis) or indicators of high-risk behaviors and environments (hazardous alcohol use, assault, and transportation accidents; Table 1). The Charlson Comorbidity Index (CCI) was used to...
comorbidity; and CCI score
no comorbidities; CCI score=1 indicates presence of only 1
summary measure of 19 comorbid conditions that has been
measure chronic disease comorbidity. The CCI is a widely used
classified as remote residents.

**Study Outcomes**
The primary outcomes for patients with a first ARF episode were recurrent ARF and progression to RHD. The RHD Register was used to identify recurrent ARF episodes. The primary outcomes for patients with RHD were progression to severe RHD, diagnosis of a serious complication (heart failure, endocarditis, stroke, or atrial fibrillation), or death. Data from the RHD Register about clinical reviews by medical specialists were used to identify progression to severe RHD. Since 2004, the NT RHD control program has classified the patient’s cardiac status into 4 categories (no, mild, moderate, or severe RHD) based on Australian national guidelines. Hospital inpatient data were used to identify the occurrences of RHD complications (atrial fibrillation, endocarditis, heart failure, or stroke). All 4 are almost always investigated and treated in hospital at their first occurrence. The secondary outcomes were hospitalization for treatment of ARF or RHD after the first ARF episode or RHD diagnosis.

**Statistical Analysis**
Statistical analyses were conducted with Stata, version 13 (StataCorp, College Station, TX). The χ² test was used to compare prevalence of comorbidities, high-risk indicators, and RHD hospitalization (within 1 year) between Indigenous and non-Indigenous patients and other population subgroups.

Survival analysis was used to analyze the following adverse outcomes for ARF patients: ARF recurrence and progression to RHD. For RHD patients, the following adverse outcomes were analyzed: progression to severe RHD, the occurrence of RHD complications, and death. For each of the adverse events except death, the follow-up time was censored at the earliest of the following: date of diagnosis of each adverse event, date of death, or June 30, 2013. For the analysis of death, follow-up time was censored at June 30, 2013. The incidence rate (per 100 person-years) of adverse outcomes was calculated for years 0, 1 to 4, 5 to 9, and 10 to 14 after diagnosis. The Stata stptime command was used to calculate incidence rates, which allowed variable follow-up time for each patient. The risks of the adverse outcomes at each time point (1, 5, and 10 years) were calculated from the Kaplan-Meier failure function as the cumulative probability of each event using the Stata sts list command.

For the multivariable analysis, separate Cox proportional hazards regression analysis was used to identify factors associated with each adverse outcome. The same independent variables were included in all final regression models: Indigenous status (Indigenous compared with non-Indigenous); remoteness of residence (remote compared with urban); sex (male compared with female); age at diagnosis (per year); year of diagnosis (per year);
and comorbidity (CCI scores of 1 and 2 or more compared with 0). Cox regression and logistic regression were used to identify the factors associated with higher hospitalization for ARF/RHD treatment of ARF patients and higher hospitalization for RHD treatment (within 1 year) of RHD patients.

**Ethics**

The study was approved by the Human Research Ethics Committee of the NT Department of Health and the Menzies School of Health Research (HREC-2011-1680). Approval to access RHD Register data was obtained from the NT RHD Steering Committee and the Register’s data custodian. Approval to access the NT hospital inpatient data and Client Master Index data was obtained from the NT hospital inpatients data custodian.

**RESULTS**

After the exclusion of those who died before 1997 or had their first ARF episode or RHD diagnosis after June 30, 2013, there were 2660 potentially eligible individuals in the RHD Register.

To derive the ARF cohort, we excluded 926 patients with first ARF episode or RHD diagnosis before 1997, 193 with unconfirmed ARF, and 969 patients in whom we were unable to identify the first ARF episode, leaving 572 individuals with a first ARF episode in the study period. To derive the RHD cohort, we excluded 746 with no RHD diagnosis, 561 patients diagnosed before 1997, and 105 patients who had prior hospitalization for heart failure (n=62), endocarditis (n=11), stroke (n=10), or atrial fibrillation (n=22), leaving 1248 patients diagnosed with RHD during the study period. There were 152 RHD patients who had a clinical review for RHD before their RHD diagnosis date was recorded in the RHD Register; the diagnosis date was replaced with the first review date. Similarly, the RHD diagnosis date was replaced for 330 patients who had a prior inpatient episode with a diagnosis of RHD. For the analysis of ARF and RHD patients, there were no missing data for Indigenous status, sex, age, and remoteness.

**ARF Patient Cohort**

**ARF Patients’ Characteristics**

Of the 572 confirmed first ARF episodes, 97.0% patients were Indigenous, 43.5% were male, and 5.1% had comorbidities (CCI score ≥1; Table 2). The median age of

| Table 2. Demographic Characteristics of ARF and RHD Patients, NT, 1997 to 2013 |
|----------------------------------|-------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                   | ARF, %                        | RHD, %          |
|                                   | Indigenous (n=555)            | Non-Indigenous (n=17) | Indigenous (n=1173) | Non-Indigenous (n=75) |
| Male                              | 43.2                          | 52.9            | 35.2            | 28.0            |
| Remote resident                   | 85.4                          | 29.4            | 84.1            | 20.0            |
| Age group, y                      |                               |                 |                 |                 |
| 0–4                               | 0.9                           | 0.0             | 0.3             | 0.0             |
| 5–9                               | 21.1                          | 5.9             | 10.3            | 0.0             |
| 10–14                             | 39.5                          | 23.5            | 20.0            | 6.7             |
| 15–24                             | 25.4                          | 23.5            | 26.0            | 9.3             |
| 25–34                             | 7.9                           | 17.7            | 17.8            | 10.7            |
| 35–44                             | 4.3                           | 29.4            | 13.1            | 18.7            |
| ≥45                               | 0.9                           | 0.0             | 12.4            | 54.7            |
| Median age, y                     | 12                            | 19              | 21              | 46              |
| Year of diagnosis                 |                               |                 |                 |                 |
| 1997–2000                         | 16.8                          | 41.2            | 18.2            | 17.3            |
| 2001–2005                         | 33.7                          | 17.7            | 38.3            | 56.0            |
| 2006–2010                         | 28.7                          | 11.8            | 31.5            | 18.7            |
| 2011–2013                         | 20.9                          | 29.4            | 11.9            | 8.0             |
| CCI score                         |                               |                 |                 |                 |
| 0                                 | 94.8                          | 100.0           | 78.9            | 73.3            |
| 1                                 | 4.0                           | 0.0             | 13.0            | 14.7            |
| ≥2                                | 1.3                           | 0.0             | 8.0             | 12.0            |

ARF indicates acute rheumatic fever; CCI, Charlson Comorbidity Index; NT, Northern Territory; and RHD, rheumatic heart disease.
first presentation was 12 years for Indigenous subjects regardless of sex. Eleven ARF patients died in the 16-year study period.

**ARF Recurrence**
The incidence of recurrent ARF was greatest in the first year after the first ARF episode (Table 3); recurrences continued to be seen >10 years after the first episode (incidence rate, 1.41 per 100 person-years 10–14 years after first ARF episodes). Among the ARF patients who were ≤12 years of age, the incidence rate was 2.52 per 100 person-years (95% confidence interval, 1.05–6.04) in years 10 to 14.

The cumulative incidence of ARF recurrence at 10 years was 19.8% (Table 3). Among Indigenous people, the cumulative incidence of ARF recurrence was 3.8% at 1 year, 14.9% at 5 years, and 20.1% at 10 years. In multivariable analysis, the only factor associated with time to ARF recurrence was age at first ARF episode (incidence decreased by 9% per year of age; Table 4).

**Progression to RHD**
The risk of progression to RHD was higher than ARF recurrence, was very high in the first year after the first ARF episode, and decreased thereafter (Table 3). The cumulative incidence of progression to RHD was 27.1% at 1 year, 44.0% at 5 years, and 51.9% at 10 years. In multivariable analysis, the rate of progression to RHD decreased with age at first ARF episode (by 3% per year of age). The risk of developing RHD was higher for remote residents and those with CCI score ≥2 (Table 4).

**Hospitalization for ARF or RHD Treatment**
Among the 572 ARF patients, the date of hospital admission corresponded with the date of first ARF notification in 163 patients (28.5%), and 323 (56.5%) were hospitalized for ARF or RHD treatment within 14 days of notification. After adjustment for age, sex, remoteness of residence, and diagnosis year but not comorbidities, Indigenous people had higher hospitalization for ARF/RHD treatment than non-Indigenous people (hazard ratio [HR], 2.18; _P_ = 0.047); after further adjustment for comorbidities, the association of being Indigenous and higher hospitalization for ARF/RHD treatment became statistically insignificant (HR, 2.14; _P_ = 0.053; Table 5). In the multivariable analysis adjusting for comorbidities, the hospitalization for ARF/RHD treatment was higher for ARF patients who were younger and with later diagnosis years (Table 5).

**RHD Patient Cohort**

**RHD Patients’ Characteristics**
Of the 1248 people diagnosed with RHD, 94.0% were Indigenous and 34.8% were male (Table 2). Age at RHD diagnosis was younger for male than female patients and for Indigenous than non-Indigenous patients (median age at diagnosis: Indigenous: male patients, 17 years; female patients, 23 years; non-Indigenous: male patients, 42 years; female patients, 49 years). Of RHD patients, 13.1% had a CCI score of 1 and 8.25% had a CCI score ≥2. Urban residents had more comorbidities (19.8% with CCI score of 1 and 13.8% with a CCI score ≥2) than remote residents (11.5% with a CCI score of 1 and 6.9% with a CCI score ≥2; _P_ < 0.001). After adjustment for age, Indigenous RHD patients were more likely to have comorbidities (CCI score ≥1) than non-Indigenous patients (odds ratio, 2.50; 95% confidence interval, 1.34–4.64).

### Table 3. Incidence Rate and Cumulative Incidence Rate (95% Confidence Interval) of Adverse Outcomes for ARF Patients (n=572) at Different Years After the First ARF Diagnosis

<table>
<thead>
<tr>
<th>Year</th>
<th>ARF Recurrence</th>
<th>Progression to RHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1</td>
<td>3.72 (2.40–5.77)</td>
<td>35.92 (30.66–42.09)</td>
</tr>
<tr>
<td>1–5</td>
<td>3.01 (2.27–3.99)</td>
<td>6.66 (5.31–8.35)</td>
</tr>
<tr>
<td>5–10</td>
<td>1.31 (0.80–2.14)</td>
<td>2.99 (1.99–4.50)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>1.41 (0.63–3.14)</td>
<td>1.47 (0.55–3.92)</td>
</tr>
<tr>
<td>Total</td>
<td>2.38 (1.94–2.93)</td>
<td>9.84 (8.70–11.12)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cumulative incidence (%)</th>
<th>ARF Recurrence</th>
<th>Progression to RHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.66 (2.38–5.61)</td>
<td>27.09 (23.61–30.96)</td>
</tr>
<tr>
<td>5</td>
<td>14.43 (11.53–17.98)</td>
<td>43.95 (39.67–48.47)</td>
</tr>
<tr>
<td>10</td>
<td>19.82 (16.18–24.16)</td>
<td>51.89 (47.12–56.84)</td>
</tr>
</tbody>
</table>

**Table 4. Multivariable Cox regression for ARF Adverse Outcomes, NT, 1997 to 2013**

<table>
<thead>
<tr>
<th>Variable</th>
<th>ARF Recurrence</th>
<th>Progression to RHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous vs non-Indigenous</td>
<td>2.00 (0.26–15.27)</td>
<td>0.97 (0.38–2.43)</td>
</tr>
<tr>
<td>Remote vs urban</td>
<td>1.02 (0.55–1.91)</td>
<td>1.70 (1.14–2.55)*</td>
</tr>
<tr>
<td>Male vs female</td>
<td>0.97 (0.64–1.48)</td>
<td>0.84 (0.65–1.08)</td>
</tr>
<tr>
<td>Age at diagnosis (per year of age)</td>
<td>0.91 (0.87–0.95)</td>
<td>0.97 (0.96–0.99)*</td>
</tr>
<tr>
<td>Diagnosis year (per year)</td>
<td>1.03 (0.97–1.09)</td>
<td>1.02 (0.99–1.05)</td>
</tr>
<tr>
<td>CCI score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.21 (0.44–3.31)</td>
<td>0.88 (0.43–1.80)</td>
</tr>
<tr>
<td>≥2</td>
<td>1.71 (0.24–12.44)</td>
<td>5.63 (2.17–14.61)*</td>
</tr>
</tbody>
</table>

ARF indicates acute rheumatic fever; CCI, Charlson Comorbidity Index; CI, confidence interval; HR, hazard ratio; NT, Northern Territory; and RHD, rheumatic heart disease.

* _P_ < 0.05.
Rheumatic Heart Disease: Lessons From Linked Data

ORIGINAL RESEARCH ARTICLE

RHD Hospitalization After RHD Diagnosis
Fewer than half of the 1248 RHD patients were hospitalized for RHD treatment within 1 year of diagnosis: 345 (27.6%) on the day the diagnosis was first recorded, 422 (33.8%) within 90 days, and 492 (39.4%) within 1 year. The proportion hospitalized within 1 year for RHD treatment was greater for female than male patients (42.6% compared with 33.1%) and for Indigenous than non-Indigenous patients (40.1% compared with 29.3%) but was similar for remote and urban residents (41.9% compared with 38.8%). After adjusting for age, sex, remoteness of residence, and diagnosis year but not comorbidities, Indigenous people had higher hospitalization within 1 year for RHD treatment than non-Indigenous people (odd ratio, 2.08; \( P < 0.05 \)); after further adjustment for comorbidities, the association between Indigenous status and hospitalization within 1 year for RHD treatment became statistically insignificant (odds ratio, 1.61; \( P = 0.116 \); Table 5). In the multivariable logistic regression adjusted for comorbidities, the hospitalization rate within 1 year for RHD treatment was higher for RHD patients who were female, younger, and with comorbidities (Table 5).

Progression to Severe RHD
Seventy-eight of the 772 patients (10.1%) diagnosed with RHD between 2004 and 2013 had severe RHD at diagnosis. The rate of progression to severe RHD was highest in the first year after diagnosis and decreased thereafter (Table 6).

Complications
Heart failure and atrial fibrillation were more common complications than endocarditis or stroke (Table 6). The incidence of development of all 4 complications was highest in the first year after diagnosis and decreased

Table 5. Hospitalization for ARF/RHD Treatment of ARF Patients (n=572) and Hospitalization (Within 1 Year of RHD diagnosis) for RHD Treatment of RHD Patients (n=1248), NT, 1997 to 2013

<table>
<thead>
<tr>
<th>Exposure Variables</th>
<th>Hospitalization for ARF/RHD Treatment</th>
<th>Hospitalization for RHD Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline Plus CCI Baseline Plus CCI</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HR (95% CI) OR (95% CI) HR (95% CI) OR (95% CI)</td>
<td></td>
</tr>
<tr>
<td>Indigenous</td>
<td>2.18 (1.01–4.71)* 2.14 (0.99–4.63) 2.08 (1.18–3.67)* 1.61 (0.89–2.90)</td>
<td></td>
</tr>
<tr>
<td>Urban vs remote</td>
<td>0.98 (0.75–1.28) 0.97 (0.74–1.27) 0.79 (0.58–1.08) 0.92 (0.66–1.27)</td>
<td></td>
</tr>
<tr>
<td>Female vs male</td>
<td>1.12 (0.92–1.36) 1.12 (0.93–1.36) 1.47 (1.15–1.88)* 1.56 (1.21–2.01)*</td>
<td></td>
</tr>
<tr>
<td>Age (per year of age)</td>
<td>0.98 (0.97–0.99)* 0.98 (0.97–0.99)* 1.00 (1.00–1.01) 1.00 (0.98–1.00)*</td>
<td></td>
</tr>
<tr>
<td>Diagnosis year (per year)</td>
<td>1.04 (1.02–1.06)* 1.04 (1.02–1.06)* 1.00 (0.97–1.03) 1.00 (0.97–1.03)</td>
<td></td>
</tr>
<tr>
<td>CCI score 1 vs 0</td>
<td>0.93 (0.54–1.59) 1.81 (0.72–4.54)</td>
<td></td>
</tr>
<tr>
<td>CCI score ≥2 vs 0</td>
<td>4.95 (3.41–7.19)*</td>
<td></td>
</tr>
</tbody>
</table>

ARF indicates acute rheumatic fever; CCI, Charlson Comorbidity Index; CI, confidence interval; HR, hazard ratio; NT, Northern Territory; OR, odds ratio; and RHD, rheumatic heart disease.

*\( P <0.05 \).

Table 6. Incidence Rate and Cumulative Incidence Rate (95% Confidence Interval) of Adverse Outcomes for RHD Patients (n=1248) at Different Years After the First RHD Diagnosis

<table>
<thead>
<tr>
<th>Year</th>
<th>Severe RHD*</th>
<th>Atrial Fibrillation</th>
<th>Endocarditis</th>
<th>Heart Failure</th>
<th>Stroke</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–1</td>
<td>25.72 (21.99–30.09)</td>
<td>4.70 (3.61–6.13)</td>
<td>1.00 (0.57–1.76)</td>
<td>9.06 (7.46–11.00)</td>
<td>0.58 (0.28–1.22)</td>
<td>0.50 (0.22–1.10)</td>
</tr>
<tr>
<td>1–5</td>
<td>2.61 (1.96–3.47)</td>
<td>0.77 (0.54–1.10)</td>
<td>0.31 (0.18–0.54)</td>
<td>1.23 (0.92–1.63)</td>
<td>0.24 (0.13–0.44)</td>
<td>0.83 (0.59–1.15)</td>
</tr>
<tr>
<td>5–10</td>
<td>1.95 (1.23–3.10)</td>
<td>1.30 (0.95–1.76)</td>
<td>0.38 (0.22–0.66)</td>
<td>1.44 (1.07–1.94)</td>
<td>0.38 (0.22–0.66)</td>
<td>1.33 (1.00–1.78)</td>
</tr>
<tr>
<td>&gt;10</td>
<td>3.43 (1.11–10.65)</td>
<td>1.41 (0.85–2.33)</td>
<td>0.42 (0.18–1.02)</td>
<td>1.60 (0.98–2.62)</td>
<td>0.77 (0.40–1.48)</td>
<td>1.90 (1.26–2.85)</td>
</tr>
<tr>
<td>Total</td>
<td>6.55 (5.75–7.47)</td>
<td>1.51 (1.28–1.78)</td>
<td>0.43 (0.32–0.58)</td>
<td>2.33 (2.03–2.67)</td>
<td>0.39 (0.29–0.54)</td>
<td>1.09 (0.90–1.31)</td>
</tr>
<tr>
<td>Cumulative incidence (%)</td>
<td>1</td>
<td>20.45 (17.75–23.49)</td>
<td>4.46 (3.44–5.77)</td>
<td>0.97 (0.55–1.7)</td>
<td>8.24 (6.84–9.92)</td>
<td>0.58 (0.28–1.21)</td>
</tr>
<tr>
<td>5</td>
<td>28.04 (24.87–31.52)</td>
<td>7.42 (6.04–9.1)</td>
<td>2.17 (1.47–3.2)</td>
<td>12.63 (10.85–14.67)</td>
<td>1.51 (0.94–2.42)</td>
<td>3.75 (2.77–5.06)</td>
</tr>
<tr>
<td>10</td>
<td>34.62 (30.4–39.23)</td>
<td>13.35 (11.23–15.84)</td>
<td>4.03 (2.9–5.58)</td>
<td>18.62 (16.25–21.28)</td>
<td>3.63 (2.5–5.28)</td>
<td>10.27 (8.33–12.63)</td>
</tr>
</tbody>
</table>

RHD indicates rheumatic heart disease.

*Only for those with diagnosis year of 2004 or later (n=772).
thereafter (Table 6). In the multivariable analysis, urban residents were more likely than remote residents to develop atrial fibrillation and heart failure; atrial fibrillation and stroke incidence increased with age at diagnosis (by 5% and 4% per year of age, respectively; Table 7). RHD patients with comorbidities (CCI score ≥1) had a higher incidence of atrial fibrillation, endocarditis, and heart failure.

**Death**

After adjustment for age, sex, remoteness of residence, and diagnosis year but not comorbidities, Indigenous people had higher risk of death after a RHD diagnosis than non-Indigenous people (HR, 6.59; P<0.01). Adjustment for comorbidities (Table 7) reduced this disparity somewhat (HR, 5.19; P<0.01). In the multivariable analysis (Table 7), the death rate was higher for patients who were Indigenous, male, older, or urban-dwelling or who had comorbidities.

**Comorbidities of RHD Patients ≥18 Years Old**

The proportion of patients with any comorbidity (CCI score ≥1) was the same for Indigenous and non-Indigenous RHD patients. However, the Indigenous RHD patients had a higher prevalence of hospitalization as a result of hazardous alcohol use (13.2% compared with 0%; P=0.002) and assault (12.3% compared with 0%; P=0.003) than the non-Indigenous RHD patients (Table 8). RHD patients with hazardous alcohol use also had higher hospitalization resulting from assault (53.4% compared with 5.4%; P<0.001; Table 8). Among Indigenous RHD patients, the prevalence of any comorbidity (CCI score ≥1) was higher for urban than for remote residents (43.0% compared with 27.0%; P=0.001), as was the prevalence of many individual conditions (Table 8).

In the baseline Cox regression model, death rates were higher for Indigenous, urban, male, and older RHD adult patients (Table 9). The HR for Indigenous status decreased from 6.55 in the baseline model to 4.64 when also adjusted for renal failure, and it decreased further to 3.87 when further adjusted for hazardous alcohol use (Table 9), indicating that the higher prevalence of renal failure and hazardous alcohol use (which are both associated with higher mortality) was part of the reason for the higher mortality of Indigenous RHD patients. Further adjustment for other comorbidities (ie, adding CCI excluding renal failure to the model) did not decrease the HR for Indigenous individuals much further. Adjustment for renal failure and hazardous alcohol use similarly reduced the HR for urban compared with remote residents (Table 9).

**DISCUSSION**

The NT RHD Register has previously been used to investigate ARF and RHD occurrence, progression, and survival for a population with a very high RHD burden. This study used a data-linkage approach to refine and expand that work to include adverse outcomes other than death and the influence of chronic disease comorbidity on adverse outcomes. Linking hospital inpatient data to the RHD Register enabled us to more accurately identify the timing of the onset of RHD, to identify the occurrence of serious complications, and to document chronic disease comorbidities in ARF and RHD patients.

Our results confirm the previous findings that the progression from ARF to RHD is rapid and is occurring faster than documented ARF recurrences; RHD incidence was almost 10 times higher than the incidence of ARF recurrence in the year after the first ARF episode. Secondary prophylaxis with penicillin is important for reducing ARF recurrence and the consequent worsening of heart valve dysfunction.

**Table 7. Multivariable Cox Regression for RHD Adverse Outcomes, NT, 1997 to 2013**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Severe RHD*</th>
<th>Atrial Fibrillation</th>
<th>Endocarditis</th>
<th>Heart Failure</th>
<th>Stroke</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous</td>
<td>0.73 (0.43–1.25)</td>
<td>1.37 (0.76–2.46)</td>
<td>1.18 (0.32–4.37)</td>
<td>1.21 (0.70–2.07)</td>
<td>1.39 (0.44–4.41)</td>
<td>5.19 (1.96–13.77)†</td>
</tr>
<tr>
<td>Urban vs remote</td>
<td>1.16 (0.83–1.62)</td>
<td>1.67 (1.12–2.50)†</td>
<td>1.72 (0.84–3.50)</td>
<td>1.78 (1.29–2.45)†</td>
<td>1.67 (0.77–3.64)</td>
<td>1.70 (1.08–2.67)†</td>
</tr>
<tr>
<td>Male vs female</td>
<td>0.89 (0.68–1.17)</td>
<td>1.30 (0.92–1.84)</td>
<td>1.19 (0.64–2.23)</td>
<td>1.09 (0.81–1.45)</td>
<td>0.94 (0.46–1.90)</td>
<td>1.67 (1.13–2.47)†</td>
</tr>
<tr>
<td>Age (per year)</td>
<td>1.00 (0.99–1.01)</td>
<td>1.05 (1.04–1.06)†</td>
<td>0.99 (0.97–1.02)</td>
<td>1.01 (1.00–1.02)</td>
<td>1.04 (1.02–1.06)†</td>
<td>1.04 (1.03–1.06)†</td>
</tr>
<tr>
<td>Diagnosis year</td>
<td>1.05 (1.00–1.11)</td>
<td>1.01 (0.96–1.06)</td>
<td>1.05 (0.96–1.14)</td>
<td>0.99 (0.95–1.03)</td>
<td>0.99 (0.90–1.10)</td>
<td>0.96 (0.90–1.02)</td>
</tr>
<tr>
<td>CCI ≥1</td>
<td>1.19 (0.74–1.91)</td>
<td>1.37 (0.83–2.26)</td>
<td>3.25 (1.33–7.92)†</td>
<td>2.94 (1.95–4.44)†</td>
<td>1.28 (0.49–3.38)</td>
<td>2.85 (1.73–4.70)†</td>
</tr>
</tbody>
</table>

CCI indicates Charlson Comorbidity Index; CI, confidence interval; HR, hazard ratio; NT, Northern Territory; and RHD, rheumatic heart disease.
*Only for those with diagnosis year of 2004 or later (n=772).
†P<0.05.
Rheumatic Heart Disease: Lessons From Linked Data


ORIGINAL RESEARCH
ARTICLE

Table 8. Prevalence of Selected Comorbidities and Health Behavior–Related Hospitalizations Among Indigenous (Urban and Remote) and Non-Indigenous RHD Patients (≥18 Years Old)

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Indigenous, %</th>
<th>Non-Indigenous (n=66), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditions</td>
<td>Urban (n=128)</td>
<td>Remote (n=541)</td>
</tr>
<tr>
<td>Selected comorbidities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chronic obstructive pulmonary disease</td>
<td>14.1†</td>
<td>6.5</td>
</tr>
<tr>
<td>Congestive heart failure</td>
<td>16.4†</td>
<td>9.8</td>
</tr>
<tr>
<td>Diabetes mellitus</td>
<td>13.3†</td>
<td>6.8</td>
</tr>
<tr>
<td>Diabetes mellitus with complications</td>
<td>10.2†</td>
<td>4.4</td>
</tr>
<tr>
<td>Liver disease (mild)</td>
<td>4.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Liver disease (moderate/severe)</td>
<td>0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>Renal failure</td>
<td>7.8</td>
<td>4.3</td>
</tr>
</tbody>
</table>

Hospitalizations that indicate high-risk behavior or environment

<table>
<thead>
<tr>
<th></th>
<th>Indigenous, %</th>
<th>Non-Indigenous (n=66), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assault</td>
<td>21.9†</td>
<td>10.0</td>
</tr>
<tr>
<td>Hazardous alcohol use</td>
<td>25.8†</td>
<td>10.2</td>
</tr>
<tr>
<td>Transportation accident</td>
<td>7.8†</td>
<td>15.3</td>
</tr>
</tbody>
</table>

RHH indicates rheumatic heart disease.
*Significant difference between urban- and remote-residing Indigenous RHD patients (P<0.01).
†Significant difference between urban- and remote-residing Indigenous RHD patients (P<0.05).
‡Significant difference between Indigenous and non-Indigenous RHD patients (P<0.01).

Table 9. Multivariable Cox Regression for Mortality of RHD Patients (≥18 Years Old) With Adjustment for Different Comorbidities

<table>
<thead>
<tr>
<th>Exposure Variables</th>
<th>Baseline HR (95% CI)</th>
<th>Plus Renal Failure HR (95% CI)</th>
<th>Plus Hazardous Alcohol Use HR (95% CI)</th>
<th>Plus CCI HR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous</td>
<td>6.55 (2.45–17.51)*</td>
<td>4.64 (1.72–12.49)*</td>
<td>3.87 (1.41–10.57)*</td>
<td>3.79 (1.39–10.36)*</td>
</tr>
<tr>
<td>Urban vs remote</td>
<td>2.03 (1.26–3.27)*</td>
<td>1.76 (1.09–2.85)*</td>
<td>1.54 (0.94–2.54)</td>
<td>1.52 (0.92–2.50)</td>
</tr>
<tr>
<td>Male vs female</td>
<td>1.73 (1.13–2.65)*</td>
<td>1.71 (1.11–2.62)*</td>
<td>1.63 (1.06–2.51)*</td>
<td>1.67 (1.08–2.57)*</td>
</tr>
<tr>
<td>Age (per year of age)</td>
<td>1.05 (1.04–1.07)*</td>
<td>1.04 (1.03–1.06)*</td>
<td>1.04 (1.03–1.06)*</td>
<td>1.04 (1.02–1.06)*</td>
</tr>
<tr>
<td>Diagnosis year (per year)</td>
<td>0.99 (0.93–1.05)</td>
<td>0.99 (0.92–1.05)</td>
<td>0.97 (0.91–1.04)</td>
<td>0.96 (0.90–1.03)</td>
</tr>
<tr>
<td>Renal failure</td>
<td>4.48 (2.58–7.77)*</td>
<td>4.16 (2.38–7.25)*</td>
<td>4.16 (2.38–7.25)*</td>
<td>3.61 (2.01–6.48)*</td>
</tr>
<tr>
<td>Hazardous alcohol use</td>
<td>1.94 (1.10–3.40)*</td>
<td>1.78 (1.01–3.15)*</td>
<td>1.78 (1.01–3.15)*</td>
<td>1.61 (0.89–2.93)</td>
</tr>
<tr>
<td>CCI score 1 vs 0</td>
<td>1.37 (0.82–2.29)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCI score ≥2 vs 0</td>
<td></td>
<td></td>
<td></td>
<td>1.61 (0.89–2.93)</td>
</tr>
</tbody>
</table>

CCI indicates Charlson Comorbidity Index; CI, confidence interval; HR, hazard ratio; and RHD, rheumatic heart disease.
*P<0.05.
rence risk in each year after ARF diagnosis (9%/y) but found that the risk fell to 0 by 10 years. These findings support the longer secondary prophylaxis duration recommended in our national guidelines but indicate that continuing low-level risk may persist thereafter, again emphasizing the fundamental importance of broader control strategies than secondary prophylaxis alone. Recommending penicillin for >10 years would be difficult to justify given the need to balance morbidity related to needle administration against the low risk of further recurrences more than a decade after the last ARF episode.

This study demonstrates that although there are guidelines stating that all patients with suspected ARF should be hospitalized as soon as possible, only 56.5% of patients were admitted to hospital for the treatment of ARF or RHD within 14 days of the onset of their first ARF episode. Early hospitalization facilitates specialist review, exclusion of differential diagnosis, confirmation of the ARF diagnosis, access to echocardiography, beginning of penicillin and ARF treatments (eg, salicylates for ARF arthritis or arthralgia), and provision of education for the patient and their family about ARF/RHD. Although treatment might have been initiated in primary care, these findings suggest delays in the initiation of treatment for ARF patients that might have contributed to the early onset or progression of heart valve damage. This study demonstrates the importance of the use of routine linkage of RHD Register and hospital data to investigate this missing link, consistent with another study suggesting the effectiveness of using hospitalization data to identify ARF/RHD cases not documented in the RHD Register. Even in a jurisdiction with an effective and long-lasting register, our study has demonstrated that data are incomplete and that a linkage approach can supplement additional data on both outcomes and case ascertainment.

Our finding of higher mortality in older and Indigenous RHD patients is consistent with previous studies. In the multivariable analysis, male patients had a higher mortality despite having a similar incidence rate for all of the 4 complications compared with female patients. Although it was previously shown that males have a lower ARF and RHD incidence than females, our study found that when they do have RHD, male patients are more likely to die. Other factors may contribute to the higher mortality of male RHD patients, such as lower adherence to secondary prophylaxis, a higher prevalence of comorbidities, or hazardous alcohol use. The findings of a similar hospitalization rate for male and female ARF patients (mainly children) but a lower proportion of male RHD patients (mainly young adults) receiving RHD treatment at hospital suggest different health-related behaviors of the 2 sexes at different phases of their lives.

Our finding that the highest incidence of all 4 complications occurred in the first year after diagnosis highlights the importance of prompt action plans within this time frame to prevent or manage these complications in RHD patients (who were mainly young people with a median age at diagnosis of 21 years) because they contributed to higher death rates. Atrial fibrillation and heart failure were more common among RHD patients (5% and 8%, respectively, in first year; 13% and 19% within 10 years), whereas endocarditis and stroke were less common but still occurred in 4% of RHD patients within 10 years. Unfortunately, the development of complications was not less common in those diagnosed with RHD in more recent years.

Progression from ARF to RHD was more common for remote than for urban residents, but for RHD patients, urban residents had higher rates of atrial fibrillation and heart failure and higher mortality even after adjustment for comorbidities and hazardous alcohol use. Lower identification of atrial fibrillation and heart failure among remote residents could be an artifact of lower access to inpatient care for them, but this seems unlikely because hospitalization rates were similar for remote and urban RHD patients. Alcohol is less accessible or not accessible at all in remote communities; thus, individuals labeled as having hazardous alcohol use, who mostly live in remote locations and drink only when visiting town, would have lower risk than urban-based individuals with hazardous alcohol use. In addition, the death data, being the most reliable source of data (compared with hospital and register data), also showed that remote RHD patients have lower mortality, which is consistent with the findings of lower-than-expected cardiovascular morbidity and all-cause mortality for people residing in remote Aboriginal communities.

Because this is the first study to link hospital data with register data to investigate the incidence of long-term RHD complications, there were no available studies with which to compare our results. This study may have underestimated the incidence of complications and mortality among RHD patients because some who developed a complication might not have been admitted to hospital, either because the complication was very mild or because it was very severe and they died before arriving at hospital. The latter possibility could be investigated by linking RHD Register and hospital inpatient data with the National Death Index (which includes cause of death) to identify RHD patients who died of a complication before hospital admission; this was beyond the scope of the present study.

Our study demonstrated the high prevalence of comorbidities in RHD patients and their role in adverse outcomes. The presence of comorbidities was associated with a higher incidence of atrial fibrillation, endocarditis, and heart failure (but not stroke) and higher mortality. In particular, renal failure and hazardous alcohol use were more common among Indigenous than non-Indigenous patients, which accounted for 28% of the excess mortality of Indigenous patients (HR decreases from 6.55 to 3.87). The finding of the adverse effect of kidney failure
on Indigenous RHD patients is consistent with the finding from another study that kidney failure is the only chronic disease that is associated with 30-day and long-term mortality after RHD-related valve surgery, particularly for Indigenous RHD patients (30-day mortality odds ratio, 14.1; 95% confidence interval, 1.0–200.0). Between 2007 and 2009, renal failure was the main contributor (19%) to deaths of people with RHD. Hazardous alcohol use was the other comorbidity that contributed to Indigenous/non-Indigenous and urban/remote differentials in adverse outcomes. This is particularly relevant to NT, which has the highest alcohol consumption rates in Australia. In a 2012–2013 national survey, 30.5% of NT Indigenous adults reported alcohol consumption at risky or high-risk levels (28.4% for remote residents, 35.3% for urban residents) compared with 16.1% of NT non-Indigenous adults in a comparable national survey in 2010. Our finding that 25.8% of urban Indigenous adult RHD patients had been admitted to hospital for a condition related to hazardous alcohol consumption is consistent with these high consumption levels.

Our study suggests that hazardous alcohol use is the greatest contributor (among factors that we could investigate) to the urban/remote differentials in mortality. RHD patients with hazardous alcohol use are almost twice as likely to die even after adjustment for other comorbidities. In this case, hazardous alcohol use should not be seen only as a behavior. Rather, it could be also seen as a proxy for other unmeasured variables that are causing it. To complicate things, hazardous alcohol use could also affect other chronic diseases and other unfavorable health-related behaviors, leading to poorer outcomes.

Even after adjustment for comorbidities and hazardous alcohol use, Indigenous RHD patients are still >3 times more likely to die than non-Indigenous patients, indicating that there are other unmeasured factors contributing to this disparity that need further research. No data were available for this study on socioeconomic conditions, which may be major contributors to this disparity. Indigenous Australians, particularly in the NT, have much poorer socioeconomic conditions, including housing, than other Australians. An NT study found that socioeconomic disadvantage accounted for more than one third of the gap between Indigenous and non-Indigenous life expectancy and that >60% of the gap was explained by the combined effects of socioeconomic disadvantage plus smoking, hazardous alcohol use, obesity, pollution, and intimate partner violence.

Unlike in other Australians, ARF is a common childhood disease of Indigenous Australians, and RHD is a dangerous, sometimes fatal disease among their children and young adults. The NT RHD Register, the most comprehensive data on ARF and RHD available in Australia, has previously been used to demonstrate the very high incidence of and mortality from ARF and RHD among the NT Indigenous population. By linking the RHD Register to hospital and death records, this study has demonstrated the high frequency of serious complications in the years after the onset of RHD and the deleterious effect of the higher prevalence of chronic disease comorbidity, particularly chronic kidney disease and hazardous alcohol use, among Indigenous people (including those with RHD) on both the development of complications and survival. ARF incidence and RHD incidence have not declined in recent years. It is not clear whether this is a result of inadequate compliance with secondary prophylaxis or of the limitations of secondary prophylaxis in the absence of improvements in housing and other socioeconomic conditions. The RHD Register includes data on the administration of secondary prophylaxis that could be used to investigate the effectiveness of secondary prophylaxis, which needs to be done urgently. Data are not so readily available to investigate socioeconomic conditions; data on housing also need to be investigated with similar urgency.

**ACKNOWLEDGMENTS**

We acknowledge the support of the NT Department of Health and thank the staff of the NT Department of Health Data Management and System Reporting Branch for extracting and linking the hospital inpatient data and mortality data; Cath Milne from the NT RHD control program for extracting the RHD Register data; and Shu Qin Li from the Department of Health Health Gains Planning Branch for linking the RHD Register to the hospital inpatient data.

**SOURCES OF FUNDING**

V.Y.F. He was supported by a University Postgraduate Research Scholarship from Charles Darwin University. Dr Ralph is supported by the National Health and Medical Research Council of Australia (GNT1113638). J.L. de Dassel is supported by an Australian Postgraduate Award scholarship. Dr Zhao, Dr Roberts, Dr Currie, M. Fittock, and Dr Edwards are supported by the NT Department of Health.

**DISCLOSURES**

None.

**AFFILIATIONS**

From Menzies School of Health Research, Charles Darwin University, Darwin, Australia (V.Y.F.H., J.R.C., A.P.R., K.R., J.L.d.D., B.J.C.); Royal Darwin Hospital (A.P.R., K.R., B.J.C., K.N.E.) and Health Gains Planning Branch (Y.Z.), Northern Territory Government Department of Health, Darwin, Australia; Northern Territory Rheumatic Heart Disease Control Program, Centre for Disease Control, NT Department of Health, Darwin, Australia (M.F., K.N.E.); Telethon Kids Institute, University of Western Australia, Perth, Australia (J.R.C.); and Princess Margaret Hospital for Children, Perth, Australia (J.R.C.).
FOOTNOTES
Received December 16, 2015; accepted May 27, 2016. Circulation is available at http://circ.ahajournals.org.

REFERENCES


Long-Term Outcomes From Acute Rheumatic Fever and Rheumatic Heart Disease: A Data-Linkage and Survival Analysis Approach
Vincent Y.F. He, John R. Condon, Anna P. Ralph, Yuejen Zhao, Kathryn Roberts, Jessica L. de Dassel, Bart J. Currie, Marea Fittock, Keith N. Edwards and Jonathan R. Carapetis

Circulation. 2016;134:222-232; originally published online July 12, 2016;
doi: 10.1161/CIRCULATIONAHA.115.020966
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2016 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/134/3/222
Free via Open Access

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/
CHAPTER SIX: CONCLUDING REMARKS

In this final chapter, I will discuss the thesis’s main findings, contribution to the methodology (in the use of survival analysis and administrative data) and future application of advanced survival analysis methods in analysing administration data.

6.1 Summary of main findings

The main findings of this thesis are as follows:

6.1.1 Long-term health outcomes (mortality, recurrence, complications) are poorer for Indigenous than non-Indigenous people with serious illnesses in the NT:

a) Indigenous sepsis patients have higher excess mortality before and after adjusting for age, sex, year of follow-up, severe sepsis, bacteraemia and chronic disease comorbidity. Amongst the sepsis patients less than 65 years old, Indigenous people have lower five-year relative survival than non-Indigenous patients, with this difference being most evident in young female patients (15-44 years old). Excess mortality persisted for two years after the sepsis episode, and for three years in patients with severe sepsis.

b) Indigenous stroke patients had higher stroke recurrence and long-term death rate than non-Indigenous patients, before and after adjusting for age, sex, stroke subtype, year of diagnosis, remoteness of residence, presence of atrial fibrillation, hypercholesterolaemia and chronic disease comorbidity.

c) Indigenous cancer patients have lower net survival than non-Indigenous patients for breast, colorectal, head and neck cancers, after adjusting for age and sex. In the presence of competing risk, Indigenous cancer patients with all types of cancer have higher probability of cancer death and (except for those with lung cancer and head and neck cancers, both smoking-related cancers) higher probability of non-cancer death.

d) Indigenous RHD patients have higher mortality than non-Indigenous patients before and after adjusting for age, sex, remoteness of residence, year of diagnosis, comorbidities and hazardous alcohol use.
6.1.2 The prevalence of chronic disease comorbidities is higher for Indigenous than non-Indigenous patients with some (but not all) serious illnesses:

a) Indigenous stroke patients have higher prevalence of hypertensive diseases, rheumatic heart diseases and all the comorbidities in the modified Charlson Comorbidities Index except rheumatoid disease, peptic ulcer, metastatic cancer and AIDS.

b) In contrast to stroke patients, the proportion of RHD patients with any comorbidity (CCI≥1) was the same for Indigenous and non-Indigenous RHD patients, but Indigenous RHD patients are much younger than stroke patients (the median age at diagnosis for Indigenous RHD patients was 21 years compared to 53 years for Indigenous stroke patients). Indigenous RHD patients had higher prevalence of hospitalisation due to hazardous alcohol use and assault than non-Indigenous RHD patients, and among Indigenous RHD patients, prevalence of any comorbidity (CCI≥1) was higher for urban than remote residents, as was the prevalence of many individual conditions such as hazardous alcohol use and assault.

6.1.3 Comorbidities are partly responsible for the worse outcomes after serious illness for Indigenous compared with non-Indigenous patients:

a) Comorbidities explained part of the Indigenous disparity in long-term survival after a stroke, but did not explain the disparity in stroke recurrence.

b) Renal failure and hazardous alcohol use accounted for 28% the excess mortality of Indigenous RHD patients (and the excess mortality of those living in urban compared with remote areas). The residual Indigenous survival disparity in RHD patients, which persisted after accounting for comorbidities and hazardous alcohol use, suggests that other factors contribute to mortality, warranting further research.
6.2 Prevention is better than cure

The high prevalence of chronic disease in patients with serious illness (both Indigenous and non-Indigenous), and the contribution that the higher prevalence of comorbidities makes, to the poorer long-term outcomes of Indigenous compared with non-Indigenous patients with serious illness, demonstrates the importance of prevention. Particularly given that chronic diseases have already inflicted damage before patients suffer an episode of serious, acute illness and their comorbidities undermined the effectiveness of acute care services. As many serious chronic diseases share common preventable risk factors (1,2), the World Health Organisation (2000) has recommended that “addressing the major risk factors should be given the highest priority in the global strategy for the prevention and control of non-communicable diseases” (3). The NT PCDS (4,5) recognised the importance of socioenvironmental factors as the underlying causes of many risk factors for chronic diseases, stating that “[c]hronic disease and their risk factors are also inextricably linked with the broader socioeconomic determinants of health and quality of life, particularly education and employment” and “[l]ifestyle choices are often more reflective of unrelenting socioenvironmental constraints rather than personal preferences”. Support and commitment from all levels of the government (federal, state and local), along with non-government organisations, academia, the private sector and community groups, is essential in cultivating an environment which supports good health and primary prevention and assists people to make healthy choices and lead healthier lives (6) (i.e. school policies support healthy eating, legislation bans smoking in enclosed environments, policies control the promotion of alcohol to young people).
Although it was recognised that health promotion and primary prevention are important, there are inadequate resources allocated to this strategy. An evaluation of the NT PCDS conducted in 2007 (7) identified areas that needed improvement, including the limited resources dedicated to health promotion and prevention\(^1\). In 2009, the NT Department of Health developed the NT Chronic Conditions Prevention and Management Strategy (CCMPS) 2010-2020(8) in partnership with non-Government health services including Aboriginal Community-Controlled Health Organisations. The CCMPS identified primary prevention as one of its eight key actions areas, but the CCPMS Implementation Plan 2010-2012 stated that “primary prevention continues to have a low priority in budget allocations” and support should be given to the “allocation of appropriate funding and resources to promote population-based prevention strategies” (9).

In a time of fiscal constraints, more information and evidence is needed to guide public health policy and implementation, including resource allocation. Better use of existing data (health administrative data) can be an efficient and timely means to obtain such information, particularly when monitoring and evaluating the effectiveness of existing services and innovations. For example, using linked administrative dataset (primary health care and hospitalisation data), one NT study (10) in 2013 linking primary care with hospital data found that provision of primary care in remote Indigenous communities results in cost-savings to public hospitals (investing $1 in primary care in remote Indigenous communities could save $3.95-$11.75 in hospital costs) and health benefits to individual patients. Other studies in the NT (11,12) demonstrated a U-shaped relationship between primary healthcare visits and hospitalisations (instead of a straight-line relationship), suggesting that there is an optimal level of primary healthcare that will result in the minimum hospitalisations for remote Indigenous Australians.

\(^1\) There is evidence of a depletion in the infrastructure, dedicated staff, and funding in the important area of health promotion and prevention, as well as a reduction in the number of prevention and health promotion programs since the development of the PCDS… Increasing the focus on prevention and health promotion programs is a key issue for the future and will require genuine investment and specialised staff with diverse skills. (7)
6.3 Contribution to methodology: the use of survival analysis and administrative data

In this section, I will discuss how the four different studies use survival analysis and administrative data in a novel way to investigate the long-term health outcomes of Indigenous and non-Indigenous Australians suffering from four serious conditions: sepsis, stroke, cancer and rheumatic heart disease.

6.3.1 Sepsis study

The traditional focus of critical care has been on improving short-term mortality; in previous studies of outcomes after acute sepsis, the most common primary endpoint has been 28-day or 90-day mortality. However, survivors might develop long-term consequences that persist beyond the intensive care unit and their hospital stay, exacerbated by the complex relationship between acute illness and comorbid conditions (which are both a risk factor for, and modifier of, the acute illness). It is therefore important to assess long-term endpoints as well. This is particularly important for a population with a high prevalence of chronic diseases such as Indigenous Australians. The finding of higher excess mortality among Indigenous sepsis patients, even after adjustment for chronic disease comorbidity and bacteraemia, suggests that there are other unmeasured factors contributing to this disparity that need further research.

In the sepsis study, long-term outcomes were assessed using the relative survival methodology that was developed to measure long-term survival of cancer patients. Relative survival compares the survival of the patient cohort over time with that of the cancer patients’ background, without the need for cause of death information. The finding that excess mortality persisted for two years after the sepsis episode, and for three years in the subgroup with severe sepsis, suggests that efforts to improve outcomes for sepsis patients should focus on endpoints as long as three years post discharge, rather than be limited to the first 90 days. The use of a generalised linear model with Poisson error structure, enabled multivariable analysis of excess mortality that identified particular risk factors for lower long-term survival (i.e. lower in male, Indigenous and older patients and those with chronic disease comorbidity and sepsis-related organ dysfunction at presentation). Relative survival methods could also be used to analyse long-term outcomes for patients with other acute conditions such as cardiac...
failure, stroke, other severe infections (i.e. bone and joint infection), falls and injuries. For example, relative survival has been used to analyse long-term survival after hospital for hip fracture in Australia (13). Another novel aspect of this study was the use of a cohort of sepsis patients from a previous study (conducted in 2007-2008) and data linkage to administrative data in 2012 to obtain information about deaths - an efficient utilisation of data from a previous study that enabled examination of long-term outcomes based on clinical data. Similar examination of long-term outcomes for cohorts of patients with particular diseases assembled for short-term research projects could be a cost-effective way to provide valuable information to inform clinical practices or health policies. Previous survey studies that collected behaviours (i.e. smoking or alcohol use) or clinical variables (i.e. blood pressure or cholesterol level) could also be linked to administrative data (death register or hospital data) to investigate long-term outcomes using survival analysis methods. For example, data from a study (14) conducted in 1988-1989 that measured lifestyle and behavioural factors (i.e. diet, alcohol and smoking habits) in the Western Australia Aboriginal population was linked to hospital, cancer and death records (from 1998-2002) to investigate the effects of health-related behaviours on mortality and morbidity (coronary heart disease); this study found that higher alcohol intake, smoking and adverse dietary choices increased risk of coronary heart disease (HR 2.1, 95% CI: 1.1-4.0) and all-cause mortality (HR 2.3, 95% CI: 1.2-4.2).

6.3.2 Stroke study

Other than selecting the time point, one also has to decide on the type of outcome measures. Studies investigating disease recurrence or complications using administrative data usually measure a single combined endpoint (15,16) (i.e. treating recurrent stroke and death as a single outcome) or co-primary outcomes (analysing recurrent stroke and death separately) using conventional survival analysis methods, which ignore the competing risk. This might not be appropriate for serious illness (i.e. stroke) in which the probability of death might be higher than the primary outcome; in other words, the patients are more likely to die before they suffered from the primary outcome (i.e. recurrent stroke). In studying recurrent stroke, the stroke patients who die before having a recurrent stroke might have been more likely to have a recurrent stroke if they had not died than those patients who did not die and were followed up. If so, treating death as a censoring event rather than a competing risk would
violate the assumption that the time of the primary outcome event is independent of the censoring distribution (17) and the assumption that censoring is “non-informative” (18) (i.e. that participants’ loss to the study is unrelated to their underlying risk). Patients with comorbidity are more likely to experience death from stroke or other diseases before a recurrent stroke. In other words, the competing risk model shows that patients with no or less comorbidities are less likely to die and remain the ‘risk set’ for recurrent stroke, which lead to ‘higher risk’ of recurrent stroke. The competing risk analysis approach is particularly suitable for the analysis of stroke recurrence in Indigenous stroke patients because they have higher mortality from other causes that may bias estimation of their incidence of stroke recurrence. To our knowledge, the stroke study is the first study to use competing risk analysis to investigate the relationship between recurrent stroke and comorbidities of Indigenous Australians. Presenting the probability of recurrent stroke, death and survival as stacked cumulative incidence plots is a novel way to show how the probabilities changes over time for different age groups, which differs between Indigenous and non-Indigenous patients, that may be a more understandable way to communicate risk to patients and clinicians (19).

6.3.3 Cancer study

In studies of cancer survival, different endpoints and prognosis measures may be chosen depending on whether the objective is to provide information at a population or individual level. Net survival is a common measure of cancer prognosis for researchers and policy makers in population-based cancer studies. By removing the effect of competing causes of death, and only measuring the net effect of a cancer diagnosis in the hypothetical scenario where cancer patients die only from their cancer (19-23), net survival focuses on the outcome of cancer treatment, distinguishing between increased survival among cancer patients due to improvements in the treatment of cancer rather than reduction in other competing causes of death (19-21). It is useful for national and international comparisons of cancer survival, where cancer patients’ risk of death may vary considerably between populations because of differences in general population mortality rates. However, such information is less informative from the patient’s perspective. Another prognosis measure, the ‘crude probability of death’ (2-4, 8, 11), which takes into account competing risk,
provides more information at an individual level. This is particularly important for Indigenous Australians who have higher prevalence of chronic diseases (24) and much higher all-cause mortality rates (25,26) than other Australians. This prognosis measure, that takes into account competing risk, enables us to identify whether survival disparities are caused by cancer, non-cancer or both cancer and non-cancer deaths, as well as identify the disparities in both cancer and non-cancer deaths across time. Competing risk analysis found that Indigenous patients were more likely than non-Indigenous to die of non-cancer causes and that, in contrast to cancer deaths, this differential increased rather than decreased with time after diagnosis. Five years after diagnosis, Indigenous cancer patients had higher rates of non-cancer death than non-Indigenous patients and this differential was increasing with time since diagnosis, while there was no longer a differential in cancer deaths. This finding demonstrates an important advantage of the competing risk approach over conventional relative survival in analysing long-term outcomes for Indigenous cancer patients.

In population-based studies of survival after serious illness such as cancer the selection of endpoint is often limited by the availability and reliability of cause of death data. When cause of death data is unavailable, methods that rely on lifetables (such as the Cronin-Feuer method) can be used to estimate crude probability of death.

6.3.4 Rheumatic heart disease study

The data linkage study of RHD outcomes enabled a more comprehensive assessment of outcomes for RHD patients (recurrence, complications, worsening of clinical condition and mortality) than had been possible using data about RHD alone. By linking data from the RHD Register, hospital inpatients and the death register it was also possible to investigate the effect of comorbidities on these adverse outcomes. Presenting the occurrence of these adverse outcomes in terms of cumulative incidence and incidence rate over time since diagnosis is a novel way in which survival analysis is used to identify the conditions that were more serious at different time periods. The incidence of development of RHD was much higher than ARF recurrence for ARF patients. Heart failure and ARF were more common complications for RHD patients than endocarditis and stroke.
6.4 Future application of advanced survival analysis methods in analysing administration data

To allow for a more comprehensive understanding of patients’ prognosis, more advanced survival analysis methods offers some potential benefits. There are three types of survival analysis methods that are relevant and applicable to administrative data, which are areas of further research. They are:

1. Multiple event survival analysis methods, which consider individuals experiencing repeated occurrences of the same type of event (i.e. repeated stroke, heart attacks, acute rheumatic fever or asthma attacks)

2. Multi-state model, which considers individuals experiencing repeated events of the same type and/or events of different types over time (important for understanding the disease course and associated clinical event histories)

3. Marginal structure model, which involves causal inference (i.e. assessing whether an intervention leads to the change).

6.4.1 Multiple event survival analysis methods

Evaluation of acute care should include secondary prevention and disease recurrence, which inflicts greater damage to the patient and additional cost to the health system (27). Multiple event survival analysis methods are well suited for this purpose. They have great potential in the health sector; not only have they been used to investigate recurrent events relating to diseases such as hospitalisation due to heart failure (27-29), they are also useful in improving health services such as understanding rural family physician turnover for rural medical workforce planning (30) and identifying characteristics associated with doctors at high risk of recurrent patient complaints (31). Kelly and Lim (2000) have compared the different models and provided recommendations about the suitability of each model for different types of recurrence data (32).
6.4.2 Multi-state model

Classical survival analysis assumed that subjects experienced a single event, using a single endpoint, which ignored the reality that different kinds of events can occur. To enable a more complete picture of prognosis, disease course and associated clinical event histories, the multi-state model should be more widely used to analyse health administrative data in the future, given the previous successful applications in analysing hospital (33-40) and billing data (41). The multi-state model provides more insights into the disease/recovery process and enables clinicians to obtain more accurate predictions of survival probabilities, which can be easily updated when information on the types of events that the patients have experienced and the treatment they have received become available (42). In the multi-state model, an event marks the transition from one state to the other (i.e. the occurrence of disease is a transition from one state to another, as is disease recurrence, or development of a complication, or death); it is a stochastic process in which subjects occupy a state and move between states described by the event (43,44). The figure below is one example (Figure 6.1). The multi-state model can be used to estimate the transition probabilities (from one state to another), as well as assessing the effect of different prognostic factors on the prognosis, given that the influence of different prognostic factors may vary depending on the subjects’ current state (36,45).

Figure 6.1 Possible disease progression of ARF and RHD patients
6.4.3 Marginal structural model

Policy makers, health professionals and researchers are often interested in understanding how an intervention (e.g. administrating a new drug, or mandatory alcohol treatment) has a causal effect on health. Traditionally, randomised experiments were often used as the ‘golden standard’ to achieve this aim; however in many situations randomised trials are not feasible due to financial or ethical considerations. There is increasing use of observational data (i.e. administrative data) to evaluate interventions that cannot be tested by randomised trial. However, one of the challenges in analysing observation data is the possible presence of confounders, which obscure the relationship between the intervention and its outcome; this is avoided in randomised trials by the randomisation process. Time-dependent confounders are common in health intervention research where there are repeated treatments over time; they affect future treatment and future outcome while also being affected by previous treatment. For example, prior myocardial infarction is a time-dependent confounder in aspirin use (treatment) to reduce myocardial infarction and cardiac death risk (outcome); prior myocardial infarction is a cause of subsequent aspirin use (affecting future treatment) and a risk factor for subsequent death (affecting future outcome), but the occurrence of myocardial infarction is reduced by prior aspirin use (affected by previous treatment). In such situations, conventional statistical methods in analysing observation data produced biased estimates of the treatment effect, because the time-dependent nature of the confounders and treatments was not accounted for. The marginal structural model is one approach to this problem of isolating the association between time-dependent treatment and outcome in the presence of time-dependent confounders based on the theory of counterfactual causal inference (46-50). It has been widely used to analyse various observational data that adjust loss due to follow-up using censoring weights (51-57).
6.5 The whole is greater than the sum of its parts: linked administrative data in multiple government sectors with a culture of information sharing and proactive data release

The saying “the whole is greater than the sum of its parts” by Aristotle has now become a common phrase and has different applications in different settings. This concept is relevant to the relationship between health information and administrative data, in which data linkage is the glue that puts all the ‘parts’ (administrative data) together to bring us closer to the ‘whole’. Health information of an individual in different data sources, recorded in different times and places has more value when considered together (through data linkage) than when considered separately (in each dataset). In reality, we are unable to see the ‘whole’; the best we can do is to work towards the ‘whole’, sharing the knowledge of the past with other people, providing the information of the event and the time that it occurred, providing the platform to improve the future by reflecting on the past.

The projects in this PhD have demonstrated the wealth of information that can be generated from health administrative data, through data linkage, to improve our knowledge about long-term health outcomes of NT Indigenous and non-Indigenous Australians after serious illnesses. This was made possible by the unique patient identifier (the Hospital Registration Number, HRN) used in the NT Department of Health’s Client Master Index (CMI). The NT DoH CMI obtains the demographic information of any client of a NT government clinical service through acute care hospitals, urban and remote community health centres, mental health and other services. Clients accessing any NT DoH health services (hospitals, primary health care, etc) use the same unique patient identifier recorded in the CMI. The use of a unique patient identifier enables deterministic linkage of data (using HRN) from multiple health services, which is a less time-consuming and more accurate data linkage method than probabilistic linkage methods.

There is the potential in the NT to use non-health data sources (e.g. education, housing, justice) to investigate broader socio-environmental issues in relation to health and other human conditions. Administrative data is not only useful in producing information to facilitate better decision-making for
the health sector, but also for education (58,59) and other social services sectors (60,61). The NT could draw experience from the success of the Developmental Pathways in Western Australian (WA) Children Project (DPP). DPP takes a holistic and multidisciplinary approach to investigate risk and protective factors leading to differences in developmental outcomes (i.e. health and wellbeing, education, disability, child abuse and neglect, mental health, intentional self-harm and suicide, and juvenile delinquency) for children and youth, using de-identified linked population-level data from various WA government departments and agencies (education, housing, child protection, corrective services, police and the attorney general) (62-64).

DPP was made possible by the widespread support from participating agencies, and resulted from the collaboration between researchers in the Telethon Kids Institute and 13 WA government departments and agencies (62-64). The linkage between datasets held by different government agencies will not only provide valuable cross-agency data to generate knowledge which informs and enables future policy and prevention strategies, but will also improve the collection, utilisation and reliability of each of the contributing data sources, while enhancing whole-of-government initiatives across multiple government departments and agencies (62-64). The NT can benefit from the experience of WA, where there is “a culture of information sharing and proactive data release” (65), and they have not only overcome the technical challenges, but also the major challenges of data linkage research such as the legislative, organisational, social and political barriers (66).
6.6 Conclusion

This thesis investigated using survival analysis methods for analysing administrative data, to measure long-term health outcomes of Indigenous and non-Indigenous Australians after serious illness in the NT. It found that Indigenous patients have higher levels of comorbidity and suffered disadvantage (compared with non-Indigenous patients) in long-term (but not short-term) survival outcomes, even after adjusting for comorbidities. This reinforced the importance of prevention, given that serious illness has already inflicted damage on the patients and undermined the effectiveness of acute care services. It has demonstrated that survival analysis has great potential to produce useful information from existing data to improve our knowledge about long-term health outcomes. Other than conventional survival analysis, more advanced survival analysis methods are needed for a more comprehensive understanding of the patients’ prognosis. The competing risk models, multiple event survival analysis methods, multi-state model and marginal structural model are relevant survival analysis methods that will produce more useful information from administrative data to improve our knowledge about long-term health outcomes of Indigenous and non-Indigenous Australians, which could facilitate “better decision-making, leading to better health” (67).
6.7 Reference


2. AIHW 2012. Risk factors contributing to chronic disease. Cat. no. PHE 157. Canberra: AIHW.


**CHAPTER SEVEN: APPENDIX**

**Table 7.1: A summary of different survival analysis methods used in Chapter 4**

<table>
<thead>
<tr>
<th>Prognosis statistics</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Use cause of death data</td>
</tr>
<tr>
<td><strong>Net survival (Excludes Competing risks)</strong></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>Cause-specific survival</td>
</tr>
<tr>
<td>Possible calculation method</td>
<td>Kaplan-Mier</td>
</tr>
<tr>
<td>Possible multivariable analysis method</td>
<td>Cox proportional hazard model</td>
</tr>
<tr>
<td><strong>Crude probability (Includes competing risk)</strong></td>
<td></td>
</tr>
<tr>
<td>Measure</td>
<td>Crude probability of death</td>
</tr>
<tr>
<td>Possible calculation method</td>
<td>Fine-gray method (CR analysis)</td>
</tr>
<tr>
<td>Possible Multivariable analysis method</td>
<td>Fine-gray model (CR analysis)</td>
</tr>
</tbody>
</table>