UNIVERSITY OF Southampton

University of Southampton Research Repository
ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

http://eprints.soton.ac.uk

UNIVERSITY OF SOUTHAMPTON

Faculty of Law, Arts and Social Sciences

School of Social Sciences

"Health Inequalities in New Zealand: An Examination of Mortality and Hospital Utilisation trends, with reference to the Compression of Morbidity Hypothesis"

by

Ngaire Anne Coombs

Thesis for the degree of Doctor of Philosophy

June 2011

ABSTRACT

This thesis examines health inequalities by area-level socioeconomic deprivation, and health in later life in New Zealand. It identifies whether expansion or compression of morbidity is occurring at the end of life. It asks if overall morbidity at a population level is likely to increase or decrease in future as life expectancy increases, and if the same trend is seen for more and less deprived areas. The focus of this research is the identification and dissemination of mortality and morbidity patterns present in two large datasets, using powerful but relatively simple techniques. Large administrative datasets on morbidity and public hospital discharges in New Zealand between 1974 and 2006 are used in the analyses

The thesis consists of three papers. Each paper uses the same datasets, but addresses separate research questions using different methods. The first paper is an exploratory analysis of age-specific and age-standardised mortality and hospital bed day rates, which are used as a proxy for morbidity. The second paper explores lifetime morbidity by using period-prevalence life table functions including Hospital Utilisation Expectancies: a variation of health expectancies. The third paper uses individual record linkage between the mortality and hospital datasets to examine hospital use in the last few months of life.

Hospital bed day and mortality rates declined over the time period, and convergence was seen between more and less deprived areas. Individuals at the oldest ages (80 years and over) saw little variation in hospital or mortality rates by area deprivation. Strong evidence for compression of morbidity was observed, particularly at older ages. This was in the absence of evidence for rectangularisation of the survival curve, considered by some to be a prerequisite for compression of morbidity. Rectangularisation of the survival curve would be denoted by life expectancy increases slowing, indicating the nearing of a limit to life expectancy. Instead, compression of morbidity was achieved through a decline in the severity of morbidity in the months prior to death. No evidence of a change in the point at onset of morbidity prior to death was observed. There was however some evidence that the decline in hospital utilisation prior to death (particularly for deaths at older ages) may be partly artefactual. Further research using a different measure of morbidity is required to either support or disprove this theory.

DECLARATION OF AUTHORSHIP

I, Ngaire Coombs, declare that the thesis entitled “Health Inequalities in New Zealand: An Examination of Mortality and Hospital Utilisation trends, with reference to the Compression of Morbidity Hypothesis” and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. None of this work has been published before submission.

Signed:

Date:

ACKNOWLEDGEMENTS

I wish to acknowledge the assistance of the Population Studies Centre (PSC) at the University of Waikato, especially Prof. Ian Pool, without whom this research would not have been possible. In late 2008, following a disastrous and time consuming attempt to access New Zealand health and nutrition survey data, the PSC hosted me for three months during which time I contributed to a monograph on hospital utilisation and mortality trends in New Zealand. I then went on to use the same datasets in this study, kindly provided by the PSC. Ian Pool and his successor Prof. Natalie Jackson continue to provide ongoing support and advice, for which I am very grateful.

It is very unusual to be able to work with such a detailed and comprehensive dataset as was used in this study. The data was however of a very sensitive nature. Following the advice of Prof. 'Mac' MacDonald I contacted the Southampton Social Statistics Research Institute (S3RI), who allowed me to store the data in their secure data lab. The use of this facility is much appreciated. The size of the datasets and the format in which they were stored were also problematic. The data were initially stored in Access format, which necessitated a steep learning curve and the assistance of friend and Visual Basic (VB) maestro Steve Pilbeam.

Other challenges were largely methodological. For the third paper, uncertain how to quantify the point at onset of morbidity, Dr. Andy Hinde suggested the conversion of the data from person-event to person-month format, and the use of logistic regression. Conducting this analysis required Steve Pilbeam's VB expertise and ingenuity to automatically convert the SPSS output from hundreds of logistic regression models into just a few neat, concise, and intuitive graphs. I am also grateful to Prof. Peter Smith, whose advice rescued this section of the thesis after a panic about the distribution of the residuals.

Thanks go to my parents and to my Grandma, who have been hugely supportive of my work. Without both their belief in my ability and financial support I would not have reached the point where I could study for a doctorate, let alone complete one. While she will sadly not be able to see the completion of my PhD, my Grandma was intensely proud of my achievements and was certain that I would finish it successfully, something I often doubted.

Lastly, but perhaps most importantly I’d like to thank my supervisors Prof. Zoë Matthews and Dr. Sabu Padmadas, for their advice and unfaltering optimism over the last three and a half years.

CONTENTS

INTRODUCTION **pxxvii-xxxiv**

- The New Zealand Population Structure and Composition pxviii
- Structure of the New Zealand Health System, and Relevance to Morbidity Research pxxx
- The Implications of Health Transition Theories on Health in Later Life pxxxii
- Main Objectives and Research Questions pxxxiii
- Structure of the Papers pxxxiv

PAPER ONE – Trends in Mortality and Hospital Utilisation **p1-84**

Research Questions & Abstract p1

Section 1 – Literature Review: Socioeconomic Health Inequalities and Hospital Utilisation **p3-18**

1.1 Introduction p3
1.2 Area Deprivation p3
1.2.1 Types of Area Deprivation Measures p6
1.2.2 NZ*dep* p6
1.2.3 Is Area Deprivation a Relevant Indicator of Individual Deprivation? p8
1.2.4 Ethnicity as a Confounder of Area Deprivation p10
1.3 Health and Hospital Utilisation p13
1.3.1 Health p13
1.3.2 Hospital Utilisation as a Proxy for Morbidity p15
1.3.3 Hospital Utilisation and Area Deprivation p16
1.3.4 Limitations of Hospital Data p16
1.5 Conclusion p18

Section 2 – Data and Methods: Preparation of the Datasets and Calculation of Mortality and Hospital Bed Day Rates **p19-34**

2.1 Introduction p19
2.2 Datasets p19
2.2.1 Hospital Data p22
2.2.2 Mortality Data p23
2.2.3 Population Data p24
2.2.4 Area Deprivation Data p24
2.2.5 International Classification of Disease (ICD) Data p25
2.3 Hospital Bed Days and Discharges, Which is the Best Measure? p25
2.4 Filtering the Hospital Data p27
2.5 *De jure* and *de facto* Population Denominators p31
2.6 Calculation of Hospital Bed Day and Mortality Rates p31
2.7 Calculation of Confidence Intervals for Mortality Rates p32
2.8 Ethical Considerations p33
2.9 Conclusion p34

Section 3 – Results: Mortality and Hospital Bed Day Rates by Deprivation Decile **p35-68**

3.1 Introduction p35
3.2 National Mortality Rates (1974-2006) p35
3.2.1 Age-Specific Mortality Rates, 1974-2006 p35
3.2.2 Standardised Mortality Rates, 1974-2006 p37
3.2.3 Age-Specific Mortality Rates by Cause, 1974-2006 p38
3.2.4 Standardised Mortality Rates by Cause, 1974-2006 p40
3.2.5 Conclusions from National Mortality Analysis, 1974-2006 p42
3.3 National Hospital Bed Day Rates (1974-2006) p43
3.3.1 Age-Specific Hospital Bed Day Rates, 1974-2006 p43
3.3.2 Standardised Hospital Bed Day Rates, 1974-2006 p45
3.3.3 Age-Specific Hospital Bed Day Rates by Cause, 1974-2006 p46

3.3.4 Standardised Hospital Bed Day Rates by Cause, 1974-2006 p49
3.3.5 Conclusions from Hospital Bed Day Rate Analyses, 1974-2006 p50
3.4 Mortality rates by Area Deprivation (1991-2006) p50
3.4.1 Age-Specific Mortality Rates by Deprivation, 1991-2006 p51
3.4.2 Standardised Mortality Rates by Area Deprivation, 1991-2006 p51
3.4.3 Age-Specific Mortality Rates by Cause and Area Deprivation, 1991-2006 p53
3.4.4 Standardised Mortality Rates by Cause and Area Deprivation, 1991-2006 p56
3.4.5 Conclusions from Area Deprivation Mortality Analysis, 1991-2006 p59
3.5 Hospital Bed Day Rates by Area Deprivation (1991-2006) p59
3.5.1 Age-Specific Hospital Bed Day Rates by Area Deprivation, 1991-2006 p60
3.5.2 Standardised Hospital Bed Day Rates by Area Deprivation, 1991-2006 p60
3.5.3 Age-Specific Hospital Bed Day Rates by Cause and Area Deprivation, 1991-2006 p61
3.5.4 Standardised Hospital Bed Day Rates by Cause and Area Deprivation, 1991-2006 p65
3.5.5 Conclusions from Area Deprivation Hospital Bed Day Rate Analyses, 1991-2006 p67
3.6 Conclusion p67

Section 4 – Discussion **p69-84**
4.1 Introduction p69
4.2 National Analysis, Discussion of Results p70
4.3 Area Deprivation Analysis, Discussion of Results p72
4.4 Regional Variations in Mortality and Hospital Bed Day p74

Rates, a Comparison with Previous Research
4.5 Relevance of Findings to the Research Questions p81
4.6 Limitations of this Research p82
4.7 Conclusion p83

PAPER TWO – Examination of Compression of Morbidity using Life Table Techniques **p85-140**

Research Questions & Abstract p85

Section 1 – Literature review: Compression and Expansion of Morbidity Hypotheses **p87-102**

1.1 Introduction p87
1.2 Epidemiologic Transition p87
1.3 Compression and Expansion of Morbidity Hypotheses p89
1.3.1 Compression of Morbidity Hypothesis p89
1.3.2 Expansion of Morbidity Hypothesis p91
1.3.3 Dynamic Equilibrium p93
1.3.4 Evidence for Compression/Expansion of Morbidity p93
1.3.5 The Human Life Span p95
1.3.6 Rectangularisation of the Survival Curve p97
1.3.7 Inequalities in Health p98
1.3.8 Methodological Issues p99
1.4 Conclusion p100

Section 2 – Data and Methods: Calculation of Hospital Utilisation Expectancies **p103-108**

2.1 Introduction p103
2.2 Sullivan's Method p104
2.3 Previous Research using Hospital Utilisation Expectancies p104
2.4 Hospital Utilisation Expectancies and the Compression/Expansion of Morbidity Debate p105
2.5 Calculation of Hospital Utilisation Expectancies p106
2.6 Conclusion p107

Section 3 – Results: Evidence for/against Compression of Morbidity from Hospital Utilsation Expectancy Analysis **p109-130**

3.1 Introduction p109

3.2 National Life Expectancy and HUEs (1974-2006) p109

3.2.1 National Results, Life Expectancy and HUEs Over Time, 1974-2006 p109

3.2.2 National Results, Life Expectancy and HUEs by Age, 1974-2006 p113

3.2.3 National Results, Life Expectancy and HUEs by Cause, 1974-2006 p115

3.2.4 Conclusions from National Analysis of Life Expectancy and HUEs, 1974-2006 p117

3.3 Life Expectancy and HUEs by Area Deprivation (1991-2006) p118

3.3.1 Area Deprivation Results, Life Expectancy and HUEs Over Time, 1991-2006 p118

3.3.2 Area Deprivation Results, Life Expectancy and HUEs by Age, 1991-2006 p122

3.3.3 Area Deprivation Results, Life Expectancy and HUEs by Cause, 1991-2006 p127

3.3.4 Conclusions from Area Deprivation Analysis of Life Expectancy and HUEs, 1991-2006 p129

3.4 Conclusion p129

Section 4 – Discussion **p131-140**

4.1 Introduction p131

4.2 National Analysis, Discussion of Results p131

4.3 Area Deprivation Analysis, Discussion of Results p132

4.4 Private Health Insurance in New Zealand, Implications for Public Hospital Use by Area Deprivation p134

4.5 Relevance of Findings to the Research Questions p138

4.6 Limitations of this Research p139

4.7 Conclusion p140

PAPER THREE – Examination of Compression of Morbidity using Longitudinal Techniques **p141-208**

Research Questions & Abstract p141

Section 1 – Literature review: Ethics of using Medical Records in Research and Individual Record Linkage **p143-156**

1.1 Introduction p143
1.2 Use of Medical Records in Research: Privacy and Consent p144
1.3 Benefits and Limitations of Medical Record Linkage p147
1.4 Western Australia Data Linkage System: A Case Study p150
1.5 Previous Research using Linked Health Data to Examine Hospital Utilisation Prior to Death p152
1.6 Conclusion p155

Section 2 – Data and Methods: Conducting Record Linkage and Description of Two Longitudinal Techniques **p157-170**

2.1 Introduction p157
2.2 Conceptual Diagram p158
2.3 Reference Point Modelling p159
2.4 Person-Month Logistic Regression p162
2.5 Data Preparation p167
2.6 Conclusion p169

Section 3 – Results: Evidence for/against Compression of Morbidity from Reference Point Modelling and Person-Month Logistic Regression Analyses **p171-196**

3.1 Introduction p171
3.2 Reference Point Modelling p171
3.2.1 National Results, Hospital Bed Days in Months Prior to Death, 1990-2006 p172
3.2.2 National Results by Cause, Hospital Bed Days in Months Prior to Death, 1990-2006 p175

3.2.3 Area Deprivation Resilts, Hospital Bed Days in Months Prior to Death, 1993-2006 p179
3.2.4 Conclusions from Reference Point Modelling p183
3.3 Person-Month Logistic Regression p184
3.3.1 National Results, Association between Hospital use in a Previous Month and Death in the Current Month, 1990-1998 and 2004 p184
3.3.2 Conclusions from Person-Month Logistic Regression p193
3.4 Conclusion p194

Section 4 – Discussion **p197-210**
4.1 Introduction p197
4.2 Reference Point Modelling, Discussion of Results p197
4.3 Person-Month Logistic Regression, Discussion of Results p198
4.4 Institutionalised Healthcare in New Zealand, Implications for Public Hospital Utilisation Prior to Death p200
4.5 Changes in Place of Death over time in New Zealand, Implications for Public Hospital Utilisation Prior to Death p204
4.6 Relevance of Findings to the Research Questions p207
4.7 Limitations of this Research p208
4.8 Conclusion p209

CONCLUSION **p211-218**

- Reflection on the Data and Methods Used p212
- Key Finding 1 - Declining Hospital Bed Day and Mortality Rates over time, and Convergence by Deprivation p214
- Key Finding 2 - Evidence for Compression of Morbidity, Expecially at Older Ages p215
- Limitations and Further Research p216

REFERENCES **p219-228**

APPENDIXES **p229-301**

Appendix A – Data Preparation for Papers One and Two **p229-250**

- Introduction p229
- Step 1 - Export data from Access to Text, 1974-2000/01 p232
- Step 2 - Import data into SPSS from Text, 1974-2006/mid08 p234
- Step 3 - Add NZdep p235
- Step 4 - Convert ICD codes to numerical format and add ICD description p239
- Step 5 - Create grouped ICD and age variables p243
- Step 6 - Create filter variables for hospital data p246
- Conclusion p248

Appendix B – Data Linkage and Preparation for Paper Three **p251-294**

- Introduction p251
- Step 1 – Individual Record Linkage p251
- Step 2 – Probabilistic matching Process p255
- Step 3 – National and Area Deprivation Filters p262
- Step 4 – Duplicate and Overlapping Hospital Stays p264
- Step 5 – Length of Stay variables for Months Prior to Death p266
- Step 6 – Aggregate Length of Stay Variables and Reduce to One Record per Individual p268
- Step 7 – Incorporate Linked Dataset Back into Mortality Dataset (Reference Point Modelling) p270
- Step 8 – Baseline Hospital Use (Reference Point Modelling) p270
- Step 9 – Data Preparation for Person-Month Logistic Regression p275
- Step 10 – Production of Output for Reference Point Modelling p281
- Step 11 – Production of Output for Person-Month Logistic Regression p282
- Conclusion p292

Appendix Figures **p295-300**

Appendix Tables **p301-305**

LIST OF FIGURES

INTRODUCTION

Figure 0.1: Percent of New Zealand Population by Age Group pxxix

Figure 0.2: New Zealand Population Structure, Selected Deprivation Deciles, 2002 pxxx

PAPER ONE – Trends in Mortality and Hospital Utilisation

Figure 1.2.1: Model of the Social and Economic Determinants of Health p5

Figure 1.2.2: Percentage Maori/non-Maori in each Deprivation Decile and Total Population, New Zealand, 2001 p11

Figure 2.2.1: Combining the Datasets, Hospital Discharges p21

Figure 2.3.1: Number of Unfiltered Hospital Discharges, 1974-2007 p27

Figure 2.3.2: Number of Unfiltered Hospital Bed Days, 1974-2007 p27

Figure 2.4.1: Reasons for Excluding Cases from the Hospital Dataset p29

Figure 2.4.2 Number of Hospital Bed Days, 1974-2007, Various Filters p30

Figure 3.2.1: Age-Specific Mortality Rates, 1974-2006, Selected Ages p36

Figure 3.2.2: Age-Specific Mortality Rates, 1974 p37

Figure 3.2.3: Age-Specific Mortality Rates, 2006 p37

Figure 3.2.4: Standardised Mortality Rates, 1974-2006 p38

Figure 3.2.5: Selected Age-Specific Mortality Rates by Cause, 1974 and 2004, 95% Confidence Intervals p39

Figure 3.2.6: Standardised Mortality Rates by Cause, Selected Years, 95% Confidence Intervals p42

Figure 3.3.1: Age-Specific Hospital Bed Day Rates, Selected Ages, 1974-2006 p44

Figure 3.3.2: Age Specific Hospital Bed Day Rates, 1974 p44

Figure 3.3.3: Age Specific Hospital Bed Day Rates, 2006 p44

Figure 3.3.4: Standardised Hospital Bed Day Rates, 1974-2006 p46

Figure 3.3.5: Selected Age-Specific Hospital Bed Day Rates by Cause*, 1974 and 2004 p48

Figure 3.3.6: Standardised Hospital Bed Day Rates by Cause, Selected Years p49

Figure 3.4.1: Female Age-Specific Mortality Rates, Deprivation Deciles 1 and 10, Years 1991 and 2006 p51
Figure 3.4.2: Standardised Mortality Rates, Selected Deprivation Deciles, 1991-2006 p53
Figure 3.4.3: Age-Specific Mortality Rates by Cause, Deprivation Deciles 2 and 9, Selected Ages, 1991-1994, 95% Confidence Intervals p54
Figure 3.4.4: Age-Specific Mortality Rates by Cause, Deprivation Deciles 2 and 9, Selected Ages, 2003-2006, 95% Confidence Intervals p55
Figure 3.4.5: Standardised Mortality Rates by Cause and Selected Deprivation Deciles, 1991 and 2006, 95% Confidence Intervals p57
Figure 3.5.1: Female Age-Specific Hospital Bed Day Rates, Deprivation Deciles 1 and 10, Years 1991 and 2006 p60
Figure 3.5.2: Standardised Hospital Bed Day Rates, Deprivation Deciles 2 and 9, 1991-2006 p61
Figure 3.5.3: Age-Specific Hospital Bed Day Rates by Cause, Deprivation Deciles 2 and 9, Selected Ages, 1991-1994 p63
Figure 3.5.4: Age-Specific Hospital Bed Day Rates by Cause, Deprivation Deciles 2 and 9, Selected Ages, 2003-2006 p64
Figure 3.5.5: Standardised Hospital Bed Day Rates by Cause and Selected Deprivation Deciles, 1991 and 2006 p65
Figure 4.3.1: Regular Smoker Rates for Deprivation Deciles 1 and 10, by Gender and Age, 1996 p74
Figure 4.4.1: Average NZ*dep*2006 Quintiles by Meshblock in New Zealand p76
Figure 4.4.2: New Zealand Health Regions, as Defined by Pool *et al.* 2009 p77
Figure 4.4.3: Average NZ*dep*2006 Score by Health Region p78
Figure 4.4.4: Life Expectancy by Health Region in 2001 p80
Figure 4.4.5: Percent Maori by Health Region in 2001 p80
Figure 4.4.6: Age Standardised Mortality Rate per 1000 by Health Region in 2001 p80
Figure 4.4.7: Age Standardised Hospital Bed Day Rate by Health Region in 2001 p80

PAPER TWO - Examination of Compression of Morbidity using Life Table Techniques

Figure 3.2.1: Life Expectancy and HUE at birth for Males and Females between 1974 and 2006, Three-Year Moving Averages. p110

Figure 3.2.2: Conceptual Diagram showing Rectangularisation of the Survival Curve p112

Figure 3.2.3: Survival Curve, Selected Years, Males and Females p113

Figure 3.2.4: HUEs at Different Ages, Selected Years p115

Figure 3.2.5: Percentage of HUE_0 by Cause, Selected Years p116

Figure 3.3.1: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 1 between 1991 and 2006, Three-Year Moving Averages. p119

Figure 3.3.2: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 2 between 1991 and 2006, Three-Year Moving Averages p119

Figure 3.3.3: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 9 between 1991 and 2006, Three-Year Moving Averages p120

Figure 3.3.4: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 10 between 1991 and 2006, Three-Year Moving Averages p120

Figure 3.3.5: Survival Curve for Selected Years in Deciles 1 and 10, Males and Females p122

Figure 3.3.6: HUE_0 and Life Expectancy at Birth for Females, Selected Deprivation Deciles p123

Figure 3.3.7: HUE_{80} and Life Expectancy at age 80 for Females, Selected Deprivation Deciles p124

Figure 3.3.8: Female HUEs and Life Expectancy at Different Ages in 2004-06, by Deprivation Decile p124

Figure 3.3.9: Percentage of HUE_0 by Cause, Deprivation Deciles 1 and 10, Selected Years p128

Figure 4.4.1: Percent of Total New Zealand Health Expenditure by Source, 1970/80 to 1999/2000, Adjusted for Inflation p136

PAPER THREE - Examination of Compression of Morbidity using Longitudinal Techniques

Figure 1.4.1: Proportions of Ethics-Approved Research Projects using Name-Identified and Data-Linked Administrative Health Information in Western Australia 1990-2003 p151

Figure 2.2.1: Conceptual Diagram of the Datasets Used in the Analysis p158

Figure 2.3.1: Conceptual Diagram of the Compression/Expansion of Morbidity Debate p160

Figure 2.4.1: Interpretation of Compression of Morbidity for Person-Month Logistic Regression p163

Figure 2.4.2: Person-Month logistic Regression, Conceptual Diagram p164

Figure 2.4.3: Available Date Range for Person-Month Logistic Regression p166

Figure 3.2.1: Mean Number of Bed Days in Months Prior to Death, Actual and Baseline, 1990 and 2006, Selected Ages, by Sex p173

Figure 3.2.2: Mean Number of Bed Days in Months Prior to Death, Four Leading Causes of Death, 1990 and 2006, Females p177

Figure 3.2.3: Percentage of Deaths by Cause, All Ages, Both Sexes, 1990 and 2006 p178

Figure 3.2.4: Mean Number of Bed Days in Months Prior to Death, Actual and Baseline, Deprivation Decile 1, 1993 and 2006, Selected Ages, by Sex p180

Figure 3.2.5: Mean Number of Bed Days in Months Prior to Death, Actual and Baseline, Deprivation Decile 10, 1993 and 2006, Selected Ages, by Sex p181

Figure 3.3.1: Odds Ratio and 95% C.I. of Death in the Current Month due to Excess Hospital Bed Days in a Previous Month for Females aged 75-77, Selected Years p186

Figure 3.3.2: Odds Ratio and 95% C.I. of Death in the Current Month due to Excess Hospital Bed Days in a Previous Month for Females aged 80-82, Selected Years p187

Figure 3.3.3: Odds Ratio and 95% C.I. of Death in the Current Month due to Excess Hospital Bed Days in a Previous Month for individuals aged 85-87, Selected Years p188

Figure 3.3.4: The Changing Distribution of the Number of 1, 1 Observations in each of the 24 Previous Months. Females aged 75-77, Frequency Diagrams, p191

Selected Years
Figure 3.3.5: Number of 1, 1 Observations in the 24 Previous Months. Females aged 75-77, Bar Charts, Selected Years p192
Figure 3.4.1: Compression of Morbidity Pathways p194
Figure 4.4.1: Percent of Usually Resident Population in Residential Healthcare Institutions by Age and Sex, 1996 and 2006 p201
Figure 4.4.2: Percent of Healthcare Institution Residents within each Age Group by Institution Type, both Sexes, 1996, 2001 and 2006 p203
Figure 4.4.1: Proportion of Deaths occurring in Public Hospitals, All Ages 1989-2006 p205
Figure 4.4.2: Proportion of Deaths occurring in Public Hospitals, Selected Ages, 1989-2006 p206

LIST OF TABLES

PAPER ONE – Trends in Mortality and Hospital Utilisation

Table 1.2.1: Variables used in NZ*dep*2001 p7

Table 2.2.1: Description of Datasets p20

Table 4.2.1: New Zealand Mortality Rates for ages 35-69, 1975 and 1995, and the Percentage Attributable to Smoking p71

PAPER TWO - Examination of Compression of Morbidity using Life Table Techniques

Table 3.2.1: Ratio of 2004-06 to 1974-75 for Life Expectancy and HUE, Selected Ages p114

Table 3.3.1: Ratio of 2004-06 to 1991-93 Life Expectancy and HUE by Deprivation Decile, Selected Ages p126

Table 4.4.1: Percent of New Zealanders with Private Health Insurance Policies, 1994/95 to 2002/03 p136

PAPER THREE - Examination of Compression of Morbidity using Longitudinal Techniques

Table 3.2.1: Percentage change in Mean Hospital Bed Days in the Months Prior to Death between 1990 and 2006, for Deaths at Ages 70 Years and Above, by Sex p175

Table 3.3.1: Categories of Observations from the Person-Month Logistic Regression p190

INTRODUCTION

The population of New Zealand is ageing, and older ages experience the highest rates of morbidity and use of health services. This research examines how hospital utilisation at older ages is changing over time, and how these trends vary by socioeconomic area deprivation. This research therefore has a dual focus: health in later life, and inequalities in health. The following three papers explore aspects of health in New Zealand using administrative data on mortality and public hospitals from 1974 to 2006. Mortality and public hospital utilisation patterns are examined with an emphasis on older ages. A range of theories has been developed to describe morbidity experience at older ages; these include the compression and expansion of morbidity hypotheses. The second and third papers in this series use the New Zealand hospital data as a proxy for morbidity, and examine if morbidity at older ages and in the last few months of life is increasing or decreasing over time, and if it varies by area deprivation.

This work builds on a recently published monograph detailing regional trends in mortality and public hospital data in New Zealand from 1981 to 2006 (Pool, Baxendine *et al.* 2009)[1]. The monograph employed a new life table based measure of health over the lifecourse, Hospital Utilisation Expectancies (HUEs). The primary purpose of the monograph was to guide health professionals in New Zealand, and illuminate the efficiencies and failings of the health system in different regions of New Zealand. As such, the findings of the monograph are of limited application in settings outside New Zealand. This paper extends the research presented in the monograph by exploring mortality and hospital use patterns at the socioeconomic level: area deprivation rather than geographic region is the unit of analysis in these papers. Using area deprivation allows the findings to be more transferable to other countries, and other settings. Health patterns by level of area deprivation are also of importance to policy makers, due to the fact that subsidies for health services in New Zealand (and many other countries) are allocated based on measures of area deprivation. While this research is conducted using New Zealand data, many aspects are highly transferable to other countries and the findings are of interest to the international community. New Zealand is not an anomaly: as a developed country with a (largely) free

[1] This monograph is available online at:
http://www.waikato.ac.nz/nidea/research/hospital-care-monograph (accessed 13th December 2010).

healthcare system and persistent socioeconomic health inequalities it draws parallels with other countries, including the UK.

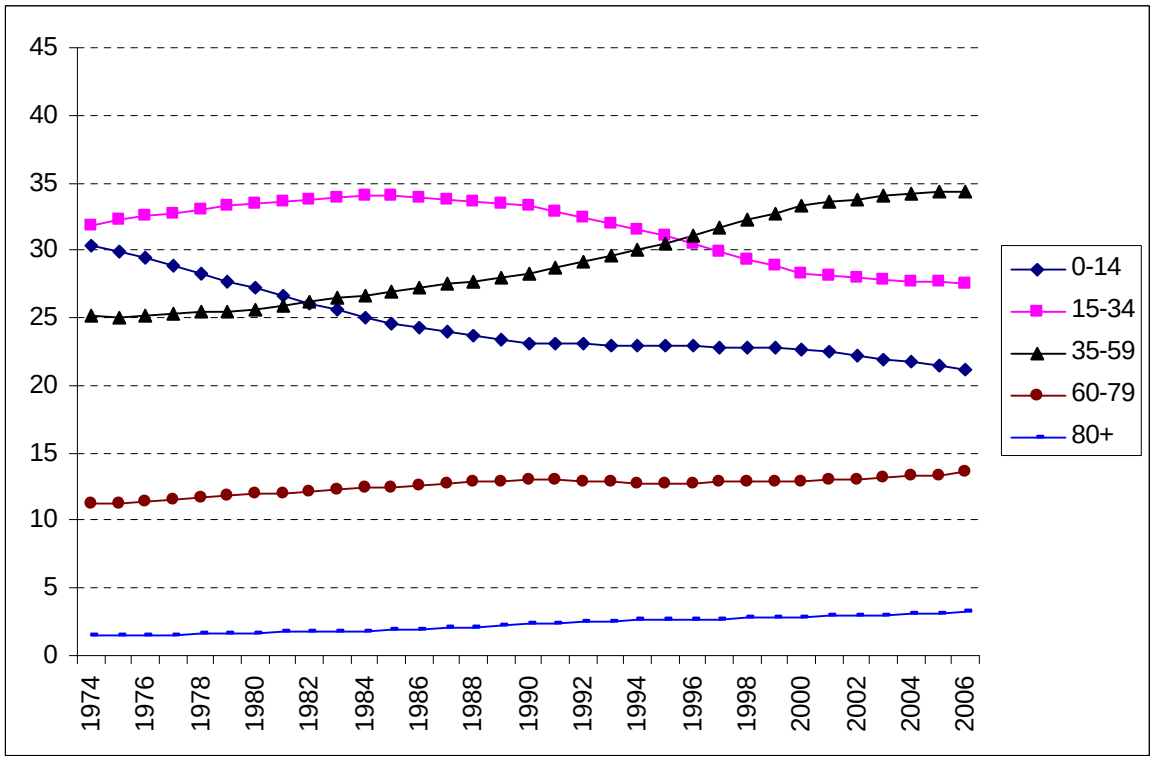
This introduction discusses the changing composition of the New Zealand population and the New Zealand health system. Health in later life is discussed in more detail, followed by the key objectives of these papers. Lastly the structure of the three papers, with reference to the research questions is described.

The New Zealand Population Structure and Composition

The population of New Zealand grew from three million to just over four million between 1974 and 2006 (United Nations Population Division Website 2009). In the space of just three decades between 1975-77 and 2005-07, life expectancy in New Zealand rose by 9.0 years for males and 6.7 years for females (Statistics New Zealand 1979; Statistics New Zealand 2008). The New Zealand population structure has changed markedly over the time period studied (see Figure 0.1). Since 1974, there have been increases in the percent of the population in middle and older age groups (35 years and over), and declines in the percent of the population at younger ages (below 35 years). The bulk of this population ageing is due to structural ageing, resulting from reduced fertility. However, in recent years increased survival at older ages has also begun to have an effect (Pool 1994). Given that morbidity and hospital use increases with age, changes in age structure may result in higher demand for health services.

There are also cohort effects in the morbidity experience of older age groups. For instance, will the post World War Two (WW2) cohort (currently in their 60s) experience similar morbidity to their parents (currently in their 80s and 90s) as they grow older? Lifestyles have changed out of recognition since WW2: labour saving devices, widespread car ownership, increasingly sedentary lifestyles, the peaking and decline of smoking rates, high consumption of processed food and increased female employment, are but a few examples of changes in lifestyle and diet since WW2. For the post war baby boomers, experiencing a lifetime of risk factors different to their parents must surely result in different morbidity patterns in old age.

Figure 0.1: Percent of New Zealand Population by Age Group, 1974-2006



Source of Data: (United Nations Population Division Website 2009)

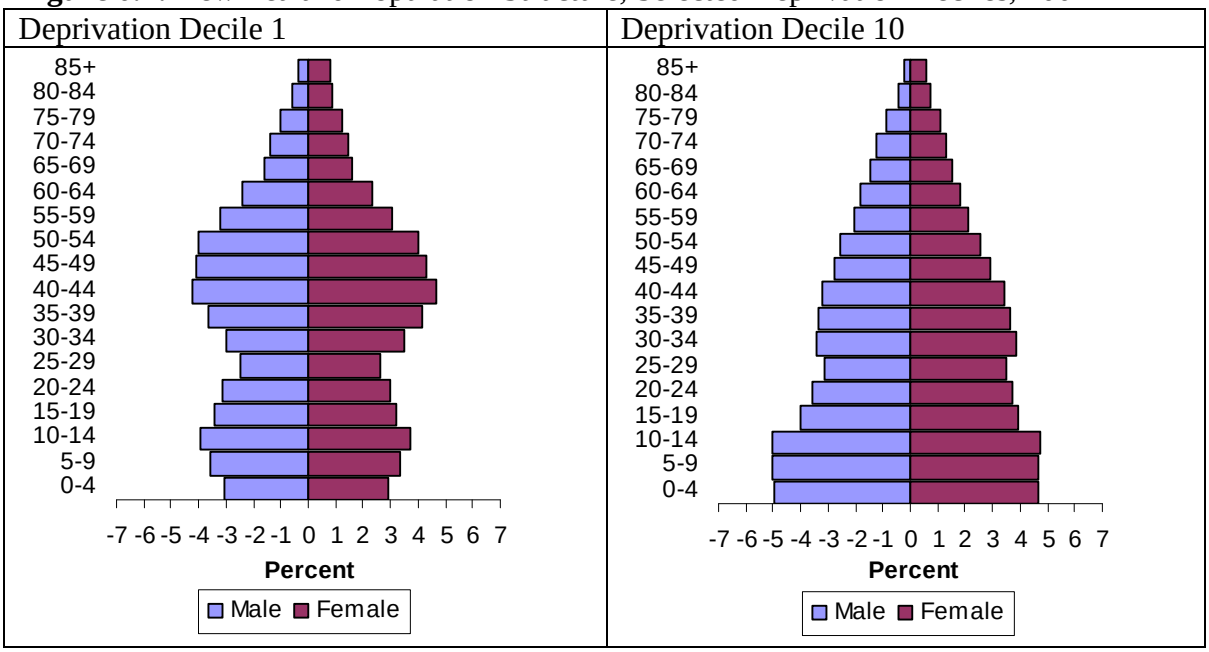
In addition to a changing age structure, the ethnic composition of the New Zealand population has also changed considerably since the 1970s. There have been several waves of immigration from Pacific Island countries since WW2, which peaked in the 1980s (Finau and Tukitonga 1999). Approximately seven percent of the New Zealand population is Pacific, and due to the recent immigration history and high fertility Pacific peoples have a much younger age structure than the general population. However Pacific peoples in New Zealand experience worse health than the general population. As the Pacific population ages it is unlikely to experience similar morbidity patterns as the general population. This may compound the cohort effect of changes in morbidity at older ages. However, due to inconsistent coding of ethnicity over the time period in both the mortality and the public hospital datasets, ethnicity will not be included in these analyses. Ethnicity is also a sensitive topic in New Zealand. A very detailed examination of mortality and hospital use by ethnicity is provided in the aforementioned monograph (Pool, Baxendine *et al.* 2009).

The population structure of New Zealand also varies substantially by deprivation. Figure 0.2 presents the population age and sex structure for the least deprived tenth (deprivation decile 1) and the most deprived tenth (deprivation decile 10) of areas in 2002. The population pyramid for decile 1 is a classic western shape with a narrow base and

large proportion of the population in the middle ages, whereas the pyramid for decile 10 is a classic developing shape, with a wide base. However, unlike developing countries a sizeable proportion of the population is present at the oldest ages. The shape of these population pyramids are driven primarily by fertility. More deprived areas have high fertility, and rapidly growing populations and less deprived areas have lower fertility. A narrowing of the base of a population pyramid would usually denote fertility below replacement level (2.1 births per woman). However in this case the narrowing of the pyramid base at ages 0-9 may be due to a small previous cohort, aged 25-29 in 2002 who would be expected in turn to have a small number of children, regardless of fertility rates.

More deprived areas also have higher mortality rates than less deprived areas, and higher proportions of Maori and Pacific peoples.

Figure 0.2: New Zealand Population Structure, Selected Deprivation Deciles, 2002



Source of Data: Statistics New Zealand

Structure of the New Zealand Health System, and Relevance to Morbidity Research

New Zealand was the first country in the world to adopt a universal welfare system (Pearce and Dorling 2006). Like the UK health system, the New Zealand health system aims to provide healthcare that is available to all, regardless of the ability to pay. This makes New Zealand a perfect example for studying hospital use as a proxy for morbidity. In countries

with no universal health insurance, uptake of health services may be artificially depressed among disadvantaged groups despite high levels of morbidity, due to an inability to afford healthcare. If hospital utilisation is used as a proxy for morbidity in these settings, it would give the misleading impression that disadvantaged groups have low morbidity levels. In New Zealand, examining uptake of healthcare services should provide a relatively balanced picture of the *demand* for healthcare services in different population groups, which is ultimately driven by morbidity levels.

The New Zealand health system is not universally free at point of contact, but there are subsidies in place to ensure universal access. At some healthcare levels (especially at the primary level) services are subsidised for certain population groups. In 2002 the New Zealand health system shifted from individual-based subsidies to area-based subsidies, based on measures of area deprivation (Hefford, Crampton *et al.* 2005). This is one of a plethora of health system reforms in New Zealand over the last three decades, and these reforms have had a negative impact on the consistency and quality of public hospital data over time (Katzenellenbogen, Baxendine *et al.* 2001; Pool, Baxendine *et al.* 2009). However, by excluding certain types of cases and by using the number of days spent in hospital rather than the number of discharges, New Zealand public hospital data can be made suitable for use as a measure of morbidity over the time period 1974 to 2006. The exclusion of particular cases is referred to in this work as filtering, and described in detail in Section 2 of the first paper and in Appendix A.

There is a relatively large body of literature documenting socioeconomic differences in health service use (Billings, Zeitel *et al.* 1993; Hofer, Wolfe *et al.* 1998; Jackson, Kelsall *et al.* 1998; Salmond and Crampton 1999). This study differs from previous research in three ways. Firstly, most of these studies use hospital discharge rates, which are vulnerable to changes in health system policy. The number of discharges from public hospitals was found to increase over the time period, but the number of hospital bed days decreased. Secondly, most of these studies did not exclude any types of hospital stay, except childbirth. Even *if* the hospital data are consistent over time with few changes in coding and policy, there is still a wide range of diagnoses that does not represent ill health. The third difference is that many of the studies on health service use by socioeconomic deprivation aim to assess the quality, efficacy and efficiency of the health service. This study does not use hospital data to assess the quality of the health system; instead it uses hospital data to directly infer morbidity for different population groups.

The Implications of Health Transition Theories on Health in Later Life

The population of New Zealand is ageing, both structurally, and through increased life expectancy and reduced mortality at older ages. If life expectancy is increasing and mortality at older ages is decreasing, it is logical to assume that morbidity is also decreasing. However, when it is considered that many causes of morbidity are not life threatening, that the highest rates of morbidity occur at older ages, and that more people are surviving to older ages, this assumption ceases to be quite so logical.

Quantifying the onset and level of morbidity in later life is of great importance, both at a human level (can we expect to live longer, healthier lives?) and at a policy level (should the health system brace itself for increased demand?). The compression of morbidity theory argues that medical advances will raise the age at onset of morbidity at a faster pace than raises in life expectancy, resulting in a reduction of morbidity at a population level (Fries 1980; Fries 1983). A fixed limit to life expectancy resulting in rectangularisation of the survival curve is seen by some as a necessary prerequisite for compression of morbidity. It is argued that increases in healthy life expectancy cannot outstrip increases in total life expectancy if total life expectancy continues to increase.The opposing argument is the expansion of morbidity theory, which sees the age at onset of morbidity remaining relatively stationary, and increases in life expectancy resulting in an increase of morbidity at a population level (Gruenberg 1977). A third, middle of the road theory is the dynamic equilibrium theory, age at onset of morbidity and life expectancy will increase in tandem, resulting in no change in the level of morbidity at a population level (Manton, Stallard *et al.* 1991).

These papers use the New Zealand mortality and hospital datasets to quantify the onset and severity of morbidity, and identify which of these health transitions (compression, expansion or equilibrium) were occurring in New Zealand between 1974 and 2006, and to examine variation by area deprivation between 1991 and 2006.

Main Objectives and Research Questions

There are two main objectives to this study. Firstly, to identify how recent mortality and hospital bed day rates in New Zealand differ by area deprivation. Secondly, to identify the health transition (compression, expansion or equilibrium) occurring in New Zealand, and examine if this is consistent between areas with different socioeconomic characteristics. These objectives translate into four research questions:

Two research questions were identified to operationalise the first objective:

1. How do the **patterns of mortality** in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of mortality differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

2. How do the **patterns of hospital use (**bed day rates) in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of hospital use differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

These two research questions are addressed in the first paper. Age-specific and age-standardised rates are used, and mortality and hospital bed days are examined separately. Inequalities by area deprivation are examined. For the second objective regarding the health transition, two further research questions were identified:

3. Taking hospital utilisation as a proxy for severe morbidity, does the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through the use of Hospital Utilisation Expectancies (HUEs)?
    a. Does the lifetime expectation of hospital use at birth and at different ages differ by deprivation decile?
    b. How have these trends changed over time?

4. Taking hospital utilisation as a proxy for severe morbidity, does the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through analysing time spent in hospital in the last months of life?
    a. Does hospital use in the last months of life differ by age at death?
    b. Does hospital use in the last months of life differ by deprivation decile?
    c. How have these trends changed over time?

The third and fourth research questions are addressed in papers two and three respectively. They examine evidence for compression of morbidity, but use different types

of methods. For the third research question, Hospital Utilisation Expectancies (HUEs) at different ages are examined. HUEs draw on life table methodology, and combine both the hospital and mortality datasets. This analysis is presented in the second paper.

The fourth research question examines hospital use in the last few months of life. This involves the combination of hospital and mortality data, but unlike for HUEs, requires record linkage between the hospital and the mortality datasets. Record linkage allows the identification of hospital utilisation in the months prior to death, for each individual who died over a certain time period. The third paper in the series presents the aggregated results from this analysis.

Structure of the Papers

For ease of navigation each of the three papers is set out in the same way and consists of four sections. The first section in each paper reviews the relevant literature. For the first paper this involves literature concerning deprivation and health, and explores themes such as the relevance of hospital utilisation as a proxy for morbidity and the potential ecological fallacy of using area deprivation measures to represent individual socioeconomic status. The literature review for the second paper addresses the epidemiological transition and the compression, expansion of morbidity and dynamic equilibrium health transition theories. For the third paper the literature review focuses on the ethics of using medical records in public health research, the ethics of individual record linkage in public health research, and examines the literature on the concentration of morbidity towards the end of life.

The second section in each paper describes the data and methods used in each paper. While the same datasets are used in all three papers, different methods are used, requiring preparation of the datasets in different formats. Technical descriptions of the preparation of the datasets for analysis are found in the appendixes.

The third sections present the results from the analyses, sticking to primarily descriptive commentary. Discussion of the results takes place in the fourth section of each paper, examining the findings and how they fit within the existing literature, examining potential confounders, suggesting possible reasons for the findings, and listing any caveats and limitations of the analyses.

Following the third paper, a conclusion discusses the key findings and implications of all three papers.

PAPER ONE

Trends in Mortality and Hospital Utilisation

RESEARCH QUESTIONS 1&2

1. How do the **patterns of mortality** in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of mortality differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

2. How do the **patterns of hospital use (**bed day rates) in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of hospital use differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

ABSTRACT

Both mortality and hospital bed day rates declined over the time period, and the gap between male and female rates converged slightly for mortality and considerably for hospital bed days. More deprived areas had higher mortality and hospital bed day rates than less deprived areas over the time period. There is little difference by deprivation in mortality rates at the oldest ages (85+): the overall difference is driven by excess mortality in more deprived areas at middle ages (50-84). In contrast, the difference by deprivation in hospital bed day rates did *not* narrow at the oldest ages (85+).

Much of the decline in mortality (for all deciles and both sexes) was driven by circulatory conditions. Decreases in male smoking related mortality and increases in female smoking related mortality may completely explain the convergence between male and female mortality seen over the period, but cigarette smoking does *not* explain the difference in mortality rates by deprivation. Hospital bed day rates followed similar trends to mortality rates, supporting the use of hospital bed days as a proxy for morbidity. There was however an interaction between deprivation and gender: age-standardised mortality and hospital bed day rates were higher for females than males in the least deprived areas, yet higher for males than females in the most deprived areas and at a national level.

SECTION 1 - LITERATURE REVIEW: Socioeconomic Health Inequalities and Hospital Utilisation

1.1 Introduction

We live in increasingly unequal societies. In the last few decades income inequality in many developed countries has increased dramatically. A recent UK report found income inequality to be as high as it was in the 1970s, despite a string of government policies to reduce inequality (Hills, Brewer *et al.* 2010). Where you are born is strongly associated with your future education, income and health prospects. In the midst of this inequality, national health systems are becoming increasingly target driven and are under pressure to be more efficient and accountable. Health and education subsidies are targeted towards disadvantaged areas, based on census-derived deprivation indexes. This work builds on a recent monograph (Pool, Baxendine *et al.* 2009) that detailed regional mortality and hospital trends in New Zealand over a 20 year period from 1981 to 2001, and introduces a new aspect: area deprivation. This paper presents an exploratory analysis of national mortality and hospital data in New Zealand over a 32 year time period, from 1974 to 2006. Sub-national trends by area deprivation are also examined from 1991 to 2006, providing a background for further papers that address themes such as expected hospital utilisation over the life course and health in later life.

This section introduces the key concepts of deprivation and health, with area deprivation and measures of morbidity discussed in the context of New Zealand. Emphasis is placed on NZ*dep*, which is a New Zealand measure of area deprivation. The relevance of using area deprivation measures instead of individual deprivation measures in health research is explored. Consideration is also given to ethnicity, which confounds the relationship between area deprivation and health in the New Zealand context. Lastly the validity of hospital utilisation as a proxy for morbidity is examined.

1.2 Area Deprivation

The relationship between socioeconomic disadvantage and health has been well trodden in research, which is not surprising given that individual and area measures of deprivation are

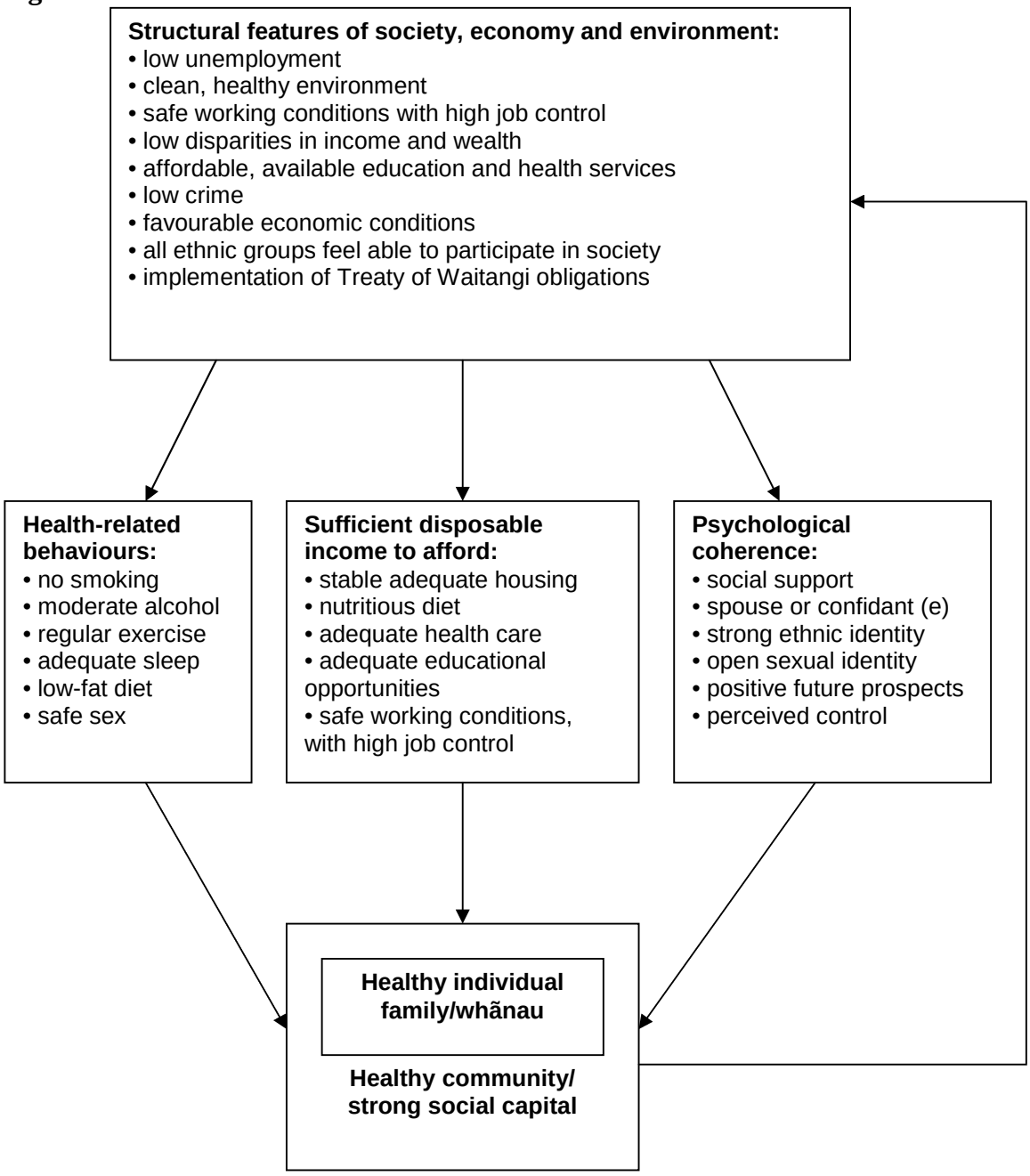
used to direct resource allocation in many countries. The rationale for studying health by deprivation includes looking for clues that point to causes of disease, and identification of groups in the community who have excess mortality that could potentially be prevented (Paul 2002).

Previous research of health gradients by both individual and area deprivation measures are extensive. Banks *et al.* used a range of large health surveys in the US and UK to compare prevalence of self-reported disease with a range of individual socioeconomic and behavioural indicators included in the survey (smoking, overweight, obesity, alcohol drinking) and found a clear relationship between increasing socioeconomic status and decreasing chronic disease (Banks, Marmot *et al.* 2006).

Blakely *et al.* compared all-cause and cause-specific mortality in New Zealand with a variety of individual socioeconomic indicators, including income, education, occupational class, labour force status and small-area deprivation, and found a clear gradient of increasing mortality with increasing socioeconomic disadvantage. The relative index of inequality (RII) between the 10^{th} and the 90^{th} percentile found mortality to be 1.94 times greater for high compared to low area deprivation (Blakely, Woodward *et al.* 2002). Tobias and Cheung looked at life expectancy by deprivation decile in New Zealand, finding a consistent gradient between the most and least deprived 10 percent of 9 years for males and 7 years for females in the mid to late 1990s (Tobias and Cheung 2003). However not all research has found a gradient in health by area deprivation. Eachus *et al.* compared individual- and area-based measures of deprivation in their study of self-reported and medically assessed severity of hip disease (Eachus, Chan *et al.* 1999). They found that difference in disease severity by deprivation, clearly observable using individual-level deprivation measures, was not significant when using area-level measures. On balance however, the evidence is overwhelmingly in favour of socioeconomic gradients in health and mortality for area-based deprivation measures.

Living in a deprived area may influence health in many ways. Barnett *et al.* proposed three causal pathways for how social inequality may affect health. Comparing your own situation with that of others who are better off may lead to psychosocial factors that influence health. Social capital and cohesion may be better in less deprived areas, offering the support needed both psychologically and materially, and more deprived areas may suffer from a lack of investment leading to fewer opportunities and poor amenities (Barnett, Moon *et al.* 2004).

Figure 1.2.1: Model of the Social and Economic Determinants of Health



Note: Arrows indicate probable causality
Source: (Ministry of Health 2000, p9)

Figure 1.2.1 presents a diagram from a Ministry of Health report showing the web of factors through which area deprivation can adversely affect health in a New Zealand setting. Many of these factors have measurement issues. How can you quantitatively assess the 'healthiness' of an environment, the level of social support, or measure how able to participate in society people of different ethnic groups feel? Area deprivation measures capture what *can* be measured (smoking rates, average income, housing quality) given

available census data, with the aim of inferring what cannot be measured, or easily quantified.

This section looks at different measures of area deprivation, focussing on NZ*dep* in particular. Confounders such as ethnicity are discussed, followed by a critique of using area deprivation as a proxy for individual deprivation and ultimately, the underlying 'deprivation' variable itself.

1.2.1 Types of Area Deprivation Measures

Area deprivation measures typically calculate a score from a range of census variables. Variables that indicate income, material situation and education are used, such as ownership of white goods, housing tenure, car ownership, number of residents per room, proportion on means tested benefits or unemployed and (where available) educational attainment and income. An average score is derived from these variables, which is then split into deciles or quintiles to denote the relative deprivation of each area on the national scale. For instance, the bottom quintile would represent the most deprived 20 percent of areas in a country, containing (approximately) 20 percent of the population.

Area deprivation scores may be commissioned by the government, or calculated by individual researchers. The former is more common, given data access, the clumsiness of aggregated summaries and the sensitive nature of individual-level census data. Area deprivation measures are tailored to suit the population, the census questions available, and the research it is to be used for. In the UK Carstairs and Townsend are common area deprivation scores. In New Zealand a Ministry of Health designed area deprivation score, NZ*dep* is commonly used.

1.2.2 NZ*dep*

NZ*dep* is an area-based measure of deprivation that has been calculated every census year since 1991. For this study NZ*dep*91, NZ*dep*96 and NZ*dep*2001 are used. Many countries, such as the UK, have censuses every 10 years. The characteristics of a small area can change considerably over a decade. New Zealand has the advantage of censuses every 5 years, allowing deprivation indexes to be calculated more regularly. The answers to nine census variables are combined and weighted to give a number for each small area. Areas of either about 100 (meshblocks) or 2,000 people (Census Area Units) are used. The variables (in order of weighting) for NZ*dep*2001 are shown in Table 1.2.1.

Table 1.2.1: Variables used in NZ*dep*2001

		Variable (proportion in small areas) in order of decreasing weight in the index
1	Income	People aged 18-59 receiving a means tested benefit
2	Employment	People aged 18-59 unemployed
3	Income	People living in equivalised* households with income below an income threshold
4	Communication	People with no access to a telephone
5	Transport	People with no access to a car
6	Support	People aged <60 living in a single parent family
7	Qualifications	People aged 18-59 without any qualifications
8	Living Space	People living in equivalised* households below a bedroom occupancy threshold
9	Owned home	People not living in own home

*Controlling for household structure
Source: (Salmond and Crampton 2002, p6)

NZ*dep* can either be reported as a score, which is a first principal component score that is scaled to have a mean of 1000 index points and a standard deviation of 100 points, or as deciles which range from 1 to 10, where 10 represents the areas with the highest 10 percent of deprivation scores, and 1 represents the areas with the lowest 10 percent of deprivation scores (Salmond and Crampton 2002). This research uses the decile version of NZ*dep*. When the term deprivation 'score' is used, it refers to deprivation deciles.

From 1996 on there were several small changes in how NZ*dep* is calculated, including shifts in the household equivalised income threshold, changes in the range of means tested benefits included and the age threshold for these, and a small change in the living space variable (Tobias, Bhattacharya *et al.* 2008). These changes were necessary for the continued relevance and comparability of the NZ*dep* measure between censuses, and when the weights applied to each of the variables is considered it is apparent that there is very little change in the composition of the NZ*dep* measure from 1996 onwards (Tobias, Bhattacharya *et al.* 2008). A slightly larger change occurred between 1991 and 1996, two variables were dropped and a new variable (telephone access) was included. While this raises concerns about continuity over time, to exclude NZ*dep*91 from the analysis would be to exclude a third of the available data. For these reasons NZ*dep*91 will *not* be excluded, but caution will be used in interpretation of trends between 1991 and 1996.

There are two further concerns with using NZ*dep* to analyse population health in New Zealand. Firstly, the validity of area deprivation measures as a proxy for individual deprivation in general, and secondly, the ethnic composition of the New Zealand population.

1.2.3 Is Area Deprivation a Relevant Indicator of Individual Deprivation?

Individual deprivation is regarded as the ideal measurement for investigating the relationship between deprivation and health. However, deprivation scores comprise of many variables which are irrelevant for people outside of working age, such as unemployment, income, and access to a car. Salmond and Crampton tested whether area deprivation can legitimately be used as a proxy for individual deprivation (Salmond and Crampton 2001), but they were only able to calculate individual deprivation for individuals aged 18-59. Individual deprivation can often only be calculated for men. Many older studies focus exclusively on male socioeconomic health gradients due to the irrelevance of income and unemployment questions for women because many women of working age would stay at home to look after children. It could be argued that an area measure of deprivation that includes people of *all* ages and both genders, albeit based largely on the characteristics of males aged 18 to 59, allows a measure of deprivation for those people for whom *individual deprivation cannot be calculated* (children, people over retirement age and women who stay home to look after children). This interpretation necessitates the assumption that similar people cluster together in the same geographical area. While people outside of working age may not be adequately captured by deprivation measures, it can be inferred that they have similar characteristics to people of working age in the same area. This assumption is put under pressure when using deprivation to examine health and mortality trends. People over retirement age experience higher rates of mortality and hospital use, yet these are the people for whom area deprivation is the least relevant due to many questions being applicable only for people of working age. When considering mortality rates by area deprivation this may result in the numerator not matching the denominator. Put simply, the people measured in area deprivation scores are not the people who are dying, and this causes problems when researching mortality differences by deprivation. It is therefore crucial that area deprivation is a valid proxy for individual deprivation at older ages.

Deprivation is not simply comprised of compositional characteristics of individuals, but also contextual area factors such as quality of amenities. There is debate over the causality of compositional and contextual factors (Salmond and Crampton 2001). Does high income in an area bring in more amenities, or does a lack of amenities drive high income individuals out of an area? Both undoubtedly play a part in creating a climate for health. Recent research has attempted to pin down and measure contextual effects using Geographic Information System (GIS) technology. This is used to measure the distance

from an individual's home to local amenities (good or bad) such a park (good) or a chip shop (bad) (Rodgers, Lyons *et al.* 2009). However, this technology is very much in its infancy, and cannot hope to capture the overall *character* of an area. There may be a park nearby, but do residents feel safe enough to use it?

Deprivation variables, be they individual or area based, are often meaningless when interpreted on their own merits. It is impossible to conceive how having no access to a telephone, for example, may be causally related to smoking habits, or how not owning your home may be related to use of hospital services. What deprivation measures aim to do is capture latent, contextual effects which cannot be measured. The question should therefore be not whether area deprivation measures are a good proxy for individual deprivation, but whether area deprivation or individual deprivation measures are *better at capturing contextual effects*, which are ultimately related to health outcomes.

Various studies have attempted to evaluate the relative importance of individual- and area-level deprivation measures. Salmond and Crampton compared NZ*dep*96 deciles with separately calculated individual indexes of deprivation, but they were limited to people aged 18-59 (Salmond and Crampton 2001). They found that the vast majority of people had no deprivation indicators, i.e. income below a threshold, unemployed, no access to car, and that only very few people had several deprivation indicators. When contrasting this index with deprivation deciles it was inevitably found that the most deprived 10 percent of areas had a large proportion of people with no or few individual deprivation indicators. However this study, while interesting, is methodologically flawed. If, for example, 70 percent of a population have none or just one indicator of deprivation, they cannot *all* live in the least deprived decile. By definition only a tenth of the population can reside in each decile.

Eachus *et al.* compared individual- and area-level measures of deprivation in their study of self-reported and medically assessed severity of hip disease in South West England (Eachus, Chan *et al.* 1999). They found that difference in disease severity by deprivation was clearly observable using individual-level deprivation measures, but was not significant when using area-level measures. Area-based deprivation is routinely used to target health services and subsidies in various countries. This has serious implications if area-based measures of deprivation are *not* sufficiently capturing the deprivation effects that are behind the severity of disease. However, in a study of all-cause and cause-specific mortality for a range of individual- and area-based socioeconomic measures, Blakely *et al.* found strong socioeconomic gradients for all-cause mortality and most cause-specific

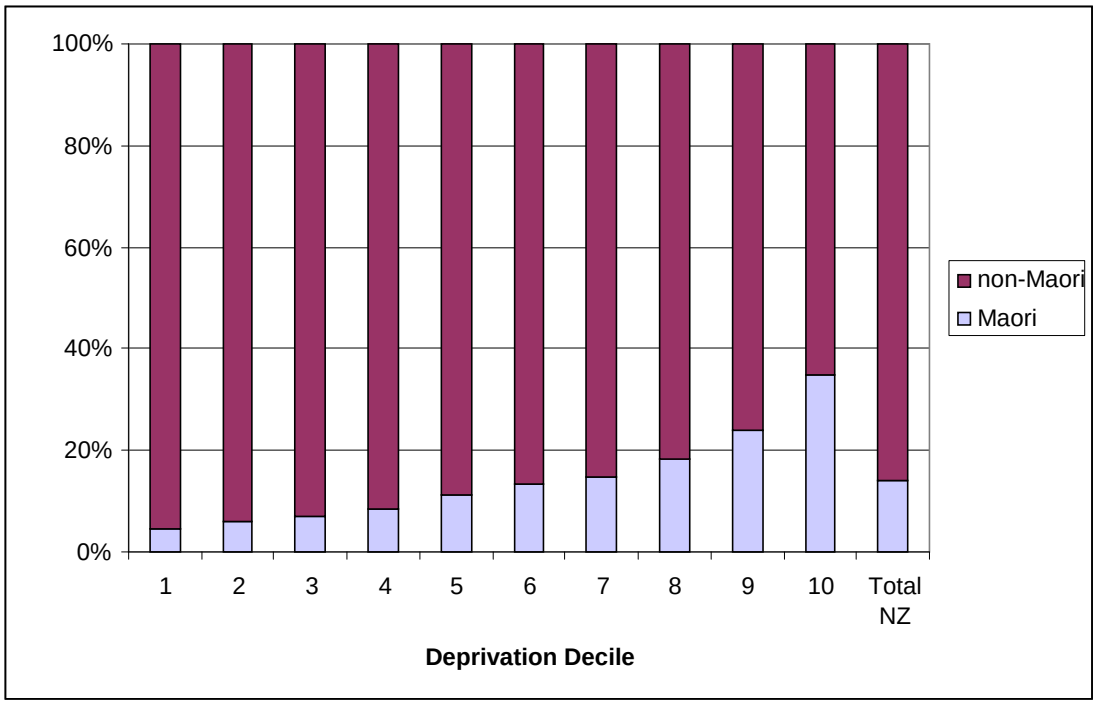
mortality regardless of which socioeconomic indicator (individual- or area-based) was used. In fact, area deprivation had the strongest association with mortality after income (Blakely, Woodward *et al.* 2002; Paul 2002). It can be concluded that while area deprivation can never be a perfect match for individual deprivation, its use in the literature is established and it is a reasonable indicator of the underlying effect of deprivation in a small area. NZ*dep* is particularly strong compared to other area deprivation indexes due to the income component, which Blakely *et al.* found to have the strongest association with mortality (when used at the individual level) (Blakely, Woodward *et al.* 2002). Income is included in the New Zealand census but not in the census for many other countries (including the UK).

A last point of interest is the issue of gender in deprivation analyses; individual measures of deprivation can be difficult to calculate for women. In their analysis of mortality trends by individual-level socioeconomic status, Wamala *et al.* raise an interesting point (Wamala, Blakely *et al.* 2006). Is it relevant to use a gender-neutral measure of socioeconomic status such as household income, or in this case area deprivation, for comparing sex-specific mortality? However, other research has found that while men and women have different social characteristics and prevalence of risk factors, such as income, marriage, physical activity, smoking and propensity to attend religious services, risk factors have a similar impact on the mortality of men and women (Rogers, Everett *et al.* 2008). A New Zealand study of mortality and socioeconomic disadvantage using a range of indicators found a strong socioeconomic mortality gradient for both men and women (Blakely, Woodward *et al.* 2002; Paul 2002).

1.2.4 Ethnicity as a Confounder of Area Deprivation

By far the single largest confounder of area deprivation in New Zealand is ethnicity. Ethnicity varies considerably by area deprivation. 14.1 percent of the total population in 2001 were Maori, yet in the most deprived decile 35.0 percent were Maori, and in the least deprived decile only 4.7 percent were Maori (see Figure 1.2.2), compared to the expected 10 percent in each decile. By definition, approximately 10 percent of the New Zealand population lives in each deprivation decile. However, only 7.0 percent of the non-Maori population lived in the most deprived decile in 2001, compared to 23.0 percent of the Maori population. Almost a quarter of Maori in New Zealand live in the most deprived 10 percent of areas.

Figure 1.2.2: Percentage Maori/non-Maori in each Deprivation Decile and Total Population, New Zealand, 2001



Source of data: (Maori Health Website 2010) data originally from Statistics New Zealand

The trend for the Pacific ethnic group was even more pronounced than for Maori. In 2006 over a third (34.8 percent) of Pacific peoples in New Zealand resided in decile 10 compared to 24.2 percent of Maori (Tobias, Bhattacharya *et al.* 2008).

Deprivation measures may not work equivalently for all ethnic groups (Tobias, Bhattacharya *et al.* 2008). Due to differences in household and family structure, overcrowding measures may not be equally applicable for Pacific and European households. As Tobias *et al.* put it, overcrowding may not have the same 'meaning' for Pacific households compared to other ethnicities and therefore the underlying latent deprivation variable may differ by ethnicity (Tobias, Bhattacharya *et al.* 2008).

Household income is the third most important variable in calculating NZ*dep*, however there are considerable differences in non-response of the income question in the census by ethnicity. In the 1996 census 88 percent of Europeans answered the income question, compared to 75 percent of Maori and 62 percent of Pacific people (Judge and Paterson 2001). This results in under-representation of Maori and Pacific income in NZ*dep*.

In summary, ethnic groups are not equally distributed across deprivation deciles in New Zealand, deprivation measures may not be equally appropriate for different ethnic

groups, and certain ethnic groups may be under-represented in area deprivation measures. But why does this matter? These differences are of concern when considering health by small area deprivation in New Zealand because ethnic groups have different health seeking behaviours and experience different rates of disease and mortality. Area measures of deprivation in New Zealand are confounded by ethnicity effects.

A study of mortality and hospital discharge rates by area deprivation in New Zealand in 1996/97 found that Maori under 25 years of age and Maori men aged 65 and over had lower hospital discharge rates than Europeans, after controlling for area deprivation (Salmond and Crampton 1999), and a study of general practitioner (GP) consultation rates in 2001 found that after controlling for area deprivation and other factors, Maori, Pacific, Asian and other ethnic groups had similar or lower rates of GP consultations than Europeans (Health Utilisation Research Alliance 2006). This is surprising given the high mortality rates of Maori compared to Europeans in all deciles, and may reflect differences in health seeking behaviours. Hefford *et al.* (Hefford, Crampton *et al.* 2005) indicated that ethnicity in New Zealand has an independent impact on health status, separate from deprivation indicators.

Despite differential health seeking behaviours between ethnic groups indicating a separate 'ethnicity effect' on health after controlling for area deprivation, and despite the over-representation of Maori and Pacific peoples in the most deprived deciles, ethnicity is highly problematic when included in a time series. There have been many changes in the recording of ethnicity in both hospital and mortality records in New Zealand in the last few decades, with limited consistency even at a single point in time. Ethnicity data are therefore unreliable at best, and particularly problematic when a time series is considered. The New Zealand Census Mortality Study (NZCMS) highlighted the problem of under-representation of Maori and Pacific in mortality statistics (Ajwani, Blakely *et al.* 2003; Blakely, Ajwani *et al.* 2004; Blakely, Tobias *et al.* 2005). Maori and Pacific peoples are often misreported on death records, and possibly also on hospital records. So using census data as a population denominator, which has much more reliable ethnicity information, may introduce numerator-denominator bias. Due to the lengthy time period considered in this study, the difficulty in obtaining population data by ethnicity for Census Area Unit, and the unreliability of ethnicity data over time, ethnicity will not be directly considered in the analysis.

No measure of area socioeconomic deprivation is perfect. However, NZ*dep* is preferable to other measures due to the inclusion of income. It has been shown that area

deprivation measures can be meaningfully used in the absence of individual deprivation measures. While NZ*dep* is confounded by ethnicity, it is the best measure of deprivation available and is considered to be of high enough quality for this research, as long as the limitations are taken into account when interpreting results.

1.3 Health and Hospital Utilisation

This section discusses measures of health, focussing on hospital utilisation in the New Zealand context. The relevance of hospital utilisation as a measure of morbidity is discussed, along with the limitations of such applications.

1.3.1 Health

Health is problematic to define, and even more so to measure. Until relatively recently, mortality was widely used as an indicator for health. Mortality is a powerful, robust indicator. As a vital statistic it is routinely collected allowing comparison over time and between countries. Compared to morbidity indicators it is easy to measure, and certain applications provide good indicators of population health in general. Infant death rates are often used to infer information on the health of the general population in countries where morbidity data are unreliable or unavailable. Applying age-specific death rates to a life table yields a range of statistics including life expectancy, which can be compared across populations with different age and sex structures.

However even a cursory examination of the epidemiological transition theory highlights the changing patterns of cause of death, with associated changes in the length of ill health preceding mortality. Although mortality is ultimately the inevitable end result, measuring mortality alone does not take into account morbidity prior to death. Not all sicknesses result in mortality, and not all mortalities are preceded by sickness. Mortality rates do not differentiate between deaths with a long period of preceding sickness, deaths with a short period of preceding sickness, and deaths without preceding sickness, and they completely miss sickness that does not result in death. And finally, mortality data represent people who have already died. They do not relate to the current population, but to the past, to people who are no longer alive. However it could be argued that the counterpart of mortality, survival *does* relate to people still alive.

Intuitively it is assumed that if mortality declines, morbidity also declines. However as is discussed in Paper Two, this is not necessarily the case. If mortality declines result in people surviving longer at older ages, this may result in an *increase* in morbidity overall because it is at older ages that the risk of morbidity is greatest (Johnstone, Cheung *et al.* 1998).

Mortality is clearly not a sufficient measure of morbidity because it measures the *quantity* of life. What are also needed are morbidity measures, which measure the *quality* of life (Ministry of Health 2000). No measure of health can be all-encompassing, but an ideal measure combines both mortality and morbidity (Johnstone, Cheung *et al.* 1998).

The most commonly used measure of morbidity is self-reported health. Health questions are included in censuses for some countries (UK) and in specially designed health surveys. Self-reported health often benefits from large sample sizes. While difficult to define, health is a concept which everyone intuitively understands, so it is logical to simply ask people about their health. However, research has repeatedly found that individuals are not necessarily the best judge of their health status. People can interpret questions differently, and perceptions of health differ by a range of socioeconomic and demographic factors such as age, gender, and particularly ethnicity.

Much research supports self-reported health. A recent study found self-reported prevalence of chronic disease to follow the same pattern as biological markers, indicating no bias in self-reported health by socioeconomic status (Banks, Marmot *et al.* 2006). However bias in self-reported health by a range of factors including gender, ethnicity, age and socioeconomic status, has been shown in many other studies. SF-36 is a widely used form of health survey that was designed to minimise bias resulting from different perceptions of health and variations in the wording of questions on health in surveys, and it has been found to be a reliable and valid indicator of health status in populations with differing health perceptions (Brazier, Harper *et al.* 1992). However the validity of SF-36 in populations with ethnic diversity has been questioned. Maori and Pacific health models in New Zealand stress collective health of the community and family, or *whanau*, rather than individual health, and have little differentiation between physical and mental wellbeing. A New Zealand study found the SF-36 form of self-reported health to be a good measure for Western health models, but its cross-cultural validity was poor for non-Western health models such as Maori and Pacific (Scott, Sarfati *et al.* 2000).

It has been suggested that hospital episodes may be better at indicating the presence of a serious illness than self-reported health, following findings that hospital use more

closely mirrors sex-specific mortality trends than self-reported health, after controlling for various factors (Case and Paxson 2005). However, hospital use is a questionable proxy for morbidity, given that it does not measure health, merely the uptake of health services which are affected by a wide range of factors other than the health of the population.

1.3.2 Hospital Utilisation as a Proxy for Morbidity

Using hospital data to infer population health characteristics is by no means a new concept. There is a growing body of research that has exploited administrative health service data for this purpose, in New Zealand and elsewhere (Salmond and Crampton 1999; Cheung, Katzenellenbogen *et al.* 2001; Hessler, Eriksson *et al.* 2003; Case and Paxson 2005; Riddell 2005; Pool, Baxendine *et al.* 2009). Public hospital data are routinely collected in many countries and present a picture of a population's use of health services over time, by cause, and by a range of demographic indicators. People do not go into hospital unless it is absolutely necessary. Time spent in hospital is very expensive, thus government and insurers aim to keep stays in hospital as few and as short as possible, while inferring maximum health benefit to the patient (Case and Paxson 2005). It can be safely assumed, with a few notable exceptions such as childbirth, that if you are in hospital you are not in the best of health, and if you are in hospital for a prolonged period, you are very unwell indeed.

Some studies have employed a range of morbidity and mortality measures, including hospital use, to examine health. Case and Paxson examined the paradox that while men have better self-reported health and use fewer health services than women, they have higher mortality rates at every age. They used self-reported health measures from the American National Health and Nutrition Examination Survey (NHANES), which includes questions on use of health services, and linked these to mortality events for each individual. It was found that after controlling for chronic diseases and other factors, differences in self-reported health by sex became negligible, but differences in hospital use and mortality remained (Case and Paxson 2005). It was suggested that men may experience more severe forms of certain diseases, particularly smoking related, than women, necessitating more hospital visits and resulting in higher mortality rates from these diseases, and that this may be largely due to greater lifetime exposure to smoking among men. However, as with many other studies of hospital use only the number of discharges is considered, without taking into account length of stay. It could be that women with the

same disease have fewer but longer stays in hospital than men, who are more often treated as day-patients due to less severe symptoms.

The number of hospital discharges is a very crude measure, which is not weighted and does not take account of the 'severity' of each episode. Eachus *et al.* stress the need to use severity of health conditions, not just prevalence as a measure of morbidity (Eachus, Chan *et al.* 1999). Pool also emphasises the need to incorporate duration of illness and not use incidence alone as a measure of morbidity, particularly when comparing morbidity and mortality trends (Pool 1994).

In addition, hospital discharge rates are highly sensitive to changes in policy. Transfers between hospitals, or even departments within a hospital, may be recorded as re-admissions, and drives towards day-patient treatment may result in an increase in the number of discharges due to people having lots of short stays in hospital instead of fewer longer stays. Neither of these scenarios would affect the number of hospital bed days, which are much more stable over time.

1.3.3 Hospital Utilisation and Area Deprivation

A New Zealand Ministry of Health commissioned study of mortality and hospital discharge rates by area deprivation found strong socioeconomic gradients of mortality and hospital use for a wide range of causes between 1996 and 1997 (Salmond and Crampton 1999). A study of avoidable hospitalisations in New Zealand also found mortality and hospital discharge rates to be positively associated with area-level socioeconomic deprivation (Jackson, Kelsall *et al.* 1998). Another New Zealand study incorporated area deprivation as a control, but did not examine hospital use trends by area deprivation itself (Raymont 2008). There has also been research on hospital discharge rates by area, and individual deprivation in the US and the UK. Most research on hospital use trends by area deprivation in New Zealand (Jackson, Kelsall *et al.* 1998; Salmond and Crampton 1999) and elsewhere (Billings, Zeitel *et al.* 1993; Hofer, Wolfe *et al.* 1998) have used hospital discharge rates without weighting for length of stay. This paper rectifies this gap in the literature, presenting carefully filtered hospital bed day rates for a range of causes in New Zealand over time and by gender, age and area deprivation.

1.3.4 Limitations of Hospital Data

Hospital data are not collected for the purposes of population health research, but to monitor spending and efficiency in the health system. It is not a measure of health, but a

measure of hospital utilisation, which is affected by a complex web of factors, population health being but one. Supply and demand are the two driving forces behind hospital utilisation. High hospital utilisation may indicate adequate levels of supply (hospital services) or high demand driven by poor population health. Low hospital utilisation could indicate poor supply due to insufficient hospital services available or low demand resulting from good population health.

The efficiency of primary and secondary health services are not taken into account. Good primary health services and uptake of these services may reduce hospital stays by treating conditions that are amenable to intervention at the primary level. If primary health services are of poor quality or underutilised, this may result in an increase in emergency and acute hospital admissions resulting from conditions that could have been treated or prevented at a primary level, but have progressed to a level of severity that necessitates hospital admission. Public hospital care in New Zealand is free[2] but GP consultations and prescription medication incur a charge. Primary health care is subsidised - for households on low incomes prior to 2002 and for areas with high socioeconomic deprivation from 2002 onwards - but despite these subsidies there is some evidence of barriers to primary health care for low income groups in New Zealand (Hopkins and Cumming 2001; Hefford, Crampton *et al.* 2005).

The New Zealand data only includes public hospital events, and publicly funded events that occurred in private facilities. The absence of private hospital data is a clear confounder of area deprivation. It would be expected that people residing in less deprived areas have higher levels of private health care. Thus low hospital bed day rates for less deprived areas may be due either to better health, or to higher use of private health care. Other health facilities that are not included are residential care homes and mental health facilities. However, given free public hospital care in New Zealand, the inability to afford healthcare for people living in deprived areas is not as great an issue as it is in other countries, such as Australia and the US. This should result in deprived areas *not* having artificially low utilisation of health services.

During the period covered, New Zealand went through extensive economic and health system reforms. There are numerous coding changes in the way different types of hospital stays are recorded, with many types of hospital stays, such as day patients only being recorded later on in the time period, or being unreliably recorded. ICD coding

[2] With some exceptions between January 1991 and July 1997, when charges were introduced

changes over the time period also makes tracking causes of hospital use over time difficult. Finally, not all hospital stays represent ill health. Childbirth is the most obvious example. Because of these reasons hospital data requires extensive filtering to exclude certain cases. This process reduces the distorting effect of coding changes, allows comparability over time, and improves the relevance of hospital data as a proxy for morbidity, thus making hospital bed days a suitable proxy of morbidity for use in this study. The filtering process is discussed briefly in Section 2, and extensively in Appendix A.

1.5 Conclusion

The suitability of NZ*dep*, mortality and hospital discharge data as a basis for examining deprivation effects on health has been established. The single largest concern is the confounding nature of ethnicity effects on health in New Zealand, and the uneven distribution of ethnic groups across deprivation deciles, and this must be considered in interpreting the results. There is a large body of research on health patterns by individual- and area-deprivation measures. This study adds a new dimension to the body of literature concerning health patterns by deprivation measures by introducing hospital utilisation as an indicator of morbidity. However hospital data must first be extensively filtered to allow for compatibility over time and to improve the validity of hospital data as a proxy for morbidity. The next section describes the datasets and methods used in calculating a range of mortality and hospital bed day rates for New Zealand.

SECTION 2 - DATA AND METHODS: Preparation of the Datasets and Calculation of Mortality and Hospital Bed Day Rates

2.1 Introduction

This section presents the data and methods used in the calculation of mortality and hospital bed day rates by area deprivation in New Zealand. Some methods of analysis were considered, such as loglinear regression, but discounted due to the very large datasets involved. With a large enough number of cases, all variables in a regression model are statistically significant, making it difficult to tell which of these effects are important, and which result from the large sample. This paper aims to present a background picture of mortality and hospital utilisation in New Zealand. Unlike previous studies it combines hospital bed day rates (not discharge rates) with area socioeconomic deprivation. The use of hospital bed day rates rather than discharge rates is a relatively novel approach. This paper also provides a solid introduction to the analyses presented in papers two and three of this series. Two different types of analysis are conducted in Section 3, a national analysis from 1974 to 2004, and an analysis by area deprivation over a shorter time period from 1991 to 2006. In this section the datasets are described, followed by aspects of the calculations. These include the decision to use hospital bed day rates in favour of discharge rates, the problems of hospital transfers in relation to classifying cause of hospital use, a recap of the filters and denominators (*de jure* and *de facto*) used for national and deprivation analyses, and a description of the confidence intervals used in presenting cause-specific mortality rates. A description of the extensive cleaning and processing of the data are described in Appendix A.

2.2 Datasets

A wide range datasets were required for this analysis (see Table 2.2.1). All of the datasets (excluding the International Classification of Disease codebooks) relate to the entire New Zealand population.

Table 2.2.1: Description of Datasets

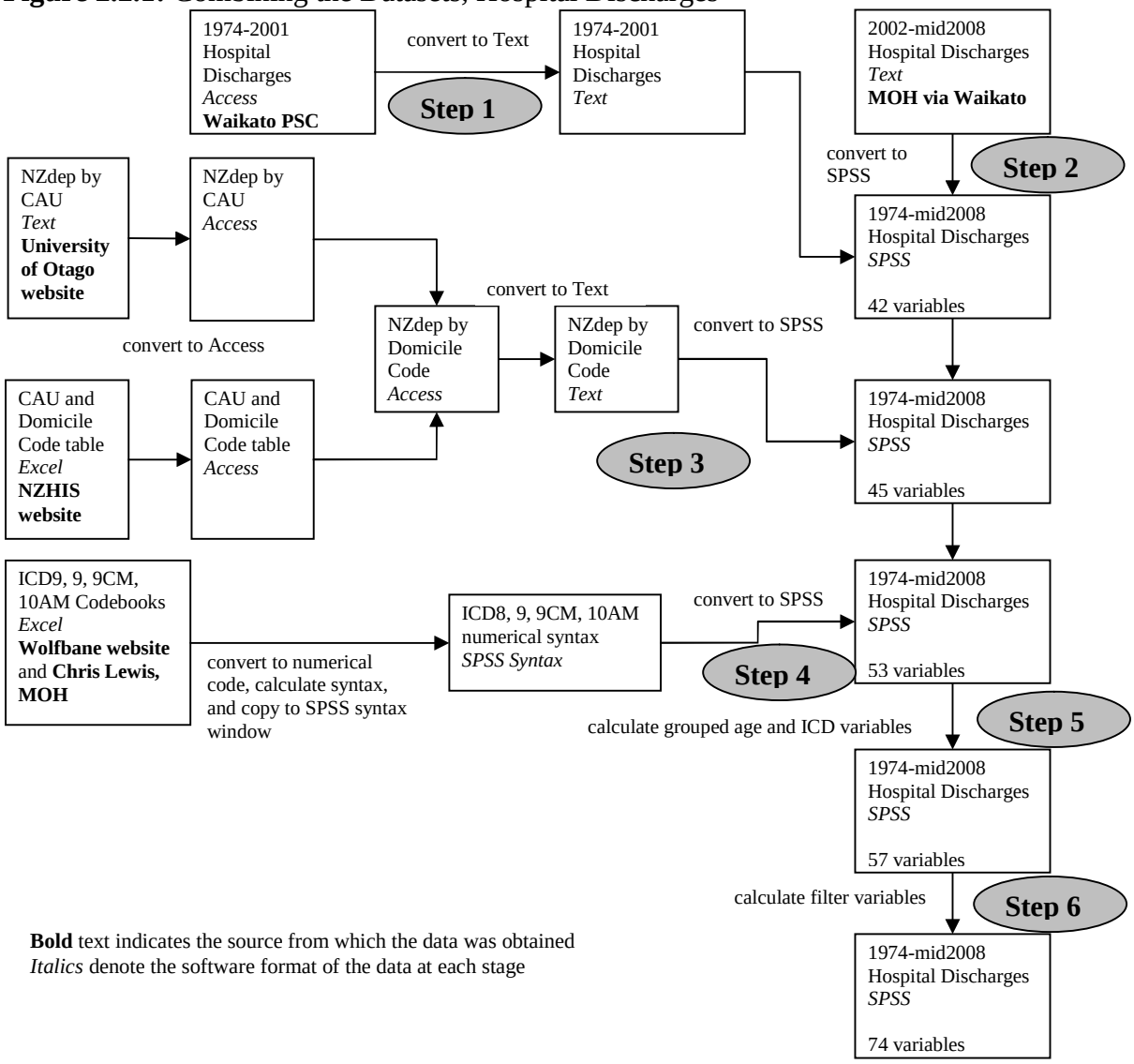
Years Covered	Data Type	Description	Format	Source
1974-2001	Hospital	Hospital discharges	Access	University of Waikato Population Studies Centre (PSC)
2002-mid2008	Hospital	Hospital discharges	Text	New Zealand Ministry of Health (MOH) (via University of Waikato PSC)
1974-2000	Mortality	Mortality registrations	Access	University of Waikato PSC
2002-2006	Mortality	Mortality registrations	Text	MOH (via University of Waikato PSC)
1970-2010 (5-yearly)	Population	Mid-Year *de facto* National Population estimates	Excel	United Nations Population Division Website (United Nations Population Division Website 2009)
1991	Population	Mid-Year *de jure* population estimate by CAU[3]	Excel	Martin Tobias and Li-Chea Yeh MOH
1996-2008	Population	Mid-Year *de jure* population estimates by CAU	Excel	Statistics New Zealand customised data request
1991, 1996, 2001, 2006	Area Deprivation	NZ*dep* by CAU	Text	University of Otago Website (University of Otago Website 1988-2001)
1991, 1996, 2001, 2006	Area Deprivation	CAU to domicile code mapping	Excel	New Zealand Health Information Service (NZHIS) Website (New Zealand Health Information Service (NZHIS) Website 2008)
	ICD[4]	ICD8 Codebook	Excel	Wolfbane Website (Wolfbane Cybernetic Ltd. Website 2007)
	ICD	ICD9 Codebook	Excel	Wolfbane Website (Wolfbane Cybernetic Ltd. Website 2007)
	ICD	ICD9CM Codebook	Excel	Chris Lewis MOH
	ICD	ICD10AM Codebook	Excel	Chris Lewis MOH

The main datasets used contained information on public hospital events and mortality for the New Zealand population. Population estimates were required to provide denominators for hospital bed day and mortality rates at both a national level from 1974 to 2006, and at a Census Area Unit (CAU) level for the analysis by area deprivation from 1991 to 2006. Two additional datasets were required to enable calculation of rates by area deprivation. The area of residence is provided by domicile code[5] in the mortality and hospital datasets, but NZ*dep* is calculated for CAUs. To incorporate an area deprivation score for each mortality and hospital event it was necessary to convert from domicile code to CAU, and to identify the NZ*dep* score for each CAU in the relevant year. Finally, International Classification of Disease (ICD) code tables were required for each ICD version in the dataset. Each death and hospital event (including procedures) is assigned an ICD code

[3] Census Area Unit. CAUs are geographical areas with approximately 2000 residents.
[4] International Classification of Disease.
[5] Domicile codes are geographical areas with approximately 2000 residents, and are designed to map one-to-one with CAUs.

based on the cause of death, diagnosis or type of procedure. These code tables made it possible to filter out certain types of hospital discharges (discussed in section 2.4) and to classify broad causal groups for the mortality and the hospital data. The process of combining the datasets is shown for hospital discharges in Figure 2.2.1.

Figure 2.2.1: Combining the Datasets, Hospital Discharges



The datasets described in Table 2.2.1, are combined into one large dataset for hospital events and one large dataset for mortality events, as shown in Figure 2.2.1. These datasets contain information on the ICD code, the CAU and the NZ*dep* score for each event, in addition to the original mortality or hospital event information.

Figure 2.2.1 only shows the process for the hospital dataset, a similar process was undertaken for the mortality data with the exception of the creation of filter variables. The

bold text in Figure 2.2.1 indicates the source from which the data was obtained. Italics denote the software format of the data at each stage.

Steps one and two, respectively, document the conversion of the hospital data into SPSS format, and the combining of the two hospital datasets into one large dataset that covers the entire time period 1974-mid 2008. At step three, three area deprivation variables are added, increasing the number of variables in the hospital dataset from 42 to 45. At step four, eight ICD variables are added, increasing the number of variables in the hospital dataset from 45 to 53. At steps five and six no new data are added to the hospital dataset, the existing variables are recoded into new variables. Step five shows the addition of four grouped age and ICD variables. This enables the calculation of hospital statistics by age group and broad diagnostic group, an essential level of aggregation for calculating age-specific hospital bed day rates for different ICD groupings. In step six, 19 filter variables are created for the hospital data for the purpose of excluding certain types of cases (this process does not occur for the mortality data). The reasons for the filtering process are discussed in section 2.4.

The population data are conspicuously absent from Figure 2.2.1. This is because the population data are not combined with the mortality or hospital data, but used as a denominator when calculating rates in Microsoft Excel.

The process of preparing the datasets for analysis as detailed in Figure 2.2.1 requires discussion, however given the technical nature of this information it is presented elsewhere. Each of the six steps in Figure 2.2.1 is examined in detail in Appendix A. In this section the datasets are each discussed in turn.

2.2.1 Hospital Data

Hospital data were obtained from the New Zealand Ministry of Health through the University of Waikato Population Studies Centre (PSC)[6].

Hospital data are available from 1974 through to mid-2008, but analyses in this paper use hospital data up until 2006. While hospital and mortality events are considered separately in this paper, in the two subsequent papers mortality and hospital data are combined, and due to the more limited time span for the mortality data, analysis is only possible up to 2006. There were 18,103,912 hospital discharges between 1974 and 2006.

[6] The Population Studies Centre (PSC) merged with Waikato Management School and Motu Economic and Public Policy Research Trust to found the National Institute of Demographic and Economic Analysis (NIDEA) in 2010. As such, the PSC no longer exists. However, given that the PSC were the providers of the data, they shall be referred to throughout as PSC, not NIDEA.

Of these, 350 cases were excluded due to unspecified sex, and just five cases were excluded due to having an ICD code that was not recognised, leaving 18,103,557 hospital discharges. Of these, 6,552,868 were excluded through filtering for the national analysis (36.2 percent). Discharges were excluded through filtering for reasons such as not being representative of ill health, and inconsistency of recording and coding over time. Excluding over a third of hospital discharges may seem severe, but this study excludes slightly fewer discharges than previous studies (Pool, Baxendine *et al.* 2009). For the area deprivation analysis, there were 11,531,239 hospital discharges between 1991 and 2006 (excluding those cases with unspecified gender or an invalid ICD code). A narrower time period is used for the deprivation analysis due to the unavailability of area deprivation data before 1991. A slightly different filtering process is used for analysis by area deprivation, as discussed in section 2.4. This resulted in 2,971,432 cases being excluded through filtering (25.8 percent) and 509,215 cases were excluded due to not having a valid NZ*dep* score (4.4 percent). These two figures are not cumulative; there was some overlap between filtering and lack of valid NZ*dep* score. Overall, exclusion through filtering and (for area deprivation analyses) lack of a valid NZ*dep* score excludes 36.2 and 28.9 percent of all hospital discharges for national and area-deprivation analyses respectively. The filtering process is discussed in more detail in section 2.4.

More variables are included in the hospital dataset in later years. This limits the number of variables that can be used for the national analyses, which goes back to 1974. The variables include date of birth, gender, ethnicity (which is not used in these analyses), date of admission, date of discharge, admission type (acute, elective, transfer), discharge type (routine, transfer, dead on discharge), various ICD codes for diagnoses, operations and procedures, and domicile code (of the patient's usual address, not the hospital). For a full list of available variables, see Appendix A.

2.2.2 Mortality Data

Mortality data were obtained from the University of Waikato Population Studies Centre (PSC), and from the New Zealand Ministry of Health (via the PSC).

The mortality dataset contains far fewer events than the hospital discharges dataset. An individual can be admitted to hospital many times in their lifetime, but only dies once. Unfortunately, due to an administrative error the dataset excludes mortality for 2001. All results presented for mortality or combined mortality and hospital measures in this paper and the two subsequent papers, exclude data for 2001. Between 1974 and 2006 (excluding

2001) there were 857,554 deaths recorded in New Zealand, 17 of which did not have a recognised ICD code and one of which had unspecified gender, leaving 857,536 deaths. Between 1991 and 2006 (excluding 2001) there were 414,009 deaths (excluding deaths without a recognised ICD or unspecified gender). For the area deprivation analysis (1991-2006) 18,475 cases did not have a valid NZ*dep* score (4.5 percent).

Fewer variables are included in the mortality data than in the hospital data. Variables include date of birth, date of death, gender, ethnicity (not used in these analyses), occupation (a free text variable only included in later years), domicile code, and various ICD codes relating to the cause of death. For a full list of available variables, see Appendix A.

2.2.3 Population Data

National mid-year *de facto* population estimates for the New Zealand population were obtained from the United Nations (UN) Population Division website. Population estimates were only available for every 5 years, and linear interpolation was used to generate yearly population estimates. For the analysis by area deprivation, mid-year *de jure* population estimates by CAU were obtained for 1991 (from contacts at the New Zealand Ministry of Health (MOH)), and for each year from 1996 to 2008 from a customised Statistics New Zealand data request. This again required linear interpolation between 1991 and 1996. Further discussion of *de facto* and *de jure* population estimates, and the reasons for using *de facto* estimates for the national analysis and *de jure* for the area deprivation analysis are addressed in section 2.5.

Mid-year estimates are required for population denominators used in calculating rates. Assuming that any change in the population size and structure occurs at a constant rate over the year, mid-year population estimates provide a proxy for the number of 'person years' at risk of experiencing an event, in this case the risk of being admitted to hospital or dying. Alternatively it can be considered as the average number of people exposed to the event at any given point in the year.

The population of New Zealand grew by over one million during the time period, from 3.0 million in 1974 to 4.2 million in 2006.

2.2.4 Area Deprivation Data

The two datasets involving area deprivation are of little merit on their own, but are necessary to identify the NZ*dep* score for each event. An Excel table mapping domicile

codes to CAUs for different census years was obtained from the University of Otago website, and an Excel table listing the NZ*dep* score by CAU for each census year was obtained from the New Zealand Health Information Service (NZHIS) website. These two Excel files were used to firstly match the domicile code to a valid CAU, and secondly to match the CAU to a relevant NZ*dep* score for each event in the mortality and hospital datasets. This process is described in detail in Appendix A. Domicile codes are updated at every census, if the population in an area has declined, the domicile code may be merged with another area, and if the population of an area has increased, it may be split into two domicile codes. The proportion of events with domicile codes that can be converted to CAUs has increased over time. 99.5 percent of deaths and hospital events in 2006 had a valid CAU, compared to just 92.0 percent of deaths and 90.9 percent of hospital events in 1991. This increase occurred largely between 1997 and 1999, due to a change in the domicile code version recorded in 1998 (see Appendix A for further discussion). NZ*dep* scores can be calculated for virtually all cases with a valid CAU, for most years and both datasets at least 99.9 percent of cases with a valid CAU had a valid NZ*dep* score. Three exceptions are 1991 and 1992 for mortality, where this figure was 99.5 percent due to coding changes (see Appendix A), and 2006 for hospital events where only 94.8 percent of valid CAUs could be matched to an NZ*dep* score. Reasons for this drop in matches in 2006 are unknown, but as the percentage of valid CAUs matched to an NZ*dep* score is still in the mid 90s, it is considered sufficient for the analyses.

2.2.5 International Classification of Disease (ICD) Data

Codebooks for each ICD version were obtained from a contact at the MOH, and from a website (Wolfbane Cybernetic Ltd. Website 2007). As with the area deprivation data, these codebooks are of little interest on their own. They were used to create grouped cause of death/hospitalisation categories that were consistent across the entire time period. This process is described in detail in Appendix A.

2.3 Hospital Bed Days and Discharges, Which is the Best Measure?

For all analyses using the hospital dataset, hospital bed days are used instead of the number of discharges. This section discusses the rationale behind this decision. The number of

hospital discharges for an individual is simply the number of times the individual is admitted and subsequently discharged from hospital. This can also be referred to as the number of hospital stays, or as the number of events. The number of hospital bed days for an individual is the total number of days that individual spends in hospital, regardless of the number of discharges. Thus an individual may have four separate hospital events, and be discharged four times, but spend 30 bed days in hospital over these four events. The length of stay in days for each event is required to calculate the number of bed days. The length of stay is obtained by calculating the difference between the admission date and the discharge date (in days). Thus a stay of one day represents a discharge one calendar day after being admitted, rather than a stay of at least 24 hours. Individuals which were admitted and discharged on the same day are coded as having a length of stay of 0.5, assuming that on average they were in hospital for half a day. The length of stay for each event is added together to give the total number of bed days. This is then divided by the population at risk to give a bed day rate. Discharge rates are simply the number of events divided by the population at risk. An individual who is discharged on the same day without staying overnight, and an individual who is admitted for a month are therefore considered the same when discharge rates are used. This is clearly not an ideal measure. In contrast bed day rates weight hospital events by the number of days spent in hospital and provide some measure of severity of the event. This is an essential adjustment when using hospital events as a proxy for morbidity.

The number of unfiltered discharges between 1974 and 2007 increased steadily, but the number of hospital bed days - the number of person-days spent in hospital - actually declined slightly (see Figures 2.3.1 and 2.3.2)[7]. This indicates a decline in the mean length of stay, in part due to a policy-driven increase in the number of day patients, which contribute heavily to the overall number of discharges but not to the overall number of bed days. The number of discharges is far more vulnerable to coding changes than the number of bed days. In addition, when a patient is transferred from one facility or ward to another they may be discharged and re-admitted, constituting two discharges. While it is misleading to classify this as two separate discharges, there are also concerns in dealing with hospital transfers when using bed day rates. The diagnosis of the first hospital admission may be different to the diagnosis given for the re-admission (when the patient is

[7] The large spike in bed days in Figure 2.3.2 between 1998 and 2006 is due to the temporary inclusion of publicly funded events occurring at private facilities, such as rest homes (Pool, Baxendine *et al.* 2009). These events had very long average stays, thus increasing total bed days, but not causing a spike in the number of discharges (Figure 2.3.1).

transferred to another ward or facility). So an individual admitted for surgery due to a heart condition may have only the first, short part of their hospital stay in an acute ward categorised as circulatory, the longer recuperation in a non-acute ward may not be categorised as circulatory. This is however, speculation.

Figure 2.3.1: Number of Unfiltered Hospital Discharges, 1974-2007

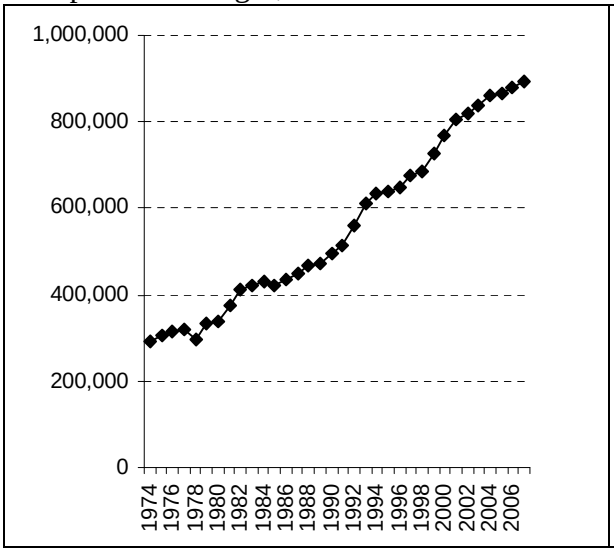
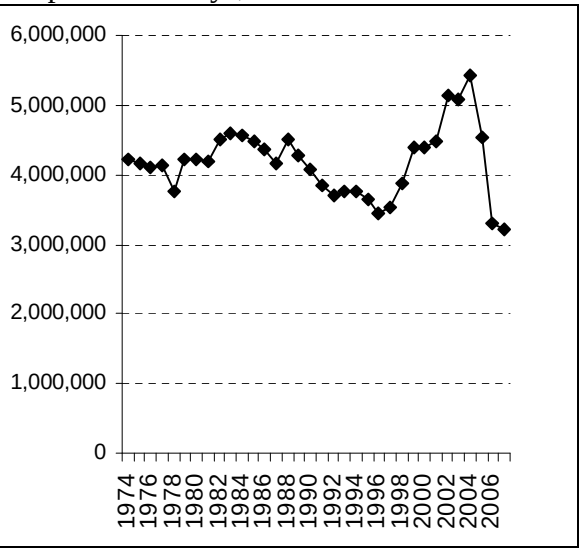


Figure 2.3.2: Number of Unfiltered Hospital Bed Days, 1974-2007



A secondary analysis linking hospital transfers and applying the initial diagnosis to the entire stay will be undertaken later on in this paper. This will identify whether hospital transfers are in reality a problem when analysing hospital use by cause. The number of hospital bed days in Figure 2.3.2 appears much more variable and less stable than the number of discharges in Figure 2.3.1. However, hospital data requires extensive filtering before it can be used in health research. The next section describes the filtering process.

2.4 Filtering the Hospital Data

The hospital data requires some types of discharges to be excluded to make it comparable over time, and to exclude cases that do not represent ill health. This is achieved through a filtering process, in which filter variables are calculated according to a set of criteria, which equal 0 if the criteria are not met, and 1 if the criteria are met. Cases are then excluded based on these filter variables. This section discusses the filtering process undertaken.

Two separate filters are used, one for the national analysis and one for the analysis by area deprivation. Because the deprivation analysis only uses data from 1991 onwards

(before which no domicile code data are available) it is not necessary to exclude day patients, which are not reported before 1990. The results from the national and deprivation analyses are not intended to be directly compared.

The cases excluded from the national analysis are:

- Non-New Zealand residents
- Boarders (healthy people staying in hospital)
- Obstetrics/pregnancy
- Well babies
- Baby boarders (healthy babies accompanying others)
- Stay > 365 days
- Day patients
- Private hospital without procedure

The cases excluded from the area deprivation analysis are:

- Non-New Zealand residents
- Boarders (healthy people staying in hospital)
- Obestetics/pregnancy
- Well babies
- Baby Boarders (healthy babies accompanying others)
- Stay > 365 days
- Same day chemotherapy/radiotherapy/renal dialysis
- Same day gastroscopies/colonoscopies/cystscopies
- Same day blood transfusions[8]
- Private hospital without procedure

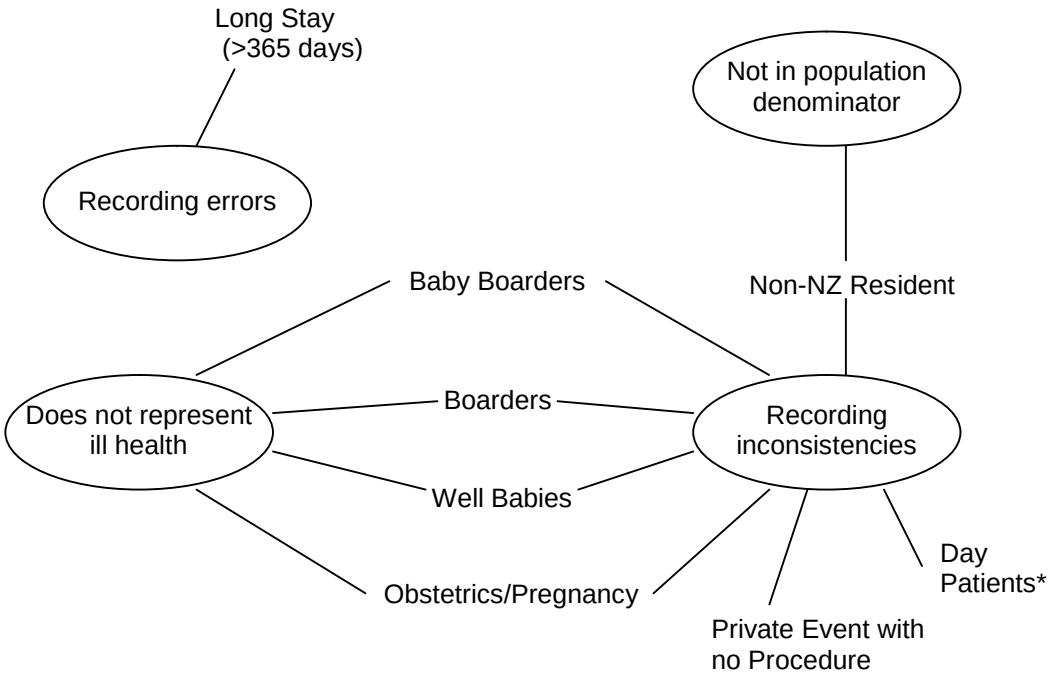
The main difference between the two filters is the inclusion of day patients in the analyses by deprivation, and the exclusion of day patients in the national analyses. Day patients were not recorded in the hospital data until 1988, but since 1988 there has been a steady increase in the number of day patients. Because they were not recorded before 1988, they must be filtered out for the national analyses (which go back to 1974) but they can be

[8] For the national analysis it was not necessary to separately exclude same day chemotherapy/radiotherapy/renal dialysis, gastroscopies/colonoscopies/cystscopies or blood transfusions because these cases are included in the day patients filter.

included in the analyses by area deprivation, which only runs from 1991. When filtering out day patients, this by default filters out same day chemotherapy, radiotherapy, renal dialysis, gastroscopies, colonoscopies, cystscopies and blood transfusions (which are inconsistently recorded over time), so there is no need to exclude these separately from the national analysis. However these cases do need to be separately excluded in the analysis by area deprivation, for which day patients are included.

Reasons for filtering out cases come under four main categories: recording errors, non-representation of morbidity, recording inconsistencies over time, and non inclusion in the population denominator, although many of the filters used fall under several of these categories. Figure 2.4.1 shows the reasons for the creation of the filter variables used in the analyses. The text in the circles gives the reasons for exclusion of certain types of hospital discharges.

Figure 2.4.1: Reasons for Excluding Cases from the Hospital Dataset



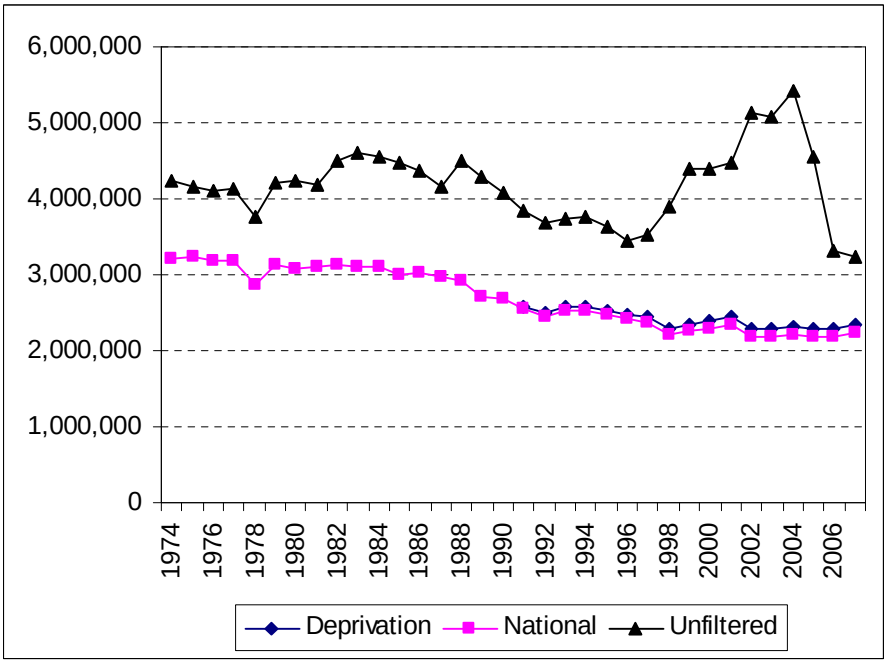
* Most Day Patients included in Deprivation analyses

The monograph this research builds on (Pool, Baxendine *et al.* 2009) uses a wider range of filters, including mental health, rehabilitation, Disability and Support Services (DSS), and non-Crown Health Enterprise (CHE) events. The number of hospital bed days that are due to mental and behavioural disorders are increasing over time. All mental health events apart from dementia were excluded from the monograph analyses, primarily due to the very long stays often found for mental health events. It is impossible to differentiate

between actual long stays and recording errors, and very long stays have disproportionate leverage over results of hospital bed day rate analyses. Similar reasons were used for excluding respite, rehabilitation, DSS, and non-CHE events. However, when stays over 365 days are excluded, the problems associated with these types of events diminish. For this reason, as well as the inability to filter out many of these types of events going back to 1974, mental health, rehabilitation, respite, DSS and non-CHE events are left in the dataset, and long stay is filtered out.

A substantial proportion of all hospital bed days are excluded through the filtering process, as shown in Figure 2.4.2. 76 percent of hospital bed days are included in the national analyses in 1974, but this drops to just 41 percent in 2004 before climbing to 69 percent in 2007. The reason for the small percentage of bed days included in the national analysis in 2004 is the temporary inclusion of publicly funded events occurring at private facilities - particularly nursing homes - with no procedure carried out (Pool, Baxendine *et al.* 2009). This is the reason for the artificial spike in unfiltered hospital bed days between 1998 and 2006, peaking in 2004. For the deprivation analysis, the percent of bed days included is slightly higher due to the inclusion of most day patients, but these events have short stays of 0.5 days and while they make up a large proportion of discharges, they do not contribute much to the overall number of hospital bed days. After filtering, the number of hospital bed days over time is smoothed and the decline in bed days over the time period is more pronounced.

Figure 2.4.2 Number of Hospital Bed Days, 1974-2007, Various Filters



2.5 *De jure* and *de facto* Population Denominators

The other difference required in the calculation of hospital bed day and mortality rates by national and area deprivation, is the denominator. It was necessary to use *de jure* mid-year population estimates for the deprivation analyses and *de facto* mid-year population estimates for the national analyses.

De jure population estimates record only usual residents, whereas *de facto* estimates include visitors, such as tourists, students and temporary migrants. For analysis by area deprivation, the mid-year population is required by Census Area Unit (CAU), from 1991 to 2006. In 1991, the denominator for Statistics New Zealand mid-year population estimates by CAU switched from *de facto* to *de jure* estimates. This necessitates the use of *de jure* estimates for the area deprivation denominator. However, national *de jure* estimates were not available for the whole time period 1974 to 2006: the Statistics New Zealand mid-year national population estimates switch from *de facto* to *de jure* estimates in 1991. These estimates are therefore not comparable over the time period 1974-2006 and could not be used for the national analysis. Other *de jure* population estimates were not available, so a solution was to use 5-yearly *de facto* mid-year population estimates available from the UN Population Division website. Linear interpolation was utilised to obtain a mid-year population estimate for each year for the national analyses, and between 1991 and 1996 for the deprivation analyses. The practice of interpolating between 5-year estimates to obtain single-year estimates is accepted and widely adopted by demographers (Siegel and Swanson 2004). While using different types of population estimates for the national and for the deprivation analyses is not ideal, these analyses are not intended to be directly compared. Rather, they form two separate pictures of mortality trends in New Zealand from different angles. The sources for these denominators are given in Table 2.2.1. Having described the data sources and preparation, attention is now given to the methods used in the analyses.

2.6 Calculation of Hospital Bed Day and Mortality Rates

This paper presents exploratory analysis of mortality and hospital use trends, both over time and by area deprivation. Trends are examined by sex, age and broad diagnostic group. To achieve this, a combination of age-specific and age-standardised rates are employed.

The methods of calculation for these rates are straightforward. The most simple type of rate is a crude rate, for crude rates the number of hospital bed days or deaths in a given year is divided by the mid-year population estimate for that year. This method is affected by the age and sex structure of a population and can be misleading when comparing different populations. For these reasons crude rates are not used in these analyses. Age-specific rates are calculated in a similar way, but both the numerator and denominator are limited to a given age group. So the number of deaths to people aged 70 to 79 in a given year would be divided by the mid-year population aged 70 to 79 in that year. Direct standardisation was used for the age-standardised rates, with the 1991 New Zealand population used as the standard population (the standard population is presented in Appendix Table 1). This method gives the crude rate that *would* be expected *if* a population had the same age structure as an arbitrary, standard population. The advantage of this method is that (unlike with non-standardised crude rates) it allows comparison of mortality levels between populations with different age and sex structures. New Zealand's population structure changed over the time period studied, and differs markedly by deprivation decile. For these reasons, age-specific and age-standardised rates are used instead of crude rates, and rates are calculated separately for males and females.

Given the large number of hospital discharges, statistical errors resulting from small numbers are not a concern. However there are far fewer mortality events than hospital events. For this reason, confidence intervals will be used in the presentation of mortality rates. Confidence intervals highlight where rates are based on small numbers of observations and are therefore unreliable. The calculation of confidence intervals for mortality rates is now discussed.

2.7 Calculation of Confidence Intervals for Mortality Rates

While confidence intervals are more commonly used to generalise from a sample to a population, it is also appropriate to use them for population data. Population data can be treated as a sample from a hypothetical 'super' population. In this case confidence intervals are employed as a warning when there are small numbers in a cell of a contingency table. New Zealand has a relatively small population of 4.2 million, and over the time period 1974 to 2006 there were only 857,554 deaths. When the total number of deaths is divided by age group, sex, deprivation decile, broad cause of death and year, many combinations

have very few numbers of deaths. Applying confidence intervals to bar charts representing rates allows one to see at a glance whether one rate is 'significantly'[9] higher than another, or whether small numbers prevent such conclusions. Confidence intervals are only used for mortality rates, not hospital bed day rates due to the relatively large numbers of hospital bed days in comparison to deaths.

Following advised common practice (Washington State Department of Health Website 2009), where the number of events (deaths) in a cell were less than 100, Poisson confidence intervals were used, and where the number of events was 100 or greater, Normal confidence intervals were used.

The formulas for calculating 95 percent confidence intervals for mortality rates are given in Equation 2.7.1:

Equation 2.7.1: Formulas Used for calculating 95% Confidence Intervals

$$If\ events < 100\ C.I. = \frac{Critical\ upper\ or\ lower\ value\ from\ Poisson\ distribution\ for\ number\ of\ events}{mid\ year\ population}$$

$$If\ events \geq 100\ C.I. = \left(\frac{events}{mid\ year\ population}\right) +/- 1.96 \times \sqrt{\frac{events}{mid\ year\ population}}$$

2.8 Ethical considerations

The datasets used contained individual mortality and hospital records for the entire New Zealand population over several decades. While the data are anonymous, containing no information on name or address, it does include some sensitive variables and care was taken in crosstabulations to ensure a sufficient number of individuals in any cell/category to protect anonymity. Only aggregated results are presented in these analyses. For data from 1991 onwards an individual identifier is included in the mortality and hospital datasets in the form of an encrypted National Health Index (NHI) number, which remained the same for each individual across hospital and mortality events. In the third paper, this

[9] Treating the population data as a sample from a hypothetical population, statistical significance at a 95 percent level of confidence allows us to infer that if repeated samples are drawn from the same population, 95 times out of 100 the true value of the statistic will lie between the upper and lower bounds of the confidence interval. Although this is purely hypothetical (we have population data, there is no larger population from which this data can be considered a sample) it does allow us to identify where small numbers of observations may impair the validity of the results.

identifier is used to track individuals' use of hospital services in the months prior to death. Date of birth, CAU of residence, and dates of hospital and mortality events were also present. See Paper Three for further discussion of the ethical issues of using hospital records in health research.

No record-linkage was conducted for the analyses presented in this paper. However, due to the sensitive nature of the data, they were stored in a secure data facility courtesy of the Southampton Statistical Sciences Research Institute (S3RI). This facility is located in a locked room in a building with electronic card entry. The computers in this facility are not networked or connected to the internet, and access is limited to only a few individuals.

Clearance for this research was obtained from the Population Studies Centre at the University of Waikato, who provided the mortality and hospital datasets.

2.9 Conclusion

A wide variety of data types from different sources were brought together to create two datasets - a mortality dataset and a hospital dataset - and accompanying population denominators for the national and for the area deprivation analyses. This paper is restricted to simple analytical techniques, and aims to set the scene, provide a background from which to conduct more technical analyses. In the second and third papers in this series combined mortality and hospital measures will be used to examine morbidity over the life course and in the last few months of life. However, before such analyses are conducted, a firm understanding of the patterns present in the data needs to be achieved. The next section presents age-specific and age-standardised mortality rates over time, by area deprivation, by gender, and by broad diagnostic group.

SECTION 3 – RESULTS: Mortality and Hospital Bed Day Rates by Deprivation Decile

3.1 Introduction

In this section mortality and hospital bed day rates are examined with emphasis on how these rates differ by sex and cause, and how they have changed over time. This analysis is conducted at a national level over a long time period (1974 to 2006) and at an area-deprivation level over a shorter time period (1991 to 2006). This requires a carefully structured set of results. There are four groups of results. National mortality and hospital bed day rates are presented first in sections 3.2 and 3.3. The same are repeated for area deprivation in sections 3.4 and 3.5. In each of these sections the rates are explored in four ways: age-specific rates, age-standardised rates, cause-specific and age-specific rates, and cause-specific age-standardised rates.

Due to administrative reasons information for 2001 is missing from the mortality dataset, so mortality results exclude 2001. There is no reason to think that mortality in 2001 differs markedly from mortality in 2000 or 2002.

Although the main interest of this research is trends in mortality and hospital use *by area deprivation*, it is important first to conduct a national analysis. The national data give a baseline against which deprivation-specific patterns can be compared. Thus the first half of this section (3.2 and 3.3) focuses on national trends in mortality and hospital use. The reader is directed to sections 3.4 and 3.5 for the findings of the area deprivation analyses.

3.2 National Mortality Rates (1974-2006)

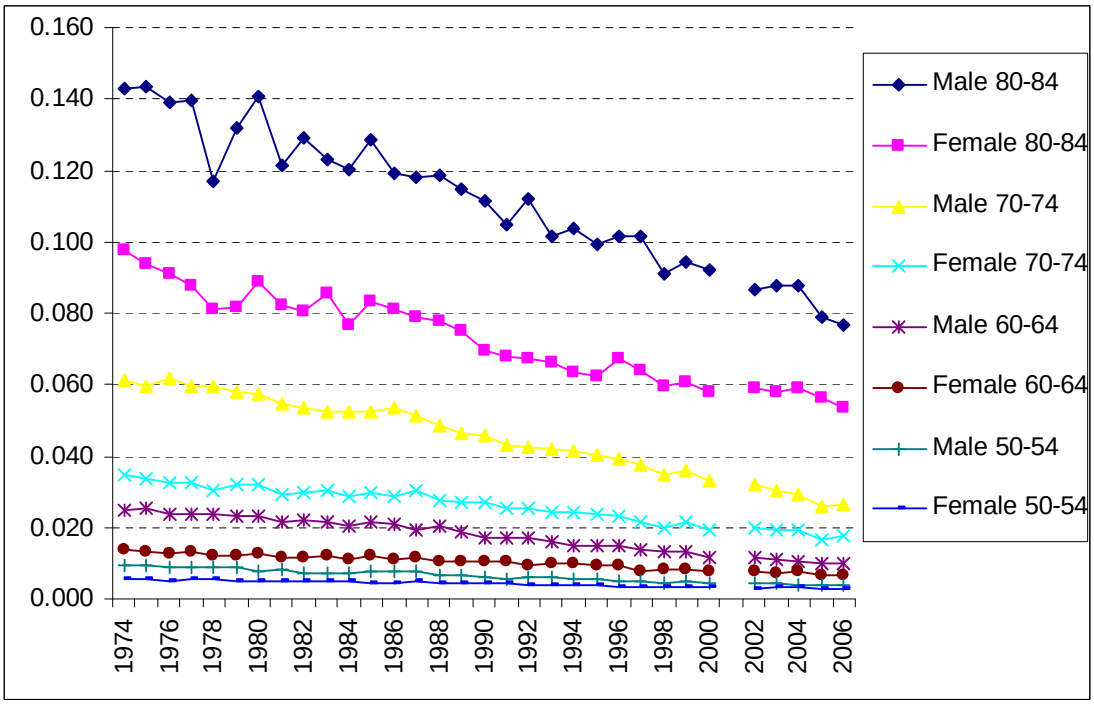
Mortality and hospital analysis by area deprivation measures can only be conducted from 1991 to 2006 giving a relatively limited time period of 16 years, however national analysis can be conducted from 1974 to 2006, allowing a longer time period of 33 years. Analysing at the national level enables longer term trends to be observed, which will help with interpretation of trends by deprivation from 1991 to 2006.

3.2.1 Age-Specific Mortality Rates, 1974-2006

This section examines crude and age-specific mortality rates. Figure 3.2.1 shows the declining trend in age-specific mortality rates at various ages, and both increased mortality rates and fluctuation in mortality rates at older ages. Males have higher mortality rates than females for every age group and every year, but male mortality rates also declined more over the time period. It is worth noting that the fluctuations in mortality rates at ages 80-84 (and at other older ages) are *not* due to small numbers of deaths. At ages 80-84 there were more than a thousand deaths every year from1974 to 2006 for males and for females.

Age-specific mortality rates approximately halved over the period, with the greatest relative declines seen at younger ages. Mortality rates at ages 0 to 14 declined by about two-thirds. Mortality for males at ages 70+ saw declines of 45 to 55 percent, however declines in mortality rates for females at ages 70+ were more modest, especially for ages 85+, which saw a decline of only 31 percent. Over the same period crude death rates saw relatively little decline, with male and female rates declining by 24 percent and 10 percent respectively. The modest declines in crude rates compared to large declines in age-specific rates indicate that the population age-sex structure changed markedly over the period 1974 to 2006.

Figure 3.2.1: Age-Specific Mortality Rates, 1974-2006, Selected Ages



Figures 3.2.2 and 3.2.3 compare mortality rates at every age group in 1974 and 2006. The most noticeable features are the large declines in, and convergence between

male and female mortality rates at ages 85+, with the male mortality rate declining from nearly 300 per thousand to just over 150 per thousand, and the female mortality rate declining from 200 to just under 150 per thousand. The gap between male and female mortality rates has also reduced in younger age groups (50 to 84).

Figure 3.2.2: Age-Specific Mortality Rates, 1974

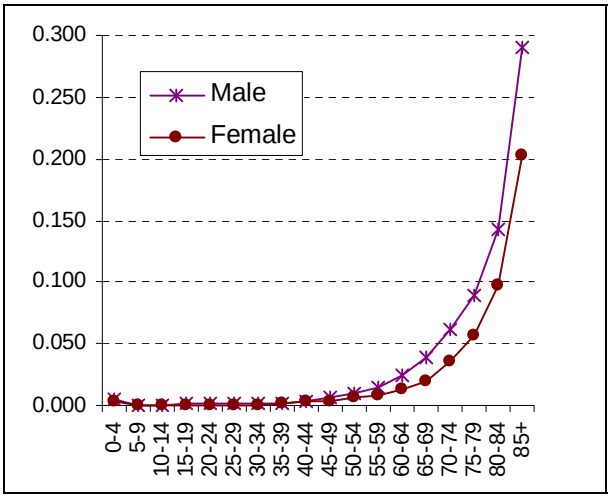
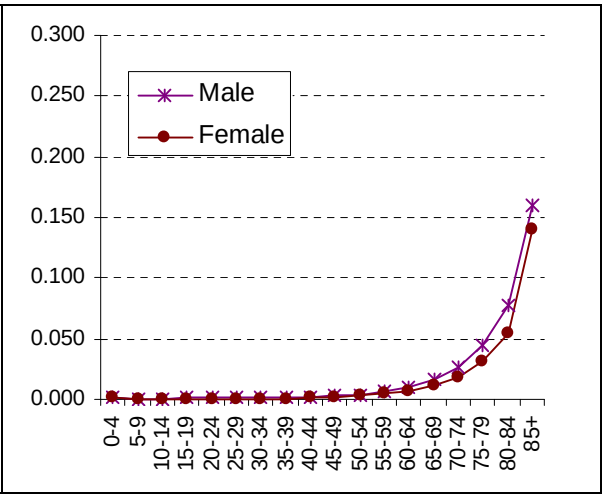


Figure 3.2.3: Age-Specific Mortality Rates, 2006

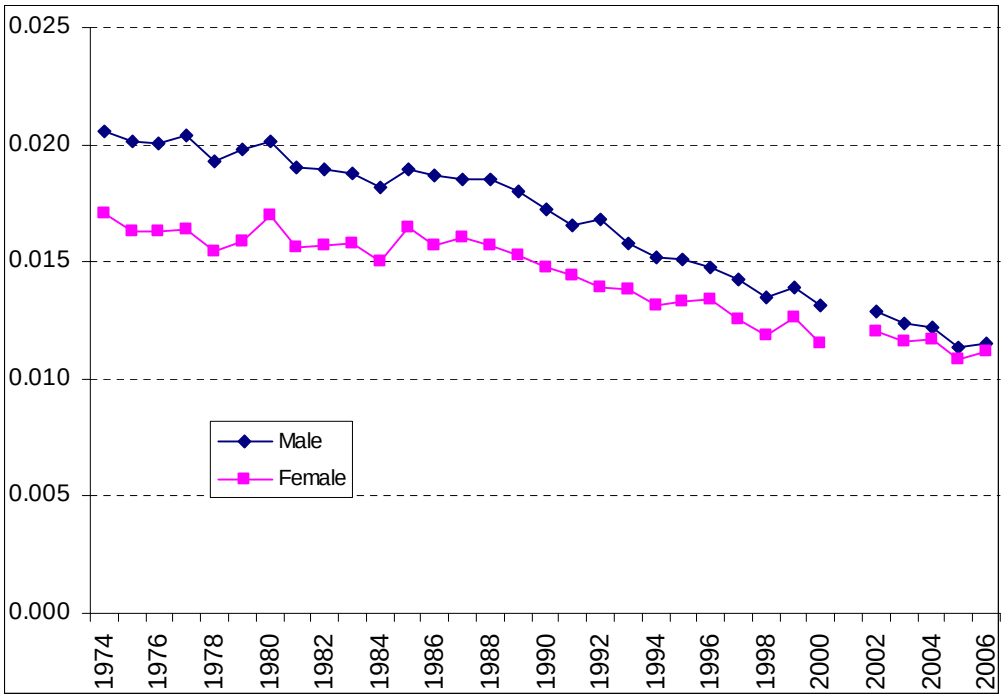


3.2.2 Standardised Mortality Rates, 1974-2006

The discrepancies between the crude and the age-specific declines in mortality discussed in section 3.2.1 highlight the need to examine age and sex standardised mortality rates. Figure 3.2.4 shows standardised mortality rates for the period 1974 to 2006.

Male standardised mortality rates showed the greatest decline, dropping from 20.5 to 11.5 per thousand. Female standardised rates declined by a smaller amount, from 17.1 to 11.2 per thousand. As with the age-specific rates in Figures 3.2.2 and 3.2.3, convergence is seen between male and female mortality rates, brought about by a large decline in male rates and a smaller decline in female rates.

Figure 3.2.4: Standardised* Mortality Rates, 1974-2006



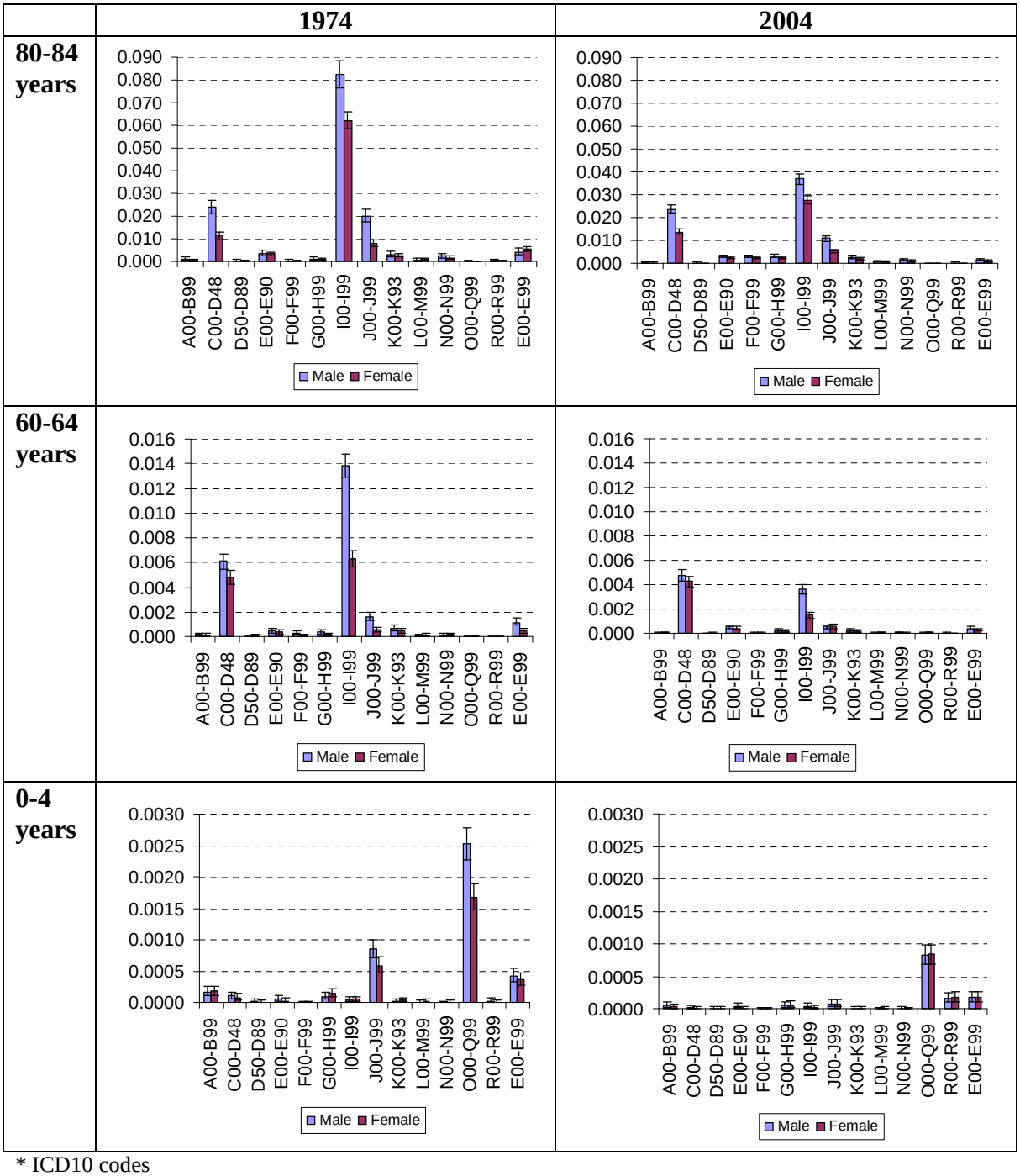
*Standardised to 1991 New Zealand total population age structure, separately for males and females.

3.2.3 Age-Specific Mortality Rates by Cause, 1974-2006

In addition to all-cause mortality, mortality rates by broad diagnostic group are considered. The graphs in Figure 3.2.5 show mortality rates for selected age groups 80 to 84, 60 to 64, and 0 to 4 in 1974 and 2004, by 14 diagnostic groupings[10]. Similar trends were found for adjacent age groups, for instance the mortality patterns of people aged 55 to 59 and 65 to 69 did not differ greatly from those aged 60 to 64. The mortality patterns were very consistent across age groups and from year to year, negating the need to present cause specific mortality rates for a wider range of ages. However, the age group 0 to 4, while constituting only a very small number of deaths, has different cause-specific patterns to other age groups, hence being presented here.

[10] While it is not technically correct to present rates in a bar chart format because rates are a point estimate, this is nonetheless common practice. Barcharts are used here because they are both aesthetically pleasing and intuitive to interpret.

Figure 3.2.5: Selected Age-Specific Mortality Rates by Cause*, 1974 and 2004, 95% Confidence Intervals



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality

The diagnosis groups are denoted by ICD10 codes (see Appendix A for details on coding). Note that there are two groups of codes starting with the letter e: E00-E9**0** denote endocrine, nutritional and metabolic disorders whereas E00-E9**9** denote external causes of morbidity and mortality. Because of the small numbers of deaths for some ICD group, age and sex combinations, 95 percent confidence intervals have been applied, using Poisson confidence intervals where the number of events is less than 100, and Normal confidence intervals where the number of events is 100 or greater. This process is detailed in Section 2.

Mortality rates for nearly every diagnostic group declined between 1974 and 2004, with the exception of symptoms, signs and abnormal clinical laboratory findings not elsewhere classified (R00-R99). The mortality rate for this cause at ages 0 to 4 increased between 1974 and 2004, however this could in part be due to changes in coding between the ICD versions, or due to changes or improvements in diagnosis. Circulatory conditions such as heart attacks and strokes (I00-I99) were the main cause of death in both 1974 and 2004, followed by neoplasms, a diagnostic group that includes cancers. However circulatory mortality rates declined dramatically over the period, for both males and females. For children aged 0 to 4, complications of childbirth, conditions originating in the perinatal period and congenital abnormalities (O00-Q99) were the main cause of death in 1974 and 2004. Respiratory diseases (J00-J99), which include pneumonia and respiratory infections but *not* infectious diseases such as tuberculosis were the second most common cause of death at ages 0 to 4 in 1974, but not in 2004. For ages 50+, men were significantly more likely to die from circulatory causes than women in 1974, but for many age groups there was no significant difference in 2004. The large declines in circulatory mortality rates (particularly for men) and the convergence in circulatory mortality rates between males and females could be due to behavioural factors such as smoking patterns, or medical factors such as improved treatment. Figure 3.2.5 matches the findings in sections 3.2.1 and 3.2.3 of a convergence of male and female mortality, particularly at older ages.

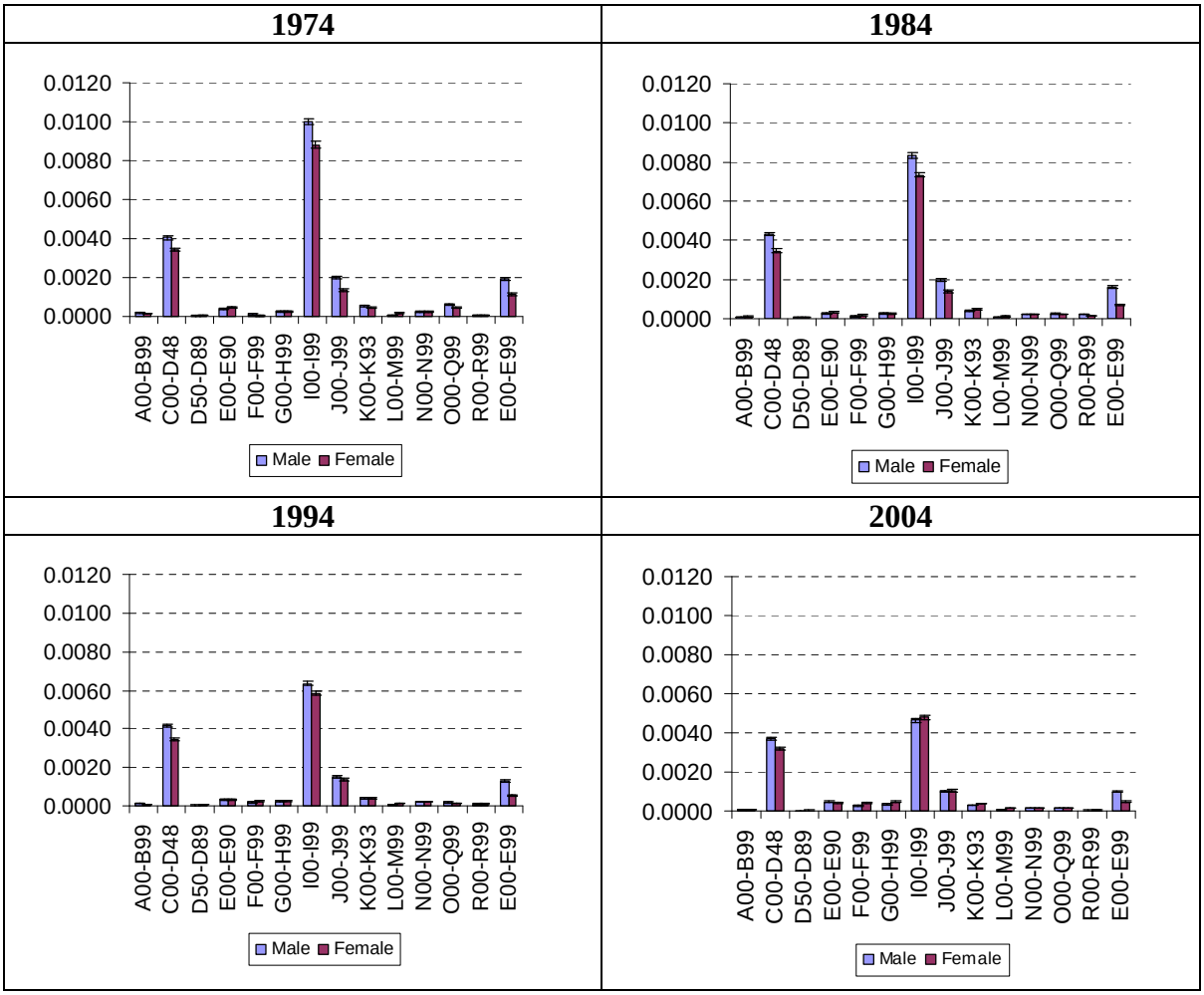
3.2.4 Standardised Mortality Rates by Cause, 1974-2006

In order to gain more than a snapshot of mortality patterns for limited age groups in 1974 and 2004, the rates are standardised by age to allow for direct comparison between cause-specific mortality rates for various years. Male and female rates are separately standardised to the total (both sexes) New Zealand population in 1991. See Appendix Table 1 for the standard population structure.

Figure 3.2.6 presents the age and sex standardised mortality rates for 1974, 1984, 1994 and 2004. It is apparent from this figure that when using age and sex standardised mortality rates, the sex differences at the start of the period are less pronounced and the cause of death pattern is much more stable between different years. Much of the large variation in cause-specific mortality seen in age-specific tables is therefore due to changes in the age and sex structure of the population over time. A decline in most causes of death was seen over the time period, with men generally having higher mortality rates than women. In 2004, females had a slightly higher mortality rate due to circulatory causes (I00-I99) than men, although the difference is not significant at the 95 percent level of significance. Circulatory causes dominate as the most common cause of death across the time period, followed by neoplasms (C00-D48), respiratory (J00-J99) and external causes (E00-E99). Deaths due to mental and behavioural disorders (F00-F99) increase over the period, increasing 10 fold for females. Increases of smaller magnitude are seen for several other minor causes of death. While there are large decreases in all-cause death rates, these are primarily driven by decreases in deaths due to circulatory conditions. Mortality rates for the second most common cause of death, neoplasms, show almost no decline, and mortality rates for several minor causes of death (particularly mental and behavioural) show increases over the time period, although remaining very uncommon causes of death.

There appears to be little change in the distribution of cause of death over the time period. Although circulatory mortality rates decline by nearly half, there are still the same peaks and troughs in the distribution of mortality rates across the diagnostic groups. The change between 1974 and 2004 is therefore not in the distribution of the causes of death, but of reduced scale.

Figure 3.2.6: Standardised* Mortality Rates by Cause**, Selected Years, 95% Confidence Intervals



* Standardised to 1991 New Zealand age structure

** ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)

3.2.5 Conclusions from National Mortality Analysis, 1974-2006

The main trends to emerge from the national analysis of mortality rates over the time period 1974-2006 are:

1. Mortality rates at all ages declined substantially over the time period.

2. There was considerable convergence between male and female mortality rates, particularly at older ages, due to a large decline in male mortality and a smaller decline in female mortality.
3. The four main causes of death across the time period were circulatory, neoplasms (cancers), respiratory and external causes.
4. Circulatory conditions are the leading cause of death, and declines in circulatory mortality drove both the overall mortality decline and the sex convergence.
5. Neoplasms (cancers) are the second most common cause of death, but neoplasm death rates did not see a decline between 1974 and 2006.
6. Variance of cause of death did not appear to increase over time.

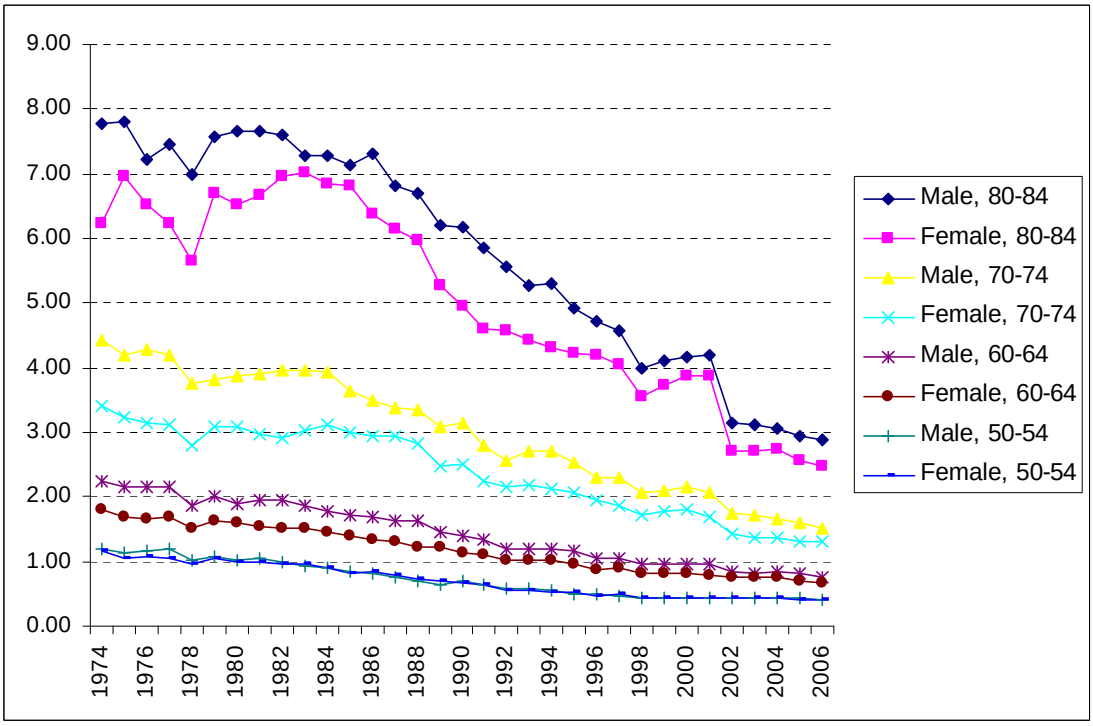
3.3 National Hospital Bed Day Rates (1974-2006)

Analyses of national mortality rates over time, by cause, by sex have highlighted striking differences in sex specific mortality. This section repeats the national mortality analyses, but for hospital bed day rates instead of mortality rates.

3.3.1 Age-Specific Hospital Bed Day Rates, 1974-2006

As with mortality rates, hospital bed day rates have steadily decreased over time (see Figure 3.3.1). Hospital use increases with age, and decreases over time. In 1974, males aged 80 to 84 spent on average nearly 8 days a year in hospital. By 2006 this had dropped to just below 3, a decrease of more than 50 percent. Males have consistently higher hospital use than females at ages below 20 and above 50 over the time period (bear in mind that obstetrics are filtered out). We can see from Figure 3.3.1 that there is no difference in hospital use between males and females at age 50-54, and that the sex gap in bed day rates increases with age. In the late 1970s there was approximately one day difference in hospital use between males and females aged 80-84, and this gap decreased over time to just half a day in 2006. As with the mortality rates, there are fluctuations in the rates at ages 70+ that are *not* due to random variation arising from a small number of observations. The sharp drop in bed day rates at ages 70+ in 2002 corresponds with the change from ICD version 9CM to ICD version 10. It is therefore likely to be artefactual, possibly arising from increased accuracy of filtering due to the wider range of diagnoses available in ICD10.

Figure 3.3.1: Age-Specific Hospital Bed Day Rates, Selected Ages, 1974-2006



Figures 3.3.2 and 3.3.3 illustrate just how much hospital use has declined, not just at ages 50+, but at all ages. Hospital use for males aged 85+ declined from over 12 days a year to less than 4, a decline of two thirds. Female rates also declined, but (as with mortality rates) they were not as high to start with. Convergence of male and female rates was observed over the time period. Declines in hospital bed day rates were considerable at ages 50 and over.

Figure 3.3.2: Age Specific Hospital Bed Day Rates, 1974

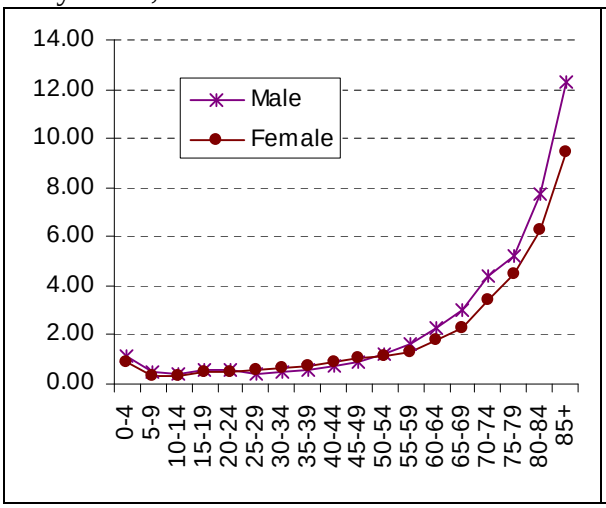
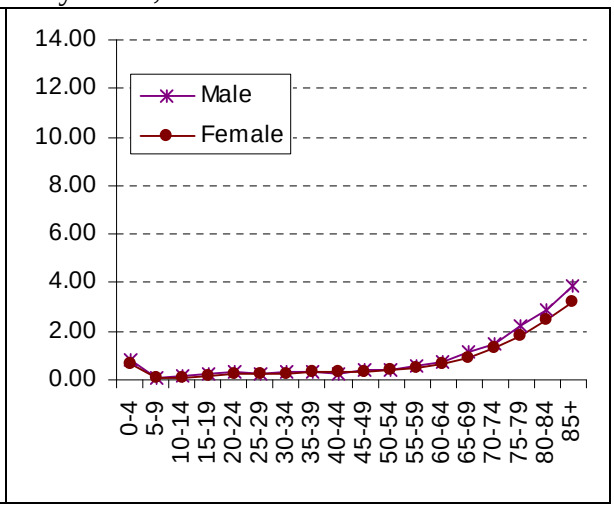


Figure 3.3.3: Age Specific Hospital Bed Day Rates, 2006



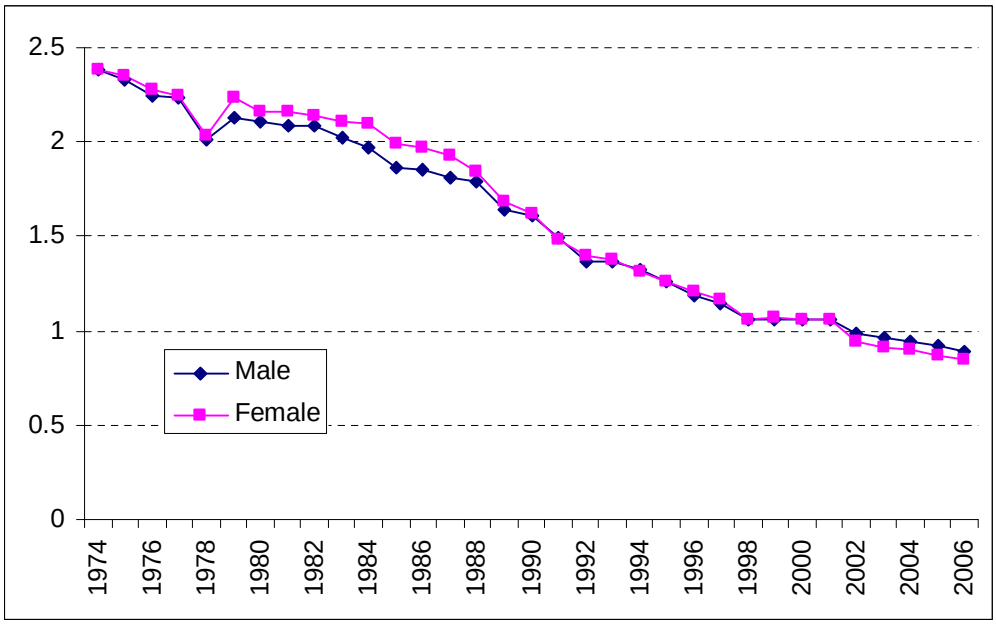
3.3.2 Standardised Hospital Bed Day Rates, 1974-2006

As with mortality rates, hospital bed day rates are affected by the population age and sex structure. Although pregnancy and childbirth have been filtered out, men and women go into hospital for different reasons, as do people of different ages. For this reason, to look at hospital bed day rates for all ages, standardisation is required to allow rates to be comparable over time.

The crude hospital bed day rates for males and females dropped from 1.1 to 0.5 days a year between 1974 and 2006. In comparison the standardised rates dropped from 2.4 to 0.8 days per year for males and females (see Figure 3.3.4). That is, on average every person in the population would have spent about 0.8 days in hospital in 2006, *if* the population age structure of New Zealand in 1991 applied. When looking at the age-specific rates, males had higher hospital use at most ages. However when standardised, we can see that overall, female hospital use was slightly higher than for men until about 2002. This is in contrast to the higher male than female standardised mortality rates seen in Figure 3.2.4. Other research has found male age standardised hospital bed day rates in New Zealand to be consistently higher than female rates (Pool, Baxendine *et al.* 2009). However, this study uses a slightly different filtering process, which on examination accounts for the different findings (see Appendix Figures 1 and 2).

The dip in hospital bed day rates in 1978 is also seen in figure 3.2.1, and has yet to be explained. Another decline occurs in 2002, and this is prolonged to the end of the period. This coincides with the change in ICD version from ICD9 to ICD10. ICD10 was very different to the preceding ICD versions, and this drop in hospital bed day rates is most likely an artefact of this change, because of a higher proportion of hospital stays falling under the criteria for filtering.

Figure 3.3.4: Standardised* Hospital Bed Day Rates, 1974-2006



*Standardised to 1991 New Zealand total population age structure, separately for males and females.

3.3.3 Age-Specific Hospital Bed Day Rates by Cause, 1974-2006

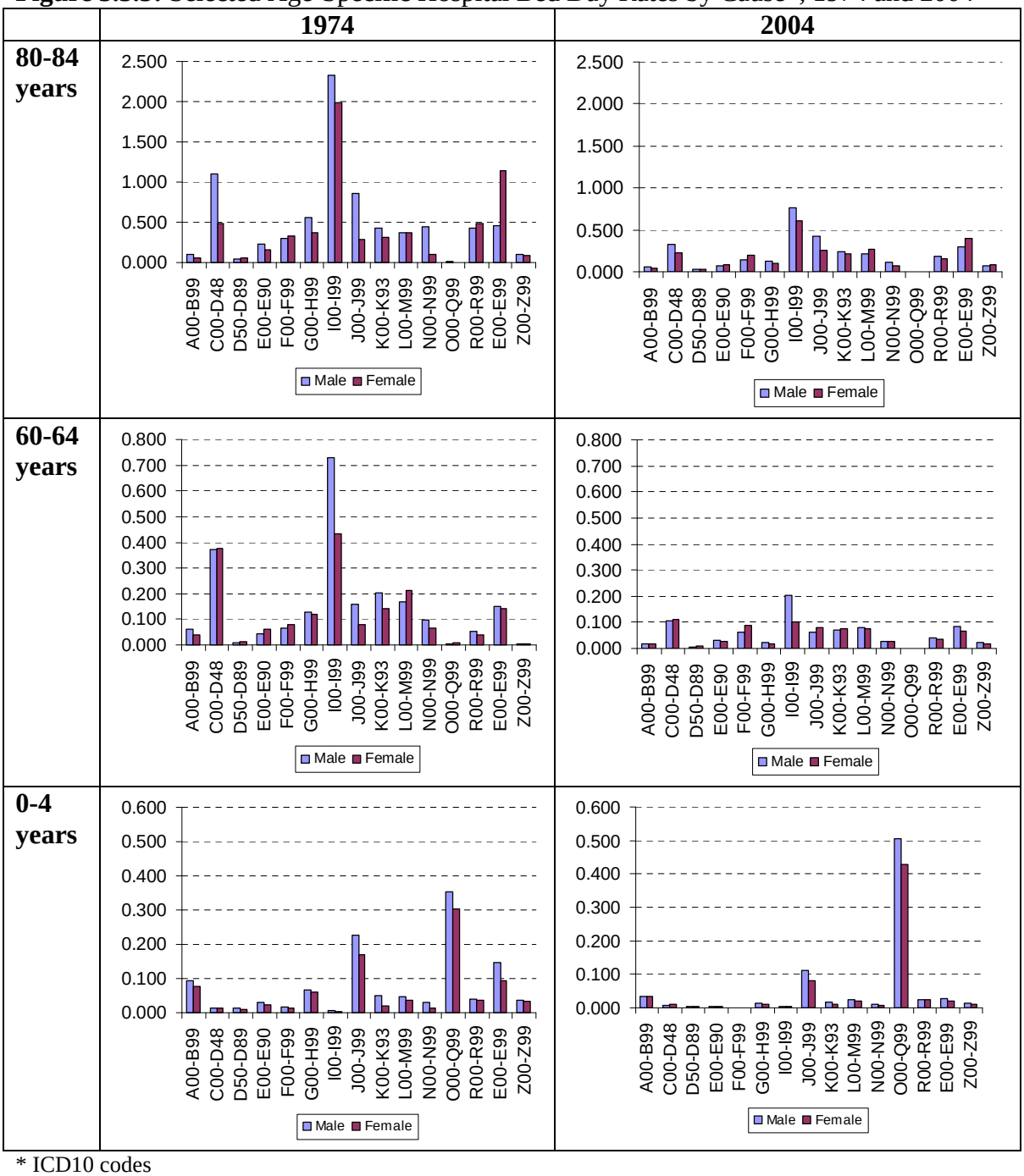
From the cause-specific mortality analyses, it was apparent that by far the main cause of death was circulatory, followed by neoplasms, respiratory, and external causes. In examining the hospital bed day rates it is interesting to compare the main causes of hospital use to the main causes of death. Do people go into hospital for the same things they die from?

Figure 3.3.5 presents hospital bed day rates by broad diagnostic group, for selected ages in 1974 and 2004. Due to the large number of hospital bed days compared to deaths, confidence intervals for these results were very small and so have not been included. As with the cause specific mortality rates, the hospital bed day rates are very stable across age groups and years, with only gradual changes occurring from year to year and between age groups. The first thing to notice is that, while circulatory conditions (I00-I99) denoting conditions such as strokes and heart disease are the most common cause of hospital use, people go into hospital for lots of other reasons as well. The bars are more spread out across the diagnostic groups than for mortality. Also, at ages 0 to 4, hospital use *increased* over the period, while mortality at ages 0 to 4 declined. This suggests that medical interventions and hospital use at ages 0 to 4 are helping to reduce mortality. Hospital use in this age group is dominated by Q00-Q99, certain conditions originating in the perinatal period and congenital abnormalities (O00-O99, obstetrics, are filtered out) followed by respiratory diseases (J00-J99), such as respiratory infections and pneumonia. Certain

infectious and parasitic (A00-B99) and external causes (E00-E99) were also prominent causes of hospital use for children ages 0 to 4 in 1974, but had declined by 2004.

The huge drop in hospital use at older ages is apparent across nearly all diagnostic groups, accompanied by an increase in the diversity of cause of hospital use. The large peaks of circulatory, neoplasms (cancers) and external causes for ages 80 to 84 in 1974 all but disappear by 2004. It is clear to see the importance of circulatory conditions as the leading cause of both mortality and hospital use. Males aged 80 to 84 in 1974 could expect to spend on average over 2 days in hospital a year due to circulatory conditions alone. External causes, such as accidents were a major cause of hospital use among women aged 80 to 84 in 1974, on average contributing one bed day a year.

Figure 3.3.5: Selected Age-Specific Hospital Bed Day Rates by Cause*, 1974 and 2004



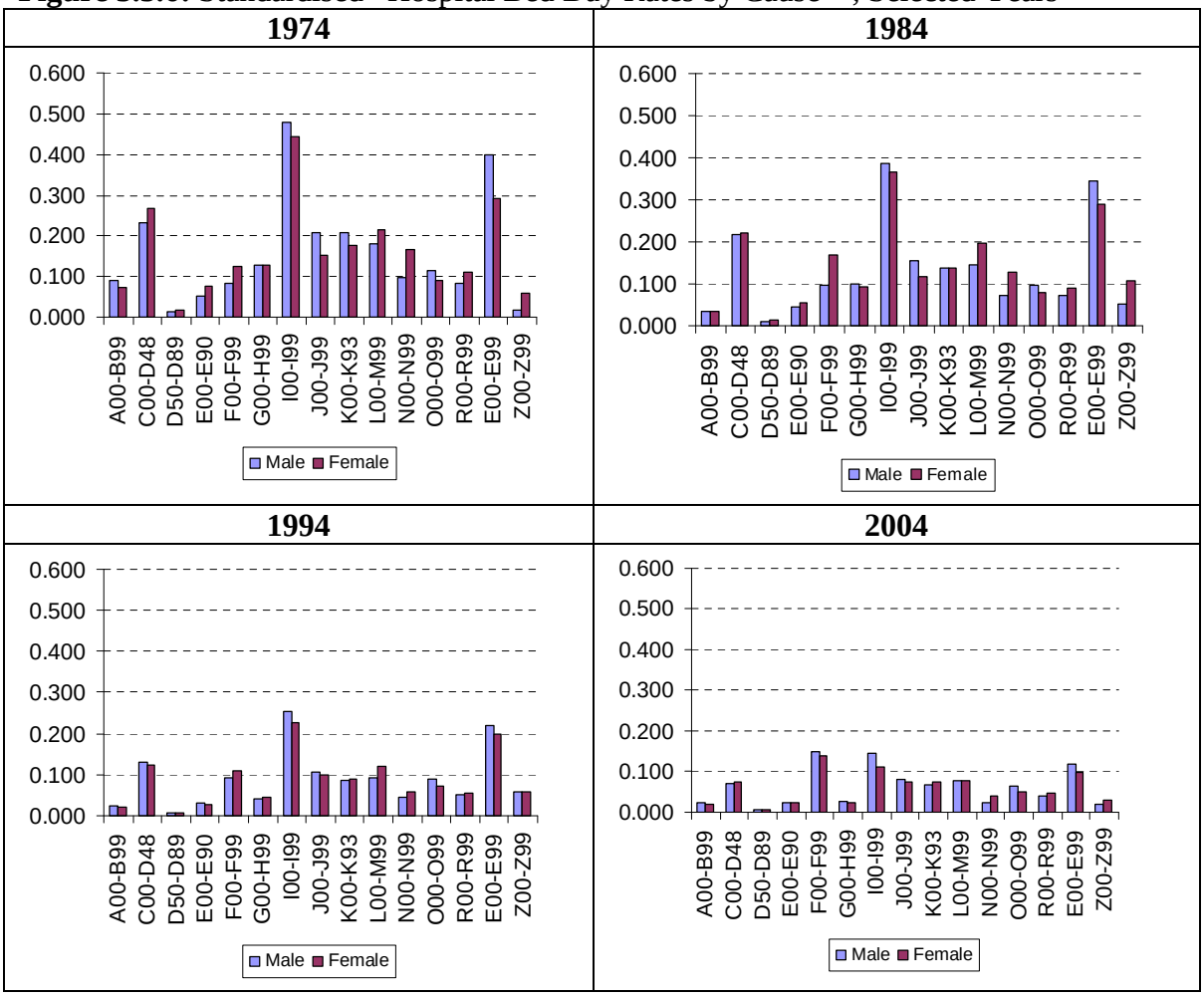
* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

3.3.4 Standardised Hospital Bed Day Rates by Cause, 1974-2006

Standardised hospital bed day rates show again a decline in the peaks and troughs of the cause of hospital use distribution, with hospital use being distributed much more evenly across diagnostic groups in 2004 than in earlier years (see Figure 3.3.6).

Figure 3.3.6: Standardised* Hospital Bed Day Rates by Cause**, Selected Years



*Standardised to 1991 New Zealand total population age structure, separately for males and females.

** ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

In 2004, circulatory conditions had ceased to be the by far the largest contributor, replaced by mental and behavioural disorders (F00-F99) which contributed nearly one bed day a year per individual in 2004. However no such trend was apparent in the age-standardised mortality analysis, or in the age-specific mortality or hospital bed day analyses, which focussed on ages 0 to 4, 60 to 64 and 80 to 84. The rise in hospital bed day rates due to mental and behavioural disorders is driven by young and middle aged adults, and is not reflected in mortality rates.

One surprise from the standardised rates is the importance of external causes in contributing to hospital use. In 1974 external causes alone contributed nearly 2 bed days a year for men and nearly 1.5 bed days a year for women, and in 2004 external causes were the third most common cause of hospital use behind circulatory and mental and behavioural disorders, contributing well over half a day in hospital for every individual per year.

3.3.5 Conclusions from Hospital Bed Day Rate Analyses, 1974-2006

The main trends to emerge from the analyses of hospital bed day rates between 1991 and 2006 are:

1. Hospital bed day rates declined over the time period, except at ages 0 to 4, where hospital bed day rates increased.
2. Males had higher age-specific bed day rates at ages below 20 years and over 50 years, but females had slightly higher age-standardised bed day rates for most of the time period (excluding 2002 onwards).
3. Male and female hospital bed day rates converged over time
4. Cause of hospital use changed considerably over the time period. The three main causes of hospital use in 1974 were (in order of importance) circulatory, external causes and neoplasms. By 2004 this had changed to mental health, circulatory, and external causes.
5. Diversity of cause of hospital use increased substantially over time.

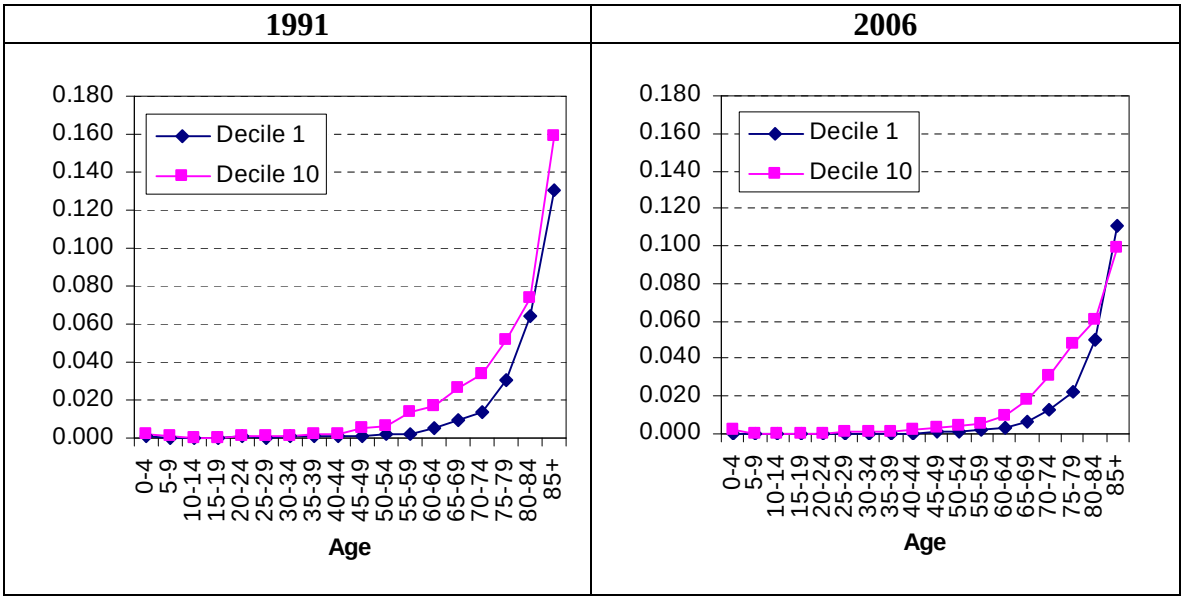
3.4 Mortality Rates by Area Deprivation (1991-2006)

Having examined trends in mortality rates and hospital bed day rates at a national level over a long time period, trends in mortality rates by NZ*dep* area deprivation deciles will now be considered.

3.4.1 Age-Specific Mortality Rates by Area Deprivation, 1991-2006

As with the national analysis, this section starts by looking at age-specific mortality rates. Figure 3.4.1 presents female mortality rates for the highest and the lowest deciles in 1991 and 2006. As expected, mortality rates declined over the period, with female mortality at ages 85+ dropping from 160 to 100 per thousand between 1991 and 2006 for deprivation decile 10. In 1991, mortality rates for decile 1 (the area with the lowest NZ*dep* score, i.e. least deprived) only start to increase from about ages 60 to 64, whereas mortality rates in decile 10 (the areas with the highest NZ*dep* score) start to increase from about ages 45 to 49. Mortality rates for deciles 1 and 10 converge at about ages 80 to 84. There appears to be relatively little difference in mortality rates between deciles 1 and 10 at ages 85+, especially in 2006. The difference in mortality rates between deciles 1 and 10 is primarily between ages 50 and 79.

Figure 3.4.1: Female Age-Specific Mortality Rates, Deprivation Deciles 1 and 10, Years 1991 and 2006



Male mortality rates follow a similar trend, but with slightly less convergence at ages 85+. Deciles 2 and 9 had similar patterns to deciles 1 and 10, but with more variation.

This may be due to relatively small numbers of deaths in some age group, decile and year combinations.

3.4.2 Standardised Mortality Rates by Area Deprivation, 1991-2006

Areas with high and low deprivation scores have markedly different age and sex structures. While age-specific trends are useful, they do not tell us about overall mortality trends for all ages combined. It is therefore vital to standardise the mortality rates to be able to compare trends between deprivation deciles, for males and females regardless of age structure.

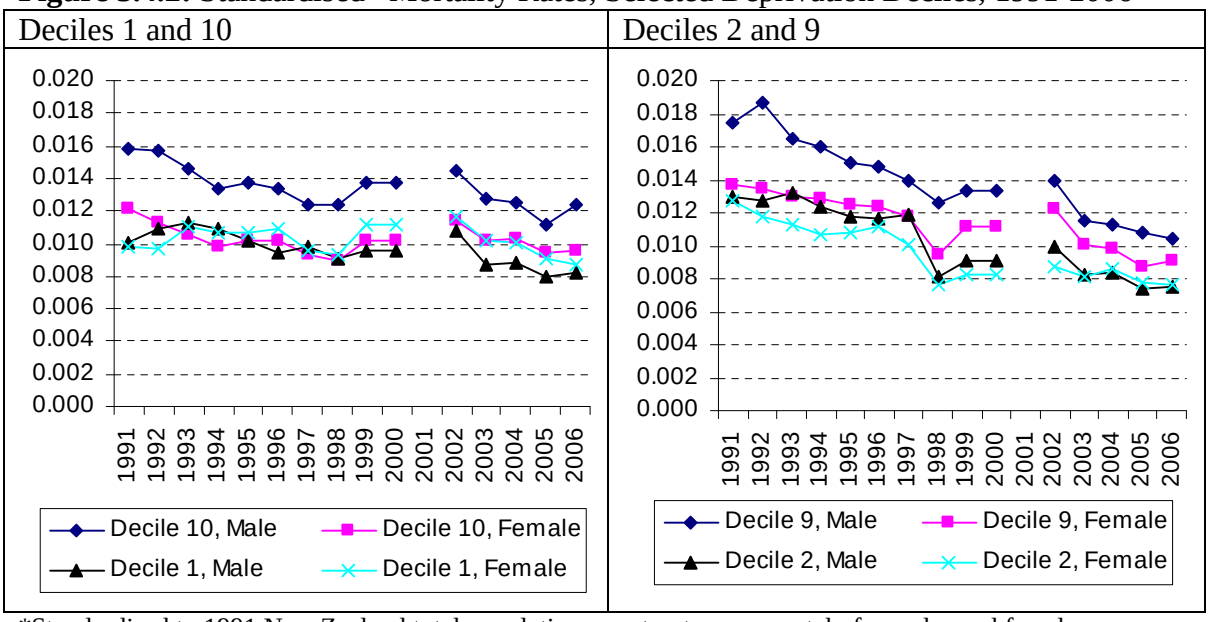
After standardisation it is apparent that there is a clear decline in mortality rates over time across all deprivation deciles, which was not so apparent from the age-specific figures. Figure 3.4.2 illustrates both this decline, and the difference in standardised mortality rates between deciles 1 and 10, and between deciles 2 and 9. The first point of interest is the difference in mortality rates between deciles 1 and 10, and between deciles 2 and 9. This difference is far greater for males than females, which suggests that area deprivation is more closely related to mortality rates for males than females: female mortality does not differ so much by level of deprivation.

The second and main point of interest is the sex specific differences in mortality rates *within* deciles. The sex difference in mortality rates for decile 10 (the most deprived decile) is approximately 2 per thousand: males in decile 10 had mortality rates approximately 2 per thousand higher than females. The same trend is apparent for decile 9, although the gap between male and female mortality rates is smaller. In contrast, females in decile 1 (the least deprived decile) have *higher* mortality rates than males in decile 1. This is seen in all years except 1992. This is a highly unexpected finding, which requires further investigation. While not replicated in other deciles, male and female mortality rates in deciles 2, 3 and 4 are very similar, and overlap, with female mortality being marginally higher than male mortality in some years in decile 2. It would appear that *either* women living in areas with low NZ*dep* scores experience abnormally high mortality (low deprivation = negative effect for women), *or* men living in areas with low NZ*dep* scores experience abnormally low mortality (low deprivation = protective effect for men). Other research has also found the top end of the socioeconomic distribution (the least deprived end) to behave abnormally. Banks *et al.* found the top end of the socioeconomic distribution to have abnormally high levels of self-reported health and chronic diseases (Banks, Marmot *et al.* 2006), but they did not find a difference by sex. This suggests that

the least deprived decile may have atypical results that do not match the gradient in mortality seen for other deciles.

There is a lot of overlapping of mortality rates by deprivation decile. There is no perfect gradient with each decile experiencing sequentially higher (or lower) mortality rates. There does however appear to be a clear divide between the bottom third and the top third of deciles.

Figure 3.4.2: Standardised* Mortality Rates, Selected Deprivation Deciles, 1991-2006

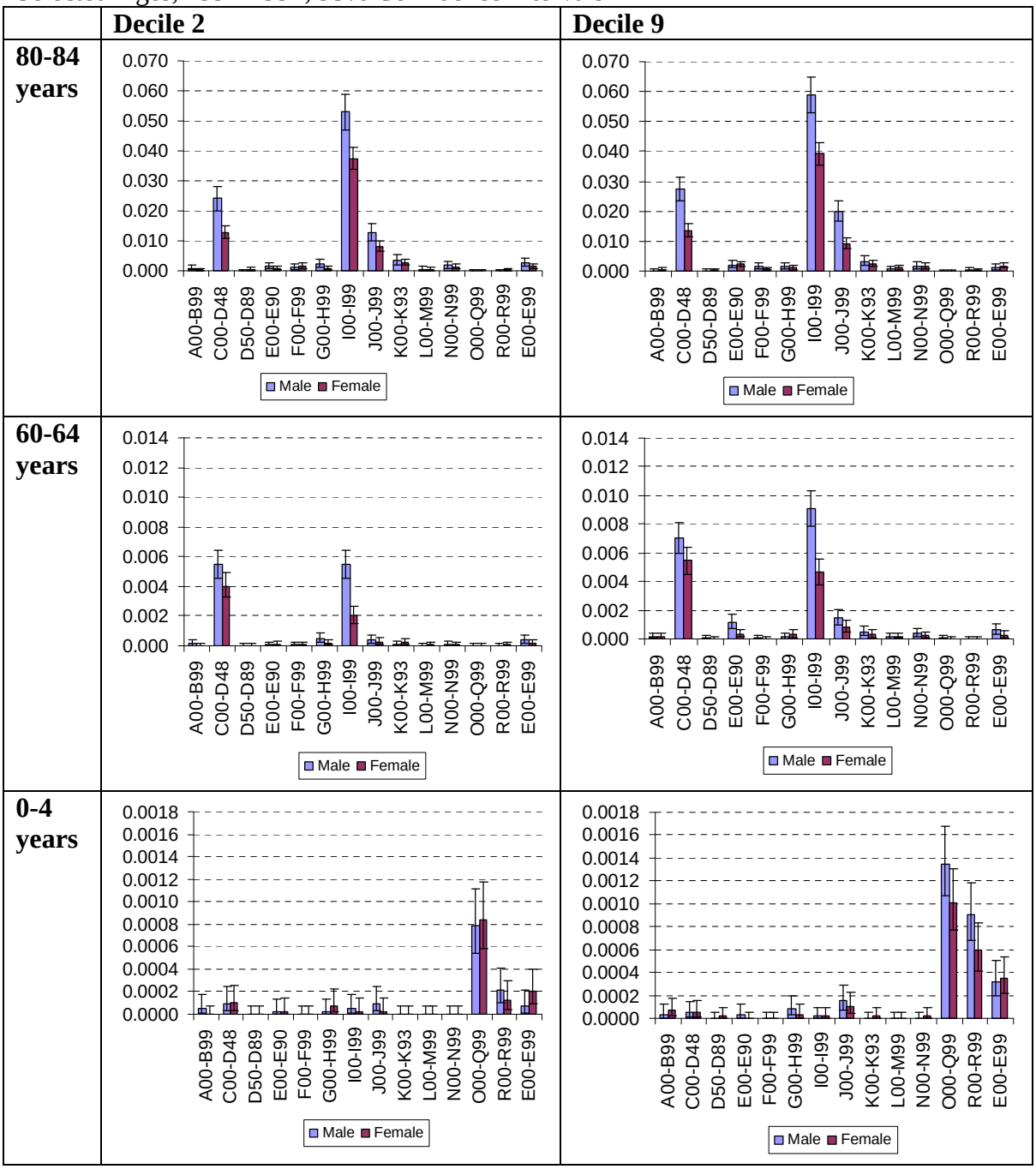


*Standardised to 1991 New Zealand total population age structure, separately for males and females.

3.4.3 Age-Specific Mortality Rates by Cause and Area Deprivation, 1991-2006

It is clear that very different patterns of mortality exist between NZ*dep* deciles in New Zealand. The age at which mortality rates begin to rise, and sex differentials in mortality rates vary by deprivation decile. Sections 3.4.3 and 3.4.4 explore the cause-specific trends in mortality by deprivation decile. There is evidence from figure 3.4.2 to suggest that deciles 1 and 10 may be abnormal, and follow different patterns than other deciles (for instance, the unusually high female mortality rate in decile 1), so for this analysis deciles 2 and 9 will be considered. Figures 3.4.3 and 3.4.4 show age-specific mortality rates by cause for years 1991 to 1994 and 2003 to 2006 respectively. Grouped years are used due to the small numbers of deaths in some age, sex, decile and diagnostic group combinations, and 95 percent confidence intervals are applied to highlight where differences in rates are significant.

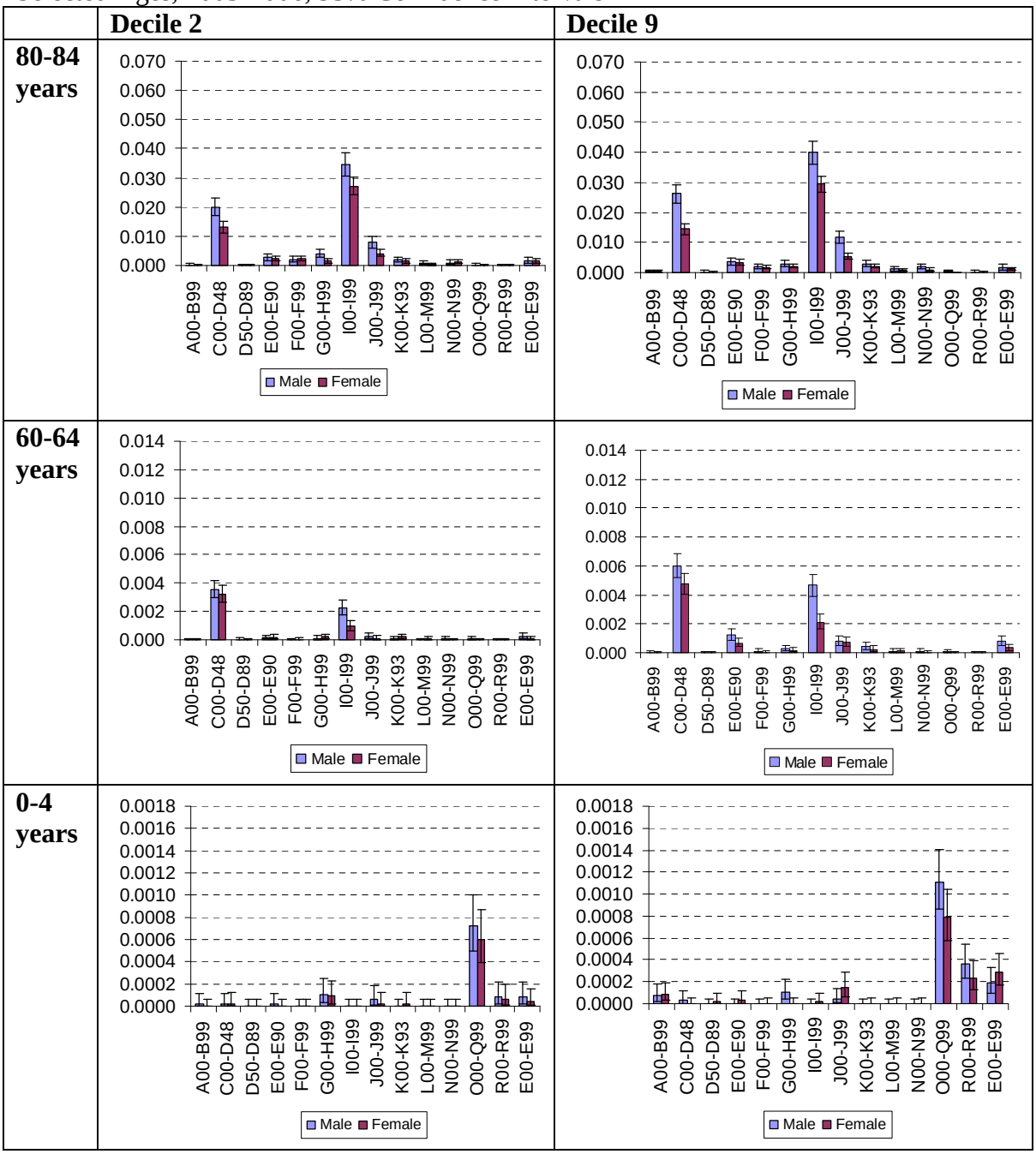
Figure 3.4.3: Age-Specific Mortality Rates by Cause*, Deprivation Deciles 2 and 9, Selected Ages, 1991-1994, 95% Confidence Intervals



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)

Figure 3.4.4: Age-Specific Mortality Rates by Cause*, Deprivation Deciles 2 and 9, Selected Ages, 2003-2006, 95% Confidence Intervals



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)

The first striking thing about Figures 3.4.3 and 3.4.4 are the large confidence intervals, particularly for mortality rates of children aged 0 to 4. From 1991 to 1994 there were only 2,034 deaths to children of this age. Two sex categories, 10 area deprivation deciles and 14 diagnostic groups gives 280 different combinations, resulting in very small numbers and large confidence intervals. While child mortality appears higher in decile 9 than in decile 2, a small number of deaths mean it is impossible to prove statistically for most causes of death.

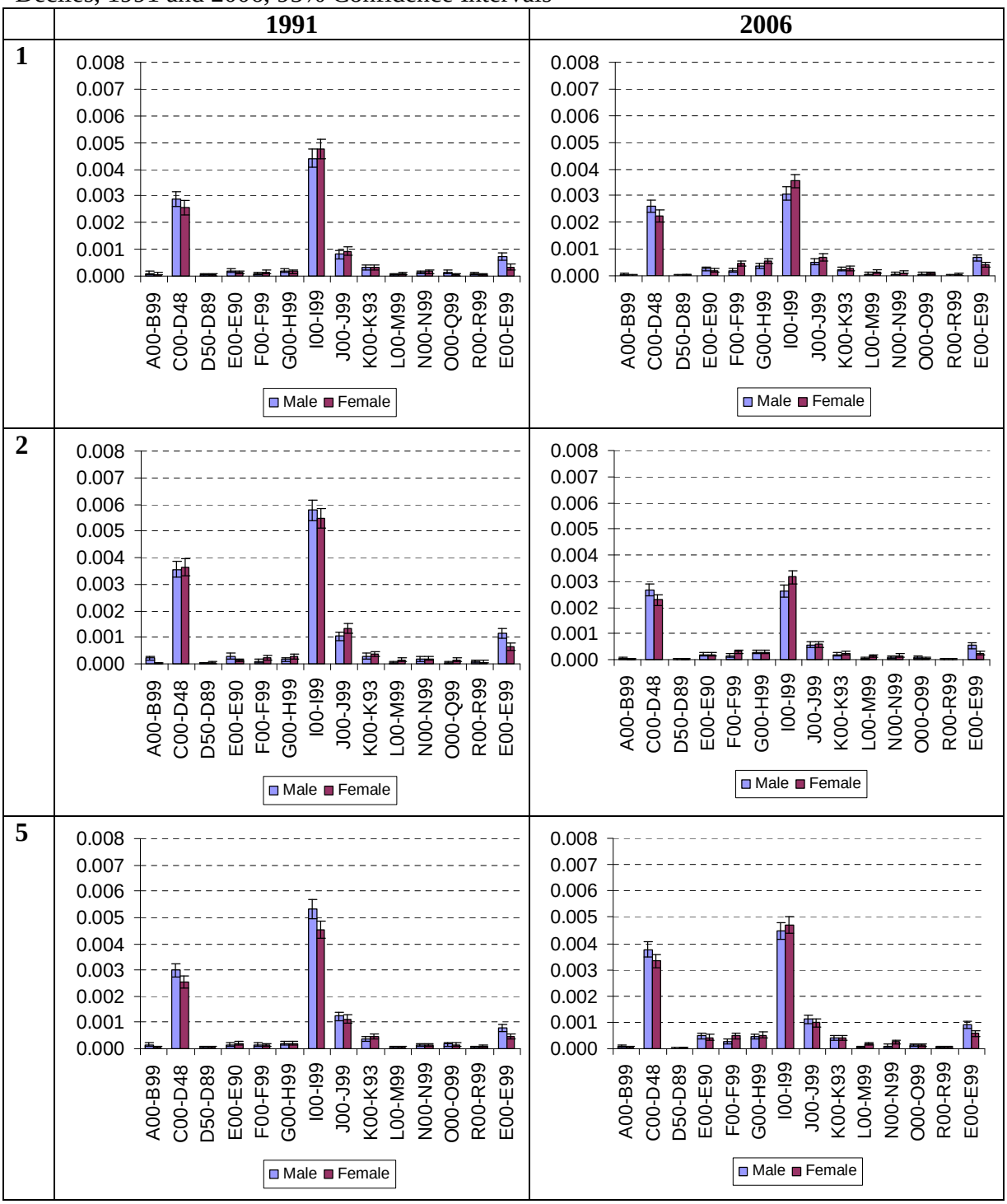
The second point of interest is that while the cause-specific pattern of mortality rates looks fairly similar for deciles 2 and 9 at both time periods, there are differences by sex. There tends to be a larger gap between male and female mortality rates for decile 9 than for decile 2, with men experiencing much higher mortality rates. This is true across the age groups and also for other high deciles (8 and 10), and despite small numbers, it is often significantly so. It would appear that men in areas with high deprivation scores experience disproportionately high mortality compared to women, particularly for circulatory and neoplasm causes.

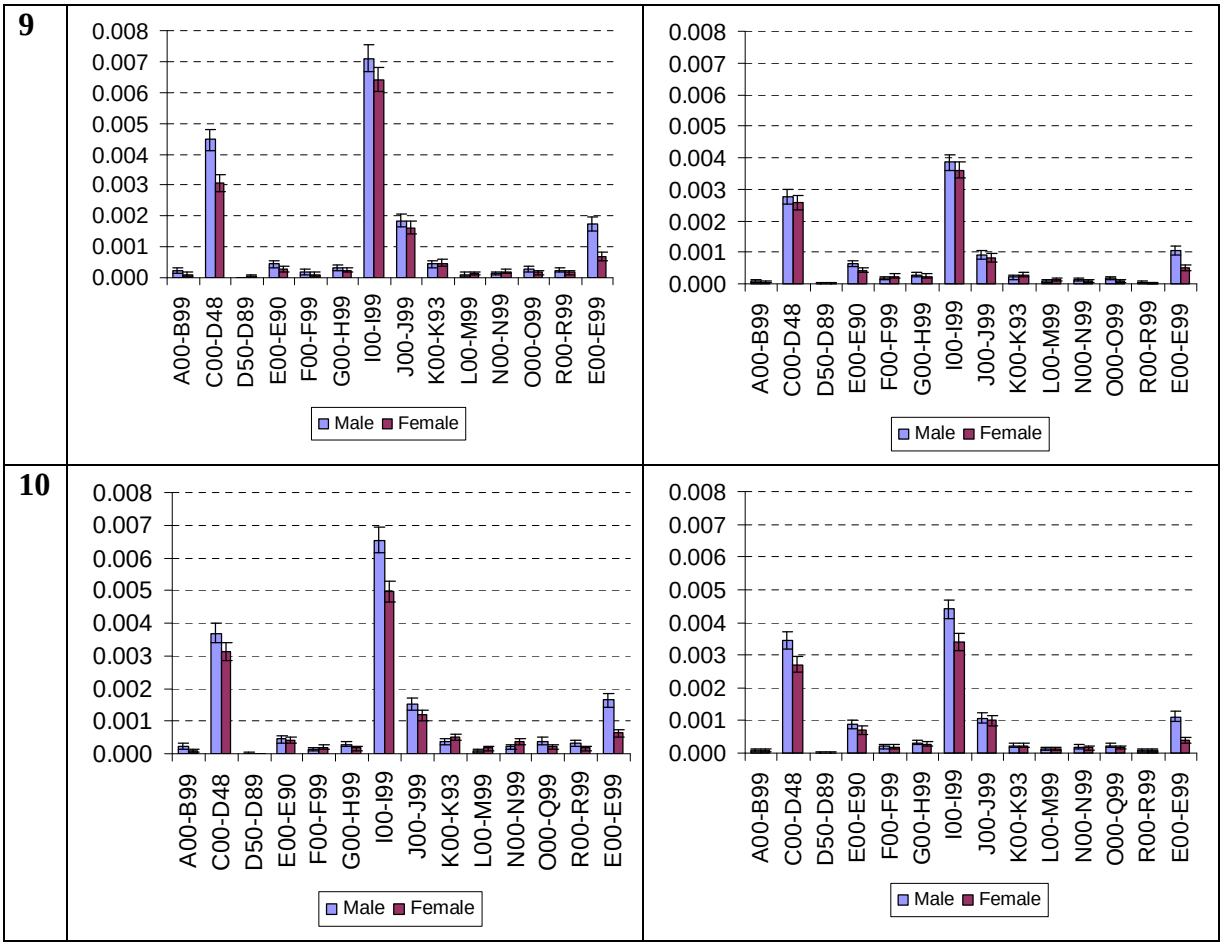
A third thing to note is the relatively small differences in mortality rates between deciles 2 and 9 at ages 80 to 84, with (as previously noted) higher mortality from circulatory conditions for women in decile 2 compared to decile 9. This is also found when comparing deciles 1 and 10. It is at younger ages that differences in mortality rates by deprivation decile manifest (as shown in Figure 3.4.1), with people in areas with high deprivation scores experiencing higher mortality rates due to circulatory conditions, neoplasms, respiratory conditions and external causes, as seen by comparing cause-specific mortality rates at ages 60 to 64. This matches with other studies that found higher mortality due to accidents, respiratory diseases, circulatory diseases and cancer in lower socioeconomic groups in New Zealand (Blakely, Woodward *et al.* 2002; Paul 2002).

3.4.4 Standardised Mortality Rates by Cause and Area Deprivation, 1991-2006

Examining mortality trends by cause and deprivation has been problematic due to small numbers in some age, sex, deprivation and diagnostic group combinations. Standardising not only allows us to directly compare mortality rates between deprivation deciles, but also reduces the number of combinations through looking at all-age mortality. Figure 3.4.5 presents the standardised mortality rates for deprivation deciles 1, 2, 5, 9 and 10, in both 1991 and 2006. The deciles are down the left hand side in Figure 3.4.5, unlike in the previous two figures.

Figure 3.4.5: Standardised* Mortality Rates by Cause** and Selected Deprivation Deciles, 1991 and 2006, 95% Confidence Intervals





*Standardised to 1991 New Zealand total population age structure, separately for males and females.
** ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)

Using standardised rates it is possible to see the differences in cause-specific mortality rates between areas with the lowest and the highest deprivation scores. Mortality due to circulatory conditions for males was just under 5 per thousand in decile 1 in 1991, but over 6 per thousand in decile 10. It can also be seen that declines in mortality between 1991 and 2006 have not been limited to areas with low deprivation scores, mortality decline was seen in all deprivation deciles. A pattern of relatively high female mortality due to circulatory conditions is apparent in Figure 3.4.5. In deciles 1 and 2 in 2006, female circulatory mortality rates were significantly higher than male circulatory rates. The

decrease in the sex mortality differential appears to be being driven by excess female circulatory mortality.

3.4.5 Conclusions from Area Deprivation Mortality Analysis, 1991-2006

The main trends to emerge from the analyses of mortality rates by area deprivation score between 1991 and 2006 are:

1. More deprived areas have higher mortality rates than less deprived areas. The difference is greatest at ages 50-79, and narrows at ages 80+. There is relatively little difference in mortality rates by deprivation score at the oldest ages.
2. The onset of increasing mortality rates is approximately 10 years later in areas with low deprivation scores than in areas with high deprivation scores.
3. Mortality in the least deprived areas is higher for females than for males and this is primarily due to differences in circulatory mortality. Male mortality from circulatory conditions dropped far more than female mortality from circulatory conditions over the time period, resulting in female circulatory mortality rates being higher than male circulatory mortality rates in areas with low deprivation scores from 1994 onwards.
4. The cause-specific pattern of mortality rates remained very similar across deprivation deciles, although areas with high deprivation scores experienced higher respiratory and external cause mortality rates, and females in areas with low deprivation scores experienced higher mortality due to circulatory conditions.

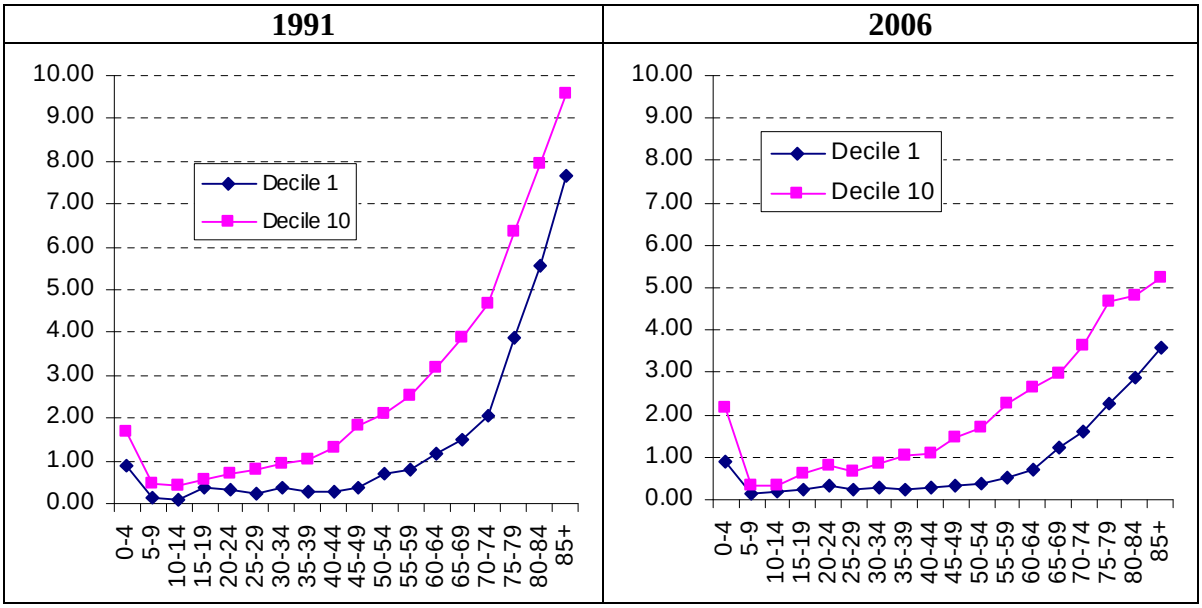
3.5 Hospital bed day rates by Area Deprivation (1991-2006)

The mortality analysis by area deprivation found a clear gradient in mortality rates by area deprivation, but how does this translate to hospital use? It must be stressed that the factors determining hospital use by area deprivation are not as clear cut as for mortality. A high bed day rate may represent higher morbidity (demand), or it may represent better provision (supply). Alternatively, better provision (supply) at the primary level, or increased substitution of healthcare from public to private hospitals in more affluent areas may result in lower public hospital use. These considerations must be taken into account when examining differences in hospital bed day rates by area deprivation.

3.5.1 Age-Specific Hospital Bed Day Rates by Area Deprivation, 1991-2006

The analysis starts by examining age-specific rates for deprivation deciles 1 and 10 in 1991 and 2006 (see Figure 3.5.1).

Figure 3.5.1: Female Age-Specific Hospital Bed Day Rates, Deprivation Deciles 1 and 10, Years 1991 and 2006



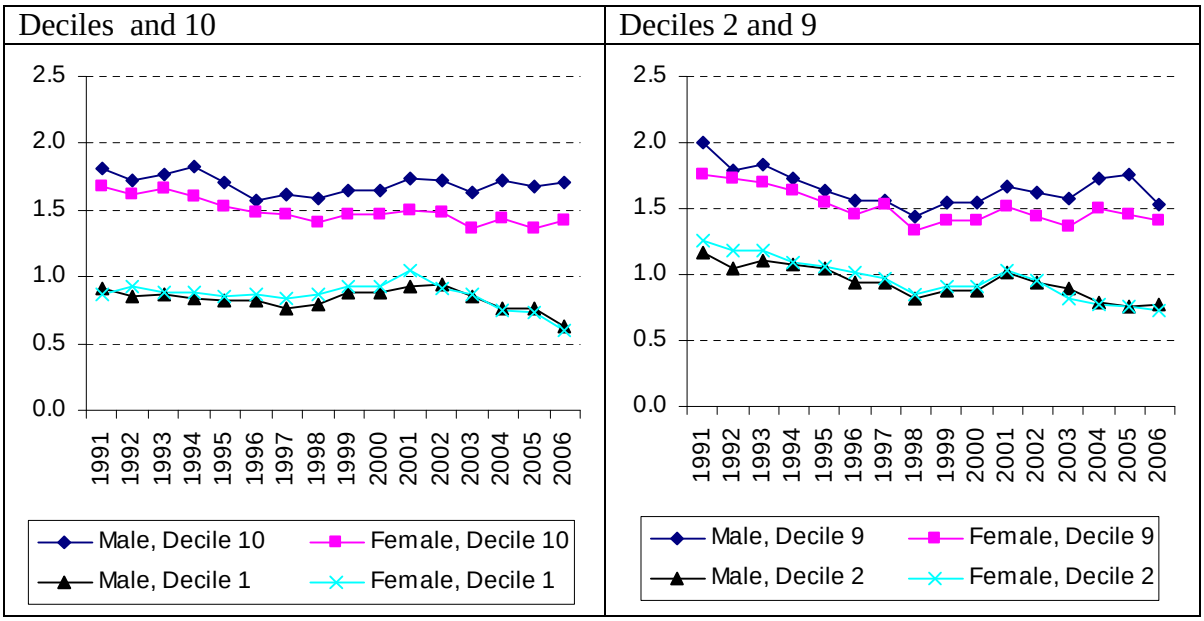
The most noticeable difference in age-specific bed day rates between 1991 and 2006 is the large drop in bed day rates at older ages (65 and above) for both deciles 1 and 10. Hospital bed day rates for those aged 85+ in decile 10 dropped from over 7 days to under 4 days over the period 1991 to 2006, and for decile 1 from just under 6 days to just under 3 days. However the gap in hospital bed day rates between deciles 1 and 10 did not decrease over the time period. The other point of interest is the kink at ages 0 to 4. Bed day rates at ages 0 to 4 did not decrease over the time period and remained about twice as high for decile 10 than decile 1. Decile 10 has higher hospital bed day rates than decile 1 in both 1991 and 2006, and the gap between deciles 1 and 10 increases with age (except in 1991, where it decreases slightly after age 70).

3.5.2 Standardised Hospital Bed Day Rates by Area Deprivation, 1991-2006

Standardised hospital bed day rates declined over the period 1991 to 2006, but the gaps between deciles 2 and 9 and between deciles 1 and 10 did not decrease, remaining at about 1 day (see Figure 3.5.2). The gap between deciles 1 and 10 however does appear to be slightly larger than the gap between deciles 2 and 9, as would be expected if each decile

had sequentially higher or lower hospital bed day rates. For decile 9, males had higher hospital bed day rates than females, but for decile 2, females had higher bed day rates than males. The same trend is apparent in deciles 1 and 10. The hospital bed day trends by decile correspond to the mortality patterns observed in Figure 3.4.2.

Figure 3.5.2: Standardised* Hospital Bed Day Rates, Deprivation Deciles 2 and 9, 1991-2006



*Standardised to 1991 New Zealand total population age structure, separately for males and females.

It is interesting to speculate on the reasons for this cross over of male and female rates. It could be due to women in deciles 1 and 2 being more health conscious than men and choosing to undergo elective preventative hospital episodes despite being quite healthy and thus not representing a real difference in morbidity. This however, does not correspond to the higher female than male mortality rates in these deciles. The mortality trends suggest that this is a real phenomenon: women in lower deciles have similar or worse health than men in those deciles. This is exaggerated by the fact that men normally experience higher mortality rates than women, at all ages. In deciles 9 and 10 men may have more dangerous occupations and risky lifestyles than women and therefore have higher bed day rates due to acute and emergency episodes, representing a real difference in morbidity.

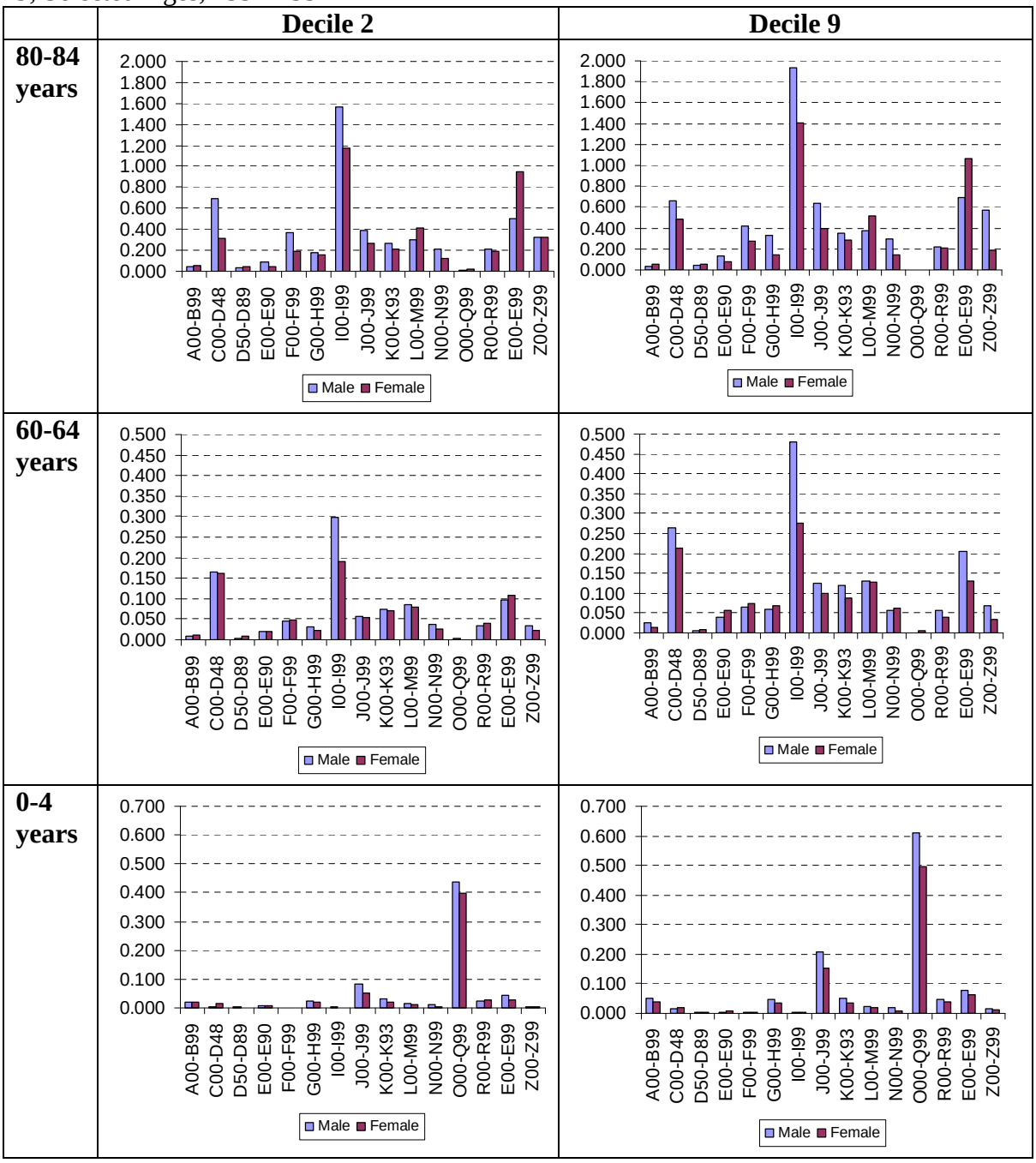
3.5.3 Age-Specific Hospital Bed Day Rates by Cause and Area Deprivation, 1991-2006

The previous two sections have determined that higher deciles have higher hospital bed day rates, that these rates declined by up to 50 percent for ages 65 and over during the

period 1991 to 2006, and that the difference in bed day rates between high and low deciles has not decreased over time. This section explores if cause-specific hospital bed day rates vary between deciles 2 and 9, and how this has changed over time (see Figures 3.5.3 and 3.5.4).

At first glance, the cause-specific pattern of hospital use is very similar for deciles 2 and 9, but closer inspection reveals several variations. There is a disproportionately high hospital bed day rate for respiratory (J00-J99) at ages 0 to 4 in decile 9, in both 1991-94 and 2003-06. At ages 60 to 64, circulatory (I00-I99), external causes (E00-E99) and neoplasms (C00-D48) are disproportionately high for men in decile 9, but not for women. The patterns in hospital bed day rates by cause do not appear to have changed over the time period.

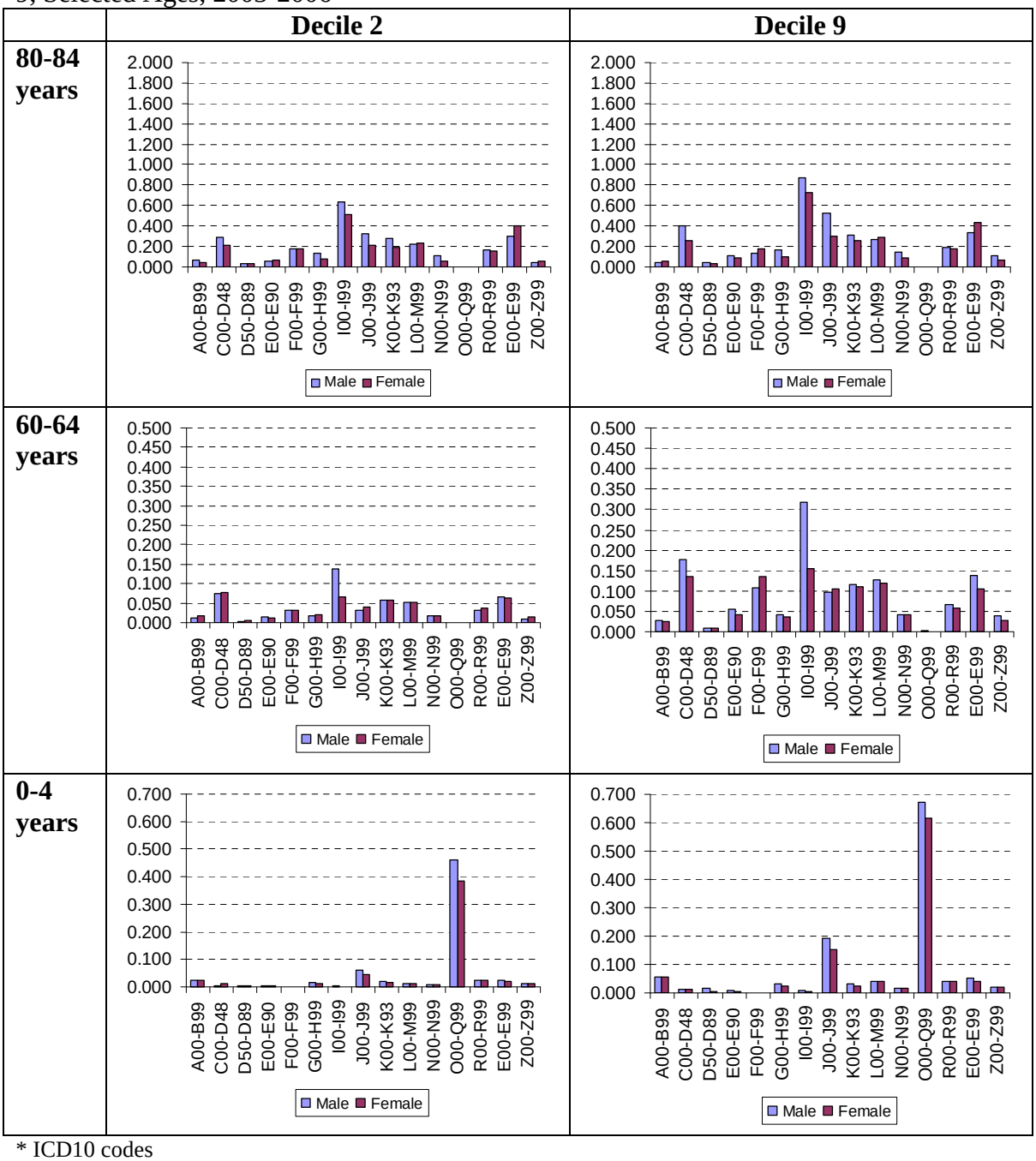
Figure 3.5.3: Age-Specific Hospital Bed Day Rates by Cause*, Deprivation Deciles 2 and 9, Selected Ages, 1991-1994



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

Figure 3.5.4: Age-Specific Hospital Bed Day Rates by Cause*, Deprivation Deciles 2 and 9, Selected Ages, 2003-2006



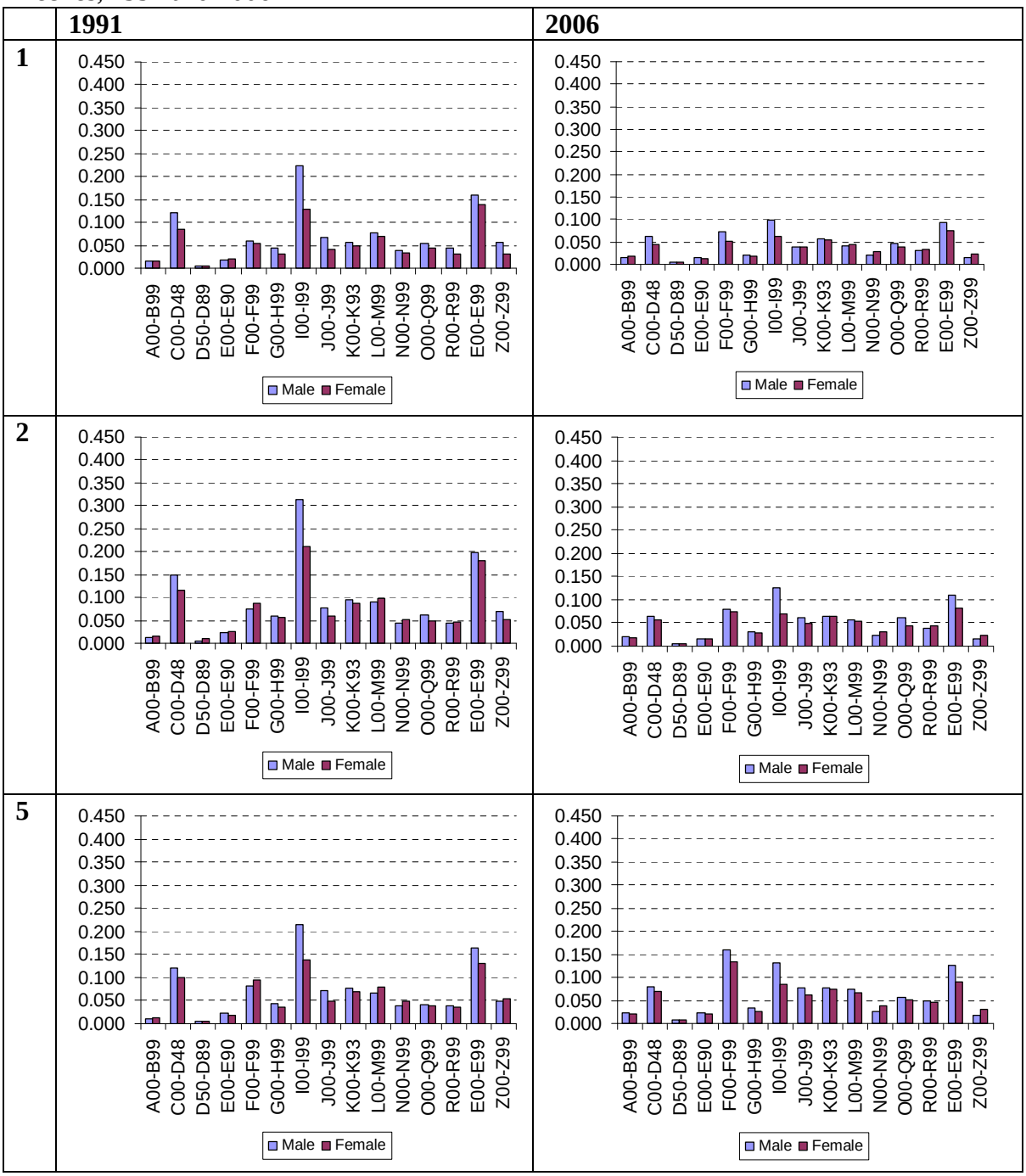
* ICD10 codes

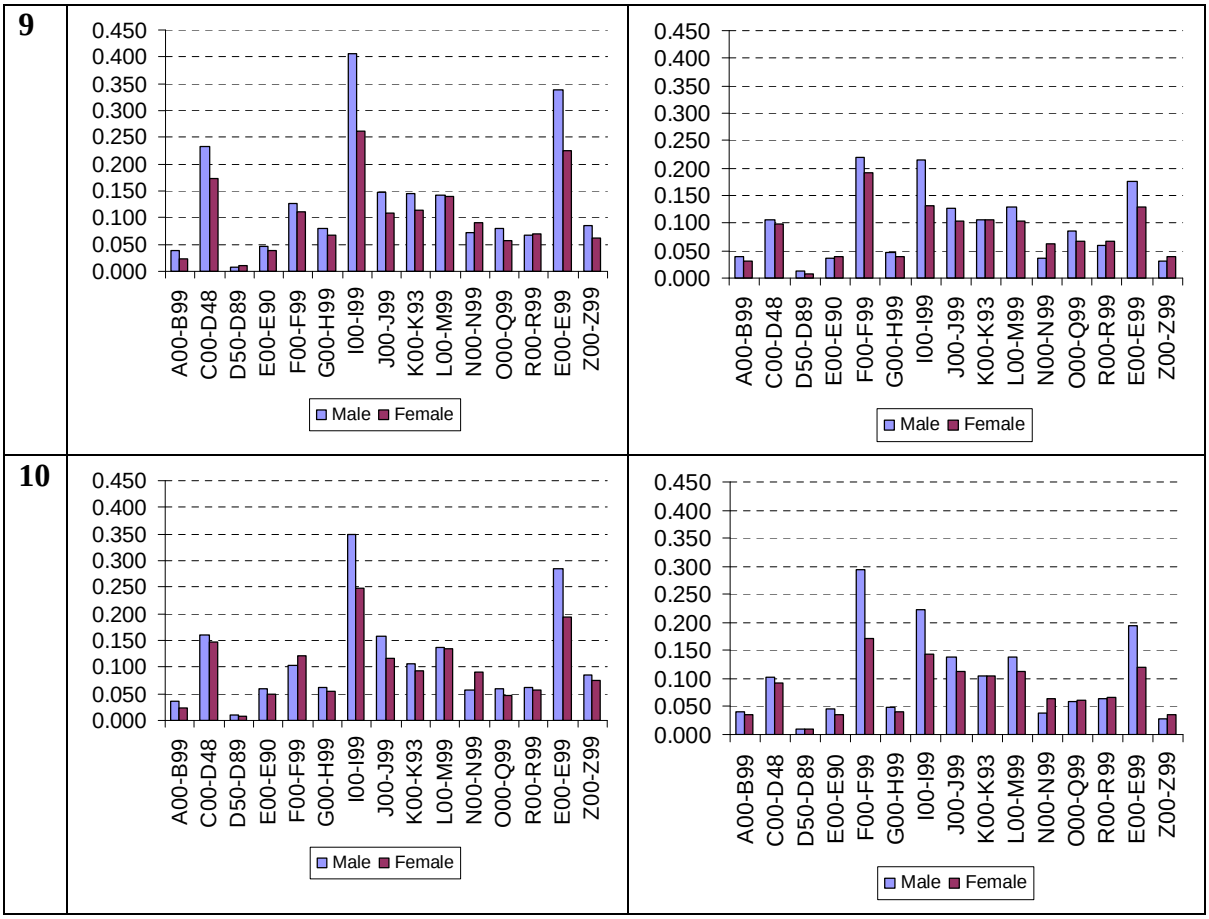
A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

3.5.4 Standardised Hospital Bed Day Rates by Cause and Area Deprivation, 1991-2006

Examining cause-specific hospital use for selected age groups only gives a snapshot of patterns. Trends in cause-specific hospital use by deprivation decile can be better explored through standardisation (see Figure 3.5.5).

Figure 3.5.5: Standardised* Hospital Bed Day Rates by Cause** and Selected Deprivation Deciles, 1991 and 2006





* Standardised to 1991 New Zealand age and sex structure

** ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

The diversity of cause of hospital use decreased markedly between 1991 and 2006 for deciles 1 and 2, but not so much for higher deciles. Bed days due to mental health (F00-F99) increased over the time period, particularly in higher deciles. In 1991, circulatory, neoplasms and external causes were the three leading causes of hospital use in all deciles examined, although there was greater variance of cause in higher deciles. By 2006, these three causes (while still making up a substantial proportion of hospital bed day rates) had decreased in importance, particularly in deciles 1 and 2. In deciles 5, 9 and 10 they had been overtaken by mental health as the single largest contributor to bed day rates in 2006.

3.5.5 Conclusions from Area Deprivation Hospital Bed Day Rate Analyses, 1991-2006

The main trends to emerge from the analyses of hospital bed day rates by area deprivation score between 1991 and 2006 are:

1. Hospital bed day rates increase with increasing area deprivation; more deprived areas have higher hospital bed day rates.
2. Hospital bed day rates decreased over the time period, regardless of the level of area deprivation. This was mainly due to declines at older ages (65 and above), which saw declines in bed day rates of up to 50 percent. The gap in bed day rates between areas with high and low deprivation scores did not decrease over the time period.
3. Males have higher hospital bed day rates than women in areas with high deprivation scores, and women have higher bed day rates than men in areas with low deprivation scores.
4. Cause-specific hospital use for areas with high and low deprivation scores is mostly quite similar, with a couple of exceptions. There are excess hospital bed day rates due to respiratory diseases at ages 0 to 4 in areas with high deprivation scores, and due to circulatory, neoplasms and external causes for males aged 60-64 in areas with high deprivation scores.
5. The diversity of cause of hospital bed days decreased over the time period due to disproportionate decreases in the main causes of hospital use, but more so for areas with low deprivation scores than for areas with high deprivation scores.
6. Bed days due to mental health overtook circulatory as the leading cause of hospital use in 2006, but only in areas with mid to high deprivation scores.

3.6 Conclusion

The analyses undertaken in this paper have highlighted stark differences in mortality and hospital bed day experiences for men and women, and for areas with different deprivation scores. The next section discusses these findings in more detail, and offers explanations for these differences. However, to reiterate, the key findings are listed once more in this section.

Both mortality and hospital bed day rates declined over the time periods in question, and that this happened for both males and females and for all deprivation deciles. However, the gap in rates between areas with higher and lower deprivation scores did not decline, inequalities in health remain prevalent. This was found to be the case for both mortality and hospital bed day rates. The bulk of the difference in rates between areas with different deprivation scores is found at ages 50-79 for mortality, and ages 65 and above for hospital use (with some convergence at ages 80+). In contrast, the gap between male and female mortality and hospital bed day rates *did* decline over time, due to a larger decrease in male rates than in female rates, particularly at older ages.

The three main causes of death over the time period were circulatory, neoplasms and respiratory, followed closely by external causes. Of these, only circulatory rates saw much of a decline over the time period. People spend time in hospital due to a wider range of causes than people die from, but the three leading causes of hospital use are similar: circulatory, neoplasms and external causes, with mental health becoming more important over time, but only in areas with mid- to high-deprivation scores. However, while neoplasm mortality rates remained much the same over the time period, neoplasm hospital bed day rates declined. The patterns of cause of death and cause of hospital use were very similar between areas with high and low deprivation scores.

Mortality rates by deprivation score appear to vary by sex, males have higher mortality rates than females except in the least deprived areas. In the least deprived areas females experience higher mortality rates than males, primarily due to circulatory conditions. An interaction between deprivation and sex is also seen for hospital bed day rates, males have higher hospital bed day rates than women in areas with high deprivation scores, whereas in areas with low deprivation scores women have higher hospital bed day rates than males. The fact that this pattern of abnormally high female rates in areas with low deprivation is present in both the mortality and the hospital data suggests that it is a real health phenomenon, not simply an artefact of the dataset.

This is only a brief summary of the findings from this analysis. The next section explores these results in greater detail.

SECTION 4 – DISCUSSION

4.1 Introduction

Mortality rates, while a valuable demographic measure of quantity of life, cannot alone describe the morbidity experience of a population. And hospital bed day rates, while providing a measure of use of health services, are influenced by a multitude of confounders and need careful adaptation (filtering) and interpretation when used as a measure of morbidity. This work combines both mortality and hospital utilisation data, drawing on the strengths of both types of data.

Mortality and morbidity measures are inherently connected. By comparing rates for mortality and hospital utilisation, an assessment of the quality of hospital utilisation in inferring mortality patterns can be achieved. Being a vital statistic, the mortality data are of better quality than the hospital data, which are liable to be inconsistent over time. Comparing hospital utilisation against mortality data also tests the quality of the hospital data over time. The results presented in Section 3 therefore examine the suitability of hospital utilisation as a proxy for morbidity.

That is not to say that if hospital utilisation trends do not exactly match mortality trends, hospital data are inadequate as a proxy for morbidity. It is not expected for cause specific mortality and hospital utilisation patterns to be identical. After all, many illnesses that require hospitalisation are not fatal. But if hospital utilisation patterns differ markedly from mortality patterns, this may suggest a failing of the hospital data in representing morbidity. Comparing mortality with hospital utilisation trends enables the identification of 'real' morbidity effects, identification of where the hospital use represents ill health, and 'false' morbidity effects, and identification of where the hospital use does not represent ill health but is driven by some other factor.

These results also provide a background understanding of mortality and hospital utilisation patterns from which to conduct more complex analyses. An understanding of the separate mortality and hospital utilisation trends is an essential footing necessary for interpreting combined mortality and hospital utilisation measures, such as Hospital Utilisation Expectancies (HUEs), which are examined in subsequent papers. This section discusses the results from Section 3 in more detail, examining trends from the national analysis and from the area deprivation analysis. In the remainder of the section, regional

results from a recent monograph on mortality and hospital use in New Zealand are compared with the results from the deprivation analysis in this paper. If the results from the deprivation analysis are robust, similar trends should be found by broad geographic region ranked in order of average deprivation score. The relevance of the findings in relation to the research questions and objectives that were outlined at the start of this paper, are examined in the penultimate section. Lastly, readers are reminded of the limitations of this analysis and caveats associated with these results.

4.2 National Analysis, Discussion of Results

The similarities between mortality and hospital bed day rates over the time period 1974 to 2006 are striking. Both the hospital and mortality analyses display a reduction in rates, particularly at older ages; a convergence between male and female rates; fluctuations in rates at older ages in the late 1970s and early 1980s; higher male than female rates at ages 50 and over; and similarities in the main cause-specific rates. This congruence adds support to the suitability of hospital bed day rates as a measure of morbidity.

There are also some notable differences. As expected, people were shown to spend time in hospital for a much wider range of causes than people die from. This is due to the fact that not all sickness episodes result in mortality, and that some types of sickness are not fatal. For hospital bed day rates, by the end of the time period the peaks of circulatory, external causes, and neoplasms sank into the general ‘noise’ of cause-specific hospital bed day rates. In contrast, for mortality rates the only diagnostic group to show considerable reduction over the time period was circulatory. Decline in circulatory mortality drove the overall decline in mortality over the time period. Male circulatory mortality rates were higher than female circulatory mortality rates in 1974, but by 2006 they had converged. Circulatory conditions therefore not only drove the overall mortality decline, but also the convergence between male and female mortality. The second most common cause of death over the time period was neoplasms. Neoplasm mortality rates showed almost no decline, despite declining neoplasm hospital bed day rates.

Smoking is known to have a large impact on circulatory mortality rates, however there is a lag between exposure and increased mortality rates. International studies have shown that recent global convergence in male and female mortality rates can be almost entirely explained by differences in the timing of increases and decreases of cigarette

smoking rates for men and women (Pampel 2002). Internationally, males adopted smoking far earlier than females, but male cessation of smoking in more recent years occurred more rapidly for males than females. Pampel examined the proportion of mortality that was attributable to smoking in developed countries between 1975 and 1995, using data previously published by Peto *et al*. (Peto, Lopez *et al.* 1992). An excerpt from this data for New Zealand is published in Table 4.2.1.

Table 4.2.1: New Zealand Mortality Rates for ages 35-69, 1975 and 1995, and the Percentage Attributable to Smoking

	Male				Female			
	Total	Smoking*	Other	% of mortality attributed to smoking	Total	Smoking*	Other	% of mortality attributed to smoking
1975	11.88	3.86	8.02	32.49	6.74	0.91	5.83	13.50
1995	6.66	1.75	4.91	26.28	4.76	1.05	3.71	22.06

*Smoking related deaths indirectly estimated from lung cancer mortality rates
Source: data published in (Pampel 2002)

It is useful to examine the patterns of smoking related mortality for males and females in New Zealand in the last few decades, and identify how much of the convergence between male and female mortality rates can be attributed to smoking. In 1975, 32.5 percent of male mortality and 13.5 percent of female mortality for ages 35-69 was related to smoking. This declined by less than a fifth for males to 26.28 percent in 1995, but *increased* by nearly two thirds to 22.1 percent for females in 1995. Females also saw a rise in smoking mortality rates over this period, from 0.91 to 1.05. This data shows that, over the period 1975 to 1995 male smoking related mortality rates at ages 35 to 69 declined substantially, while female smoking related mortality rates rose slightly despite declining all-cause mortality rates. It also highlights the effect of smoking related mortality on all mortality. Despite large declines, in 1995 smoking related mortality constituted over a quarter and a fifth of all deaths for males and females respectively.

Assuming that the data published in Pampel are reliable (both in terms of overall mortality rates and the technique used indirectly to estimate mortality attributed to smoking) it would appear that smoking trends can explain a large proportion of the convergence seen in male and female mortality at ages 35-69 over the period 1975-1995.

Returning to the results in Section 3, the most striking difference between mortality and hospital bed day rates between 1974 and 2006 is the substantially higher male than female mortality rates, and the slightly higher female than male hospital bed day rates, when standardised for age. It can be demonstrated that male mortality rates are higher than

female mortality rates at every age in most developed countries. If hospital utilisation is an adequate proxy for morbidity, should male bed day rates be expected to be higher than female bed day rates? A previous study using the same data did find standardised hospital bed day rates to be higher for males than females (Pool, Baxendine *et al.* 2009). A slightly different filter was used by Pool *et al.* which excluded a wider range of hospital stays, and resulted in higher male than female bed day rates. This emphasises how sensitive the hospital data are to the type of filtering process undertaken, gender differences in hospital use should be interpreted with caution. However, as with other measures of morbidity, gender is also a confounder. Women have been found to report worse self-reported health and have more hospital admissions than men, and yet have lower mortality rates than men at every age (Case and Paxson 2005). Women also tend to be more willing to seek health care than men. The result of this may be that males are more likely to present at hospital when the condition is critical, resulting in a short stay followed by death.

4.3 Area Deprivation Analysis, Discussion of Results

There were considerable similarities between mortality and hospital bed day trends across deprivation deciles. Rates for both mortality and hospital bed days were higher in more deprived areas, particularly at middle to older ages (50+). But although there was little difference in mortality rates by deprivation at ages 80+. Mortality rates stay very low until well into middle and early old age, and then rise fairly sharply. In the most deprived deciles this rise in rates occurs about 10 years earlier than in the least deprived deciles. For hospital bed days, rates rise much more gradually over the life course. However for the least deprived areas, hospital rates tend to stay very low before increasing from about age 60. In more deprived areas the increase in hospital bed days with age occurs much more smoothly and at an earlier age, with no real turning point from which rates begin to climb.

The most interesting finding from the area deprivation analysis was that, for both mortality and hospital bed days, standardised rates were higher for females than males in the *least* deprived deciles, and higher for males than females in the *most* deprived deciles. This pattern was less consistent for mortality. As we have learnt from the national analysis, the normal state of affairs for mortality is for male rates to be considerably higher than female rates, and the normal state of affairs for hospital bed days, is for female rates to be

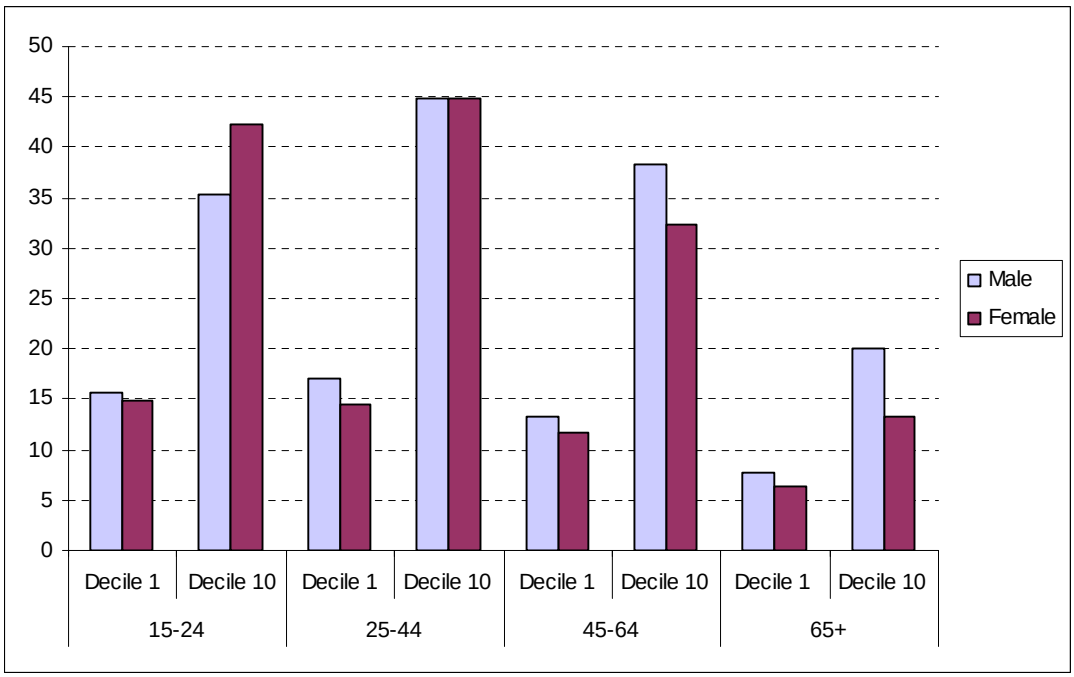
slightly higher than male rates[11]. So why is this anomaly present for both mortality *and* hospital data? The same trend occurring in the mortality dataset as well as the hospital dataset suggests that this is a 'real' morbidity effect, not an artefact of the health system and gender differentials in health seeking behaviour. In examining the cause-specific standardised rates for the time period however, there appear to be different factors at play for mortality and for hospital bed day rates.

For mortality the anomaly of higher female than male rates in less deprived areas appears to be largely due to circulatory conditions. In the least deprived deciles women have significantly higher circulatory mortality rates than males from 1994 onwards, while in the more deprived deciles men have significantly higher circulatory mortality rates for most years. For hospital bed days, the anomaly of higher male than female rates in more deprived areas is due to a range of causes. In the most deprived deciles men have much higher bed day rates than women for external, respiratory, circulatory and (in later years) mental health causes. While men also have higher rates of many of these causes in the least deprived areas, the difference tends to be smaller.

If, as has been suggested, smoking is an important factor in determining circulatory mortality and hospital bed day rates, we should expect to find higher smoking rates in more deprived areas than in less deprived areas. It would also be expected to find higher female than male smoking rates in less deprived areas, and the converse in more deprived areas. However, existing research only half supports this theory, see Figure 4.3.1. The most deprived areas have smoking rates two to three times higher than the least deprived areas. In 1996, 44.8 percent of males and females aged 25 to 44 in decile 10 were regular smokers, compared to 17 percent of males and 14.4 percent of females in decile 1. However, in less deprived areas, males have higher smoking rates than females at all ages, but in more deprived areas, females have higher smoking rates than males at young ages (15 to 24) and similar smoking rates at middle ages (25 to 44) (Salmond and Crampton 1999; Crampton, Salmond *et al.* 2000).

[11] Given the filtering process undertaken. Recall from the discussion of the national results that other research using different filters has legitimately found higher male than female bed day rates, see Pool, Baxendine *et al.* 2009.

Figure 4.3.1: Regular Smoker Rates for Deprivation Deciles 1 and 10, by Gender and Age, 1996



Source of data: (Salmond and Crampton 1999)

Differences in smoking rates by deprivation may be a factor behind the higher mortality and hospital bed day rates of circulatory, neoplasms (cancers) and respiratory conditions in more deprived areas[12]. However gender differentials in smoking cannot explain the higher female than male mortality and hospital bed day rates in the least deprived areas.

4.4 Regional Variations in Mortality and Hospital Bed Day Rates, a Comparison with Previous Research

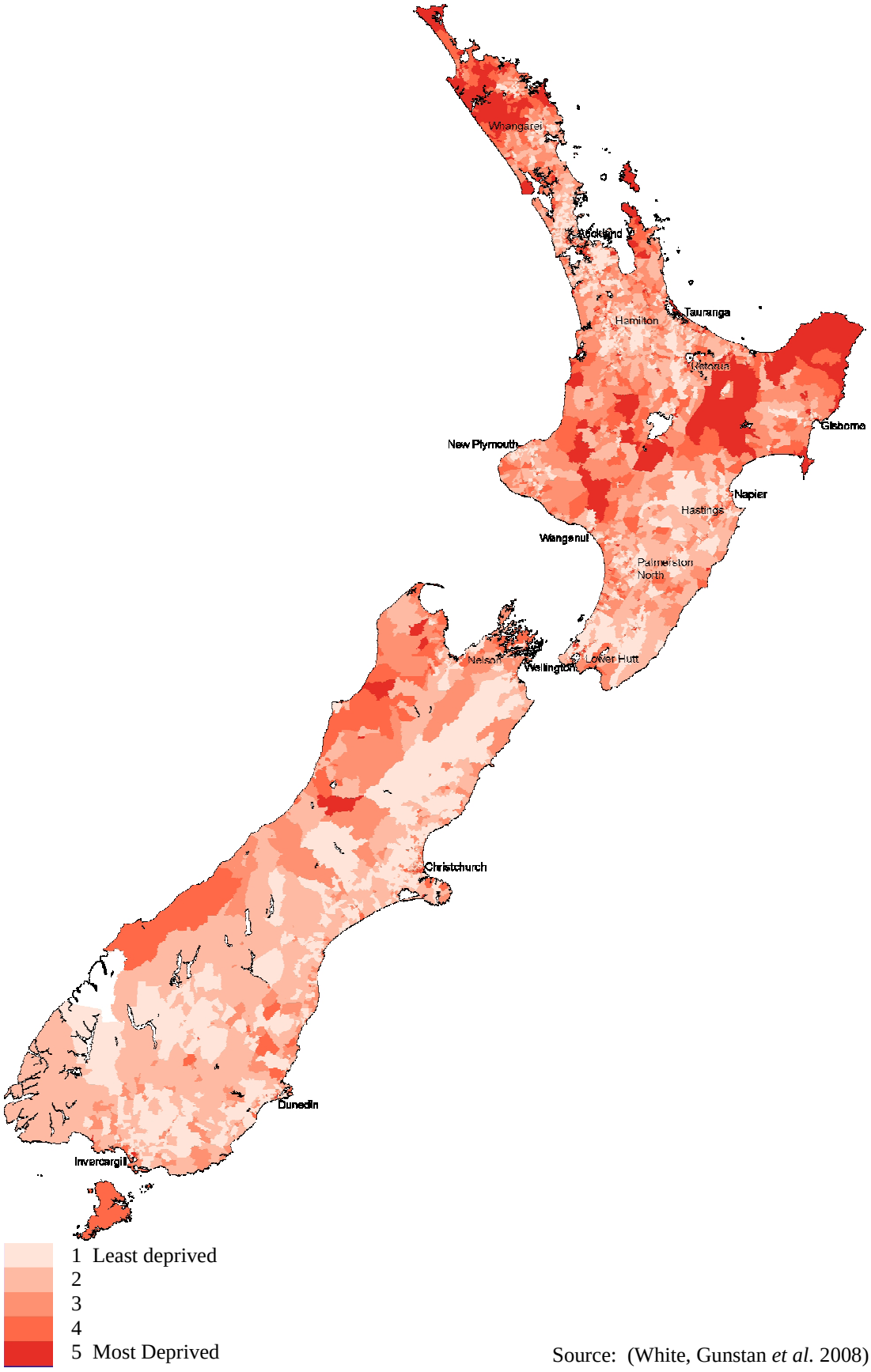
Deprivation deciles in New Zealand are not distinct geographical regions, but are made up of small areas located around the country. However, while difference in mortality and hospital use by deprivation decile have been demonstrated in this paper, there are distinct *regional* variations in mortality and hospital use in New Zealand (Pool, Baxendine *et al.* 2009). In this section, the regional findings of the Pool *et al.* study are contrasted with the

[12] Although bear in mind the lag between exposure to cigarette smoking and mortality, which can be several decades (Pampel, 2002). Smoking patterns by deprivation in 1996 may not be representative of previous decades, but unfortunately such comparisons cannot be made because NZ*dep* was only calculated from 1991.

area deprivation findings. An average NZ*dep* score is calculated for broad health regions in New Zealand, against which health and demographic indicators are compared.

Figure 4.4.1 shows the average NZ*dep*2006 score in New Zealand at a meshblock level, broken down into quintiles. While care must be taken in interpretation of such maps given the incongruity between geographical area size and population size, two comments need to be made. Firstly, there is a North South divide in deprivation in New Zealand, with far more dark areas (representing the most deprived fifth of areas) in the North Island than the South Island. The second point is that in examining geographical regions, large variations in area deprivation at a smaller level are obscured. The area around Auckland looks quite pale (not very deprived) yet South Auckland has some of the highest deprivation levels in New Zealand. The same also applies for Rotorua.

Figure 4.4.1: Average NZ*dep*2006 Quintiles by Meshblock in New Zealand



Source: (White, Gunstan *et al.* 2008)

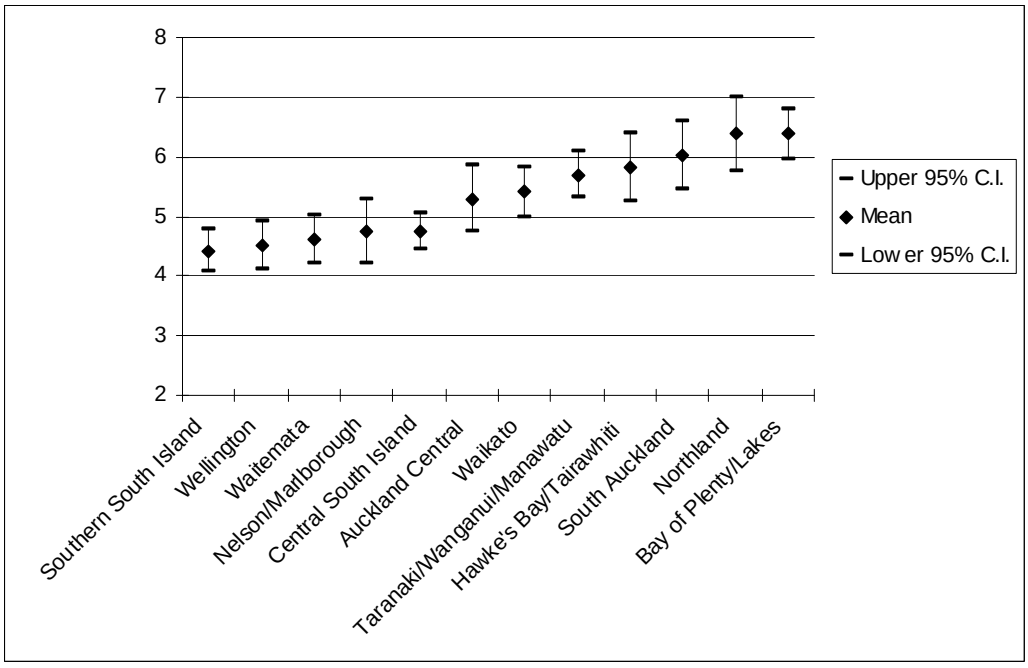
Figure 4.4.2 shows the 12 health regions drawn up by Pool *et al.* They were defined to be relatively similar in population size, and consist of neighbouring Territorial Authorities (TAs). Contrasting with Figure 4.4.1, areas in Southern South Island, Central South Island, Wellington and Waitemata tend to be in the least deprived quintiles, while areas in Northland and Hawkes Bay/Tairawhiti tend to be in the most deprived quintiles.

Figure 4.4.2: New Zealand Health Regions, as Defined by Pool *et al.* 2009

Source: (Pool, Baxendine *et al.* 2009, p17)

Contrasting the pattern of deprivation in Figure 4.4.1 with the health regions mapped in Figure 4.4.2 is rather subjective. In order to quantify this analysis, the average NZ*dep*2006 score has been calculated for each of the 12 health regions, derived from CAU NZ*dep*2006 scores (see Figure 4.4.3).

Figure 4.4.3: Average NZ*dep*2006 Score by Health Region*



*In order of increasing mean NZ*dep*2006 score.
Sources of data:

Health Regions by Territorial Authority obtained from Pool *et al.* 2009, Appendix Table 1.
NZ*dep*2006 scores by CAU and area concordance information obtained from University of Otago Website (University of Otago Website 1988-2001).

The North South divide is demonstrated in Figure 4.4.3. All but one of the five least deprived regions is located in the South Island, or at the southern most tip of the North Island (Wellington). Waitemata is the exception, as the third least deprived region, but also the second most northern region. The four most deprived regions are all located in the northern half of the North Island. South Auckland (which did not show up as particularly deprived in Figure 4.4.1 due to its small geographical area) is the third most deprived region. The five least deprived regions are all significantly less deprived than the five most deprived regions. There is however considerable overlapping of confidence intervals between individual regions.

The wide confidence intervals for the mean NZ*dep* score emphasises the diversity of the health regions. All of these regions consist of lots of small areas which have a wide range of variation in NZ*dep* scores.

Pool *et al.* used these health regions as a unit of analysis in their recent monograph of health and hospital utilisation in New Zealand. Selected findings for 2001 from the monograph are presented in Figures 4.4.4 to 4.4.7. The regions are presented in order of

increasing NZ*dep*2006 score from left to right, the same order as in Figure 4.4.3[13]. This allows for comparison between the gradient in deprivation across regions and the gradient in the indicators presented. The five most deprived areas had significantly different average deprivation from the five least deprived areas. It will be interesting to observe if the indicators shown in Figures 4.4.4 to 4.4.7 also differ in the same way.

Life expectancy decreases with increasing average area deprivation, although it is by no means a consistent pattern (see Figure 4.4.4). However, a smooth, sequential pattern in life expectancy by average area deprivation is not expected, given the overlapping confidence intervals seen in Figure 4.4.3. Waitemata, the third least deprived area, stands out as having the highest life expectancy at 79.9 years. Hawkes Bay/Tairawhiti, the third most deprived area has the lowest life expectancy at 76.5 years. Even at such a broad regional level, there is a difference in life expectancy of over 3 years. Despite the lack of a smooth gradient, the five least deprived areas all have higher life expectancy than the five most deprived areas.

The percentage of the population that is Maori ranges from 7.5 percent in Southern South Island to 29.1 percent in Northland (see Figure 4.4.5). The percentage Maori increases with increasing area deprivation. The six least deprived regions all have lower percentages Maori than the five most deprived. The lowest and highest percentages Maori by health region almost match the extremes shown for deprivation decile, of 4.7 percent Maori in the lowest decile and 35.0 percent Maori in the highest decile in 2001 (see Figure 1.2.1 in Section 1). The strength of the deprivation gradient for ethnicity is an issue of concern. Ethnicity is so closely related with deprivation in New Zealand that some studies have used percent Maori as a proxy for deprivation (Dharmalingam, Pool *et al.* 2004). It is hard to know if patterns in area deprivation or ethnicity are being represented in the results.

[13] While the NZ*dep* scores relate to 2006 and the data relate to 2001, at such a large geographical level changes in the average NZ*dep* score between 2001 and 2006 should be minimal and are not considered problematic.

Figures 4.4.4 to 4.4.7: Demographic and Health Indicators by Health Region in 2001*

Figure 4.4.4: Life Expectancy

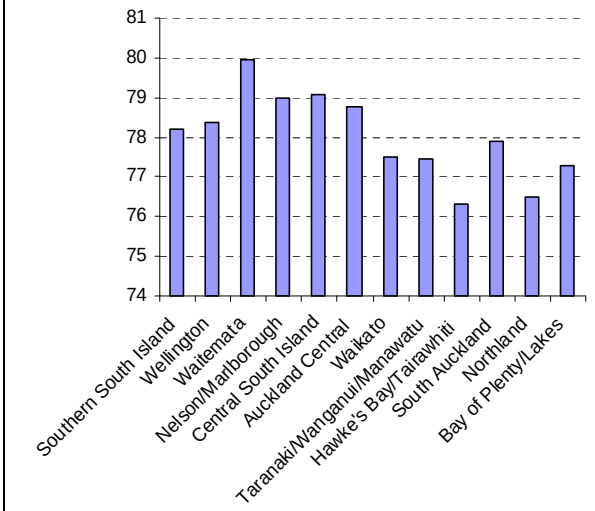


Figure 4.4.5: Percent Maori

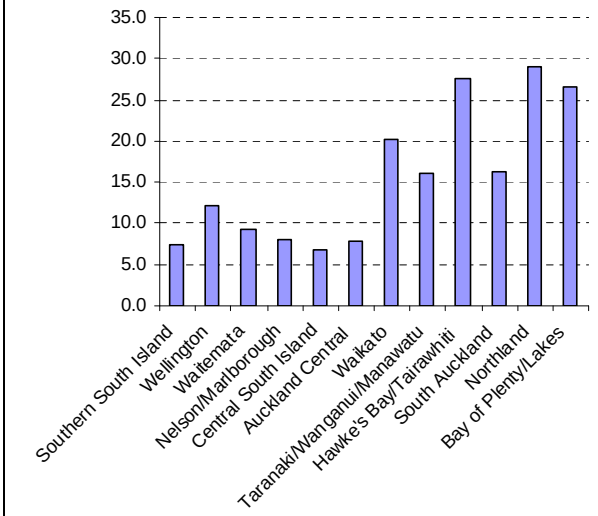


Figure 4.4.6: Age Standardised Mortality Rate per 1000

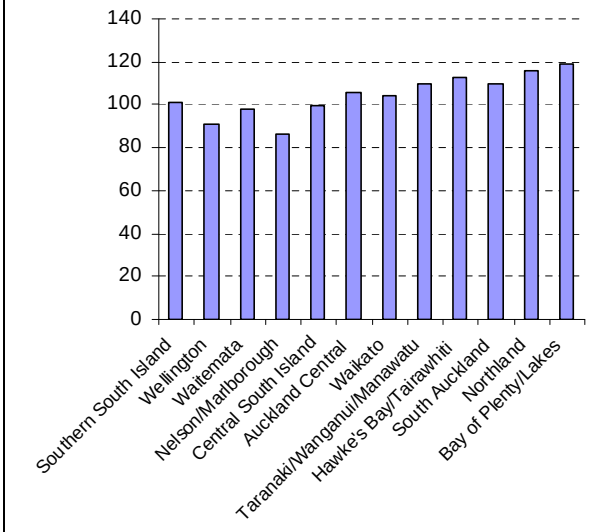
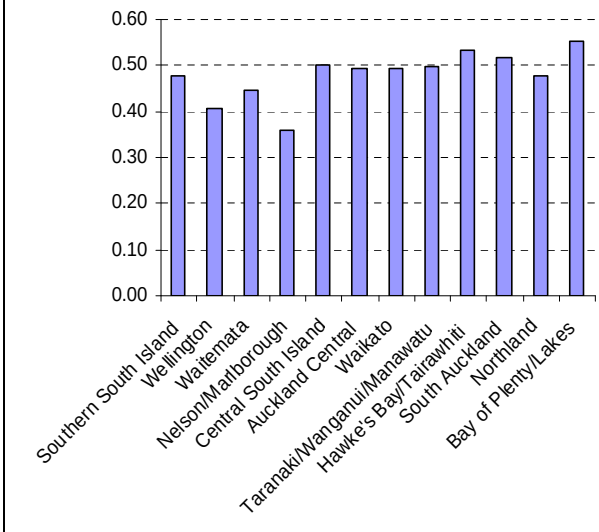


Figure 4.4.7: Age Standardised Hospital Bed Day Rate



*Regions presented in order of increasing average NZ*dep*2006 score, from left to right
Source of data: (Pool, Baxendine *et al.* 2009)

Age standardised mortality rates appear to show a less clear deprivation gradient than life expectancy and percent Maori, although this is largely due to the scale used (see Figure 4.4.6). The five least deprived regions all had lower standardised mortality rates than the five most deprived regions. The same cannot be said for age standardised hospital bed day rates (see Figure 4.4.7). There were a couple of overlaps in mortality rates between the five least deprived and the five most deprived regions. However, with the exception of one very slight overlap (standardised bed day rate of 0.477 for Southern South Island and 0.476 for Northland) the *four* least deprived regions had lower standardised mortality rates than the four most deprived regions. Unfortunately the use of a different standard

population by Pool *et al.* means the standardised rates in Figure 4.4.6 and 4.4.7 cannot be compared with the standardised rates from the previous section.

At first glance, the results from Figure 4.4.7 seem disappointing. One would expect to see a more defined trend in age standardised hospital bed day rates by average deprivation of health regions. But given the diversity of the areas within each health region, as demonstrated by both Figure 4.4.1 and the large confidence intervals in Figure 4.4.3, it would be remarkable to achieve a smooth, consistent deprivation gradient for any of the indicators in Figures 4.4.4 to 4.4.7. Hospital bed day rates have the weakest deprivation gradient, but given relative inconsistency of the deprivation gradient for the other measures (all of which have strong, well documented associations with area deprivation) this is to be expected. The overlaps between the five most deprived regions and the five least deprived regions were due to high levels of bed day rates in Southern and Central South Island. One possible reason for this is the high proportion of remote, rural areas within these regions. Patients may be kept in for an extra day in case their condition deteriorates, due to the long distances and time involved in reaching a hospital.

4.5 Relevance of Findings to the Research Questions

As is the case with research, the initial objectives are often sidetracked when unanticipated patterns emerge from the data, such as the interaction between gender and deprivation for both mortality and hospital bed day rates. The results presented in Section 3 did however answer the research questions set. To recap, the research questions for this paper were:

1. How do the **patterns of mortality** in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of mortality differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

2. How do the **patterns of hospital use** (bed day rates) in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of hospital use differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

The cause-specific distribution of mortality does not differ very much by deprivation decile. The peaks in mortality rates are for the same things (circulatory, neoplasms, respiratory, external causes) they are just a lot higher for more deprived

deciles. There is however a gender interaction, with women in less deprived deciles having higher circulatory rates than men, and men in more deprived deciles having higher circulatory rates than women. The diversity of cause of death did not differ overly by deprivation decile. If anything there is slightly more diversity in the less deprived deciles, due to lower peaks for circulatory and neoplasm causes of death. Over time circulatory mortality rates showed large declines, across all deciles, but neoplasm mortality rates (the second most common cause of death) did not decline. Circulatory rates declined more for men than for women over the time period, resulting in higher age standardised mortality rates for women than for men towards the end of the time period.

Very similar results were found for hospital bed day rates, strengthening their interpretation as a measure of morbidity. The diversity of cause of hospital use was far greater than for mortality, as was expected. Unlike the mortality data, where there were only a few key causes of death with very low rates for everything else (even in more deprived deciles), for the hospital data the rates for nearly all causes increase with deprivation. Increases in rates with deprivation are not limited to just a few causes. However, the cause-specific peaks are in the same places for more and less deprived areas, with one notable exception: in later years more deprived areas have higher hospital bed day rates due to mental and behavioural disorders, particularly for males. This is not observed in the least deprived deciles. The diversity of cause of hospital use has increased over time, particularly in the least deprived deciles. Nationally, circulatory, neoplasms and external causes gradually sink into the background ‘noise’ of other causes of hospital use. At a national level it is also possible to see the increased importance of mental and behavioural causes of hospital use over time.

4.6 Limitations of this Research

There are numerous limitations associated with this research, particularly for hospital bed day rates.

Area deprivation scores relate to CAUs, which have populations of about 2,000 people. They are not the smallest geographical unit available, meshblocks relate to areas of about 100 people. However, meshblock level information is highly sensitive and not made available outside of the Ministry of Health offices in Wellington. Pockets of deprivation and affluence within CAUs will be missed. In addition, NZ*dep* is a relative measure of

deprivation. Actual levels of deprivation may be increasing or decreasing over time, and the gap between the most deprived and the least deprived areas growing or narrowing.

While mortality data refer to *all* deaths in New Zealand, and are fairly complete and accurate, hospital data are less reliable. Only public hospital events are included in the dataset. Private hospital use (which is likely to be more prevalent in less deprived areas) is missed. This may result in artificial deflating of hospital bed day rates in less deprived areas. Hospital data are not very consistent over time, due to coding changes and health system reforms. The data were extensively filtered to allow consistency over time, with the result that many cases were excluded. In addition, cases were excluded on the basis of not representing ill health, the intended output measure. While every care has been taken in designing a filter that will not distort the data, the patterns present in the hospital data may be strongly influenced by this filtering process. For example, a recent monograph that used a slightly different filtering process observed higher age standardised bed day rates for males than females (Pool, Baxendine *et al.* 2009), while the analysis in this paper found slightly higher rates for females than males, purely as an artefact of the filtering process.

Further caution is necessary when inferring from hospital bed day rates to morbidity. This section has demonstrated that mortality and hospital bed day rates follow very similar patterns, which lends support to this inference (assuming that morbidity and mortality are associated). However, there are many non-health related factors that influence hospital utilisation, including gender, cultural, and socioeconomic differences in health care seeking behaviour, and the relative effects of supply and demand. Extensive research has gone into disaggregating the supply and demand aspects of hospital utilisation in New Zealand, see Pool *et al.* 2009 for details.

4.7 Conclusion

The results presented in this paper have shown that not only are hospital discharge rates positively associated with increasing area-level socioeconomic deprivation (as reported in other studies) but that this association is also present when hospital discharge rates are weighted by length of stay, to produce hospital bed day rates. Cigarette smoking has been identified as a powerful explanatory factor in accounting for the convergence of male and female mortality rates, but not in explaining the high female mortality and hospital bed day rates in the least deprived areas. The close association (with some exceptions) of patterns

in hospital bed day and mortality rates provides support for the use of hospital bed day rates as a proxy for morbidity. It was also useful to demonstrate the socio-economic gradient in mortality and hospital bed day rates at a much larger geographical level, demonstrating the robustness of area deprivation measures in health analysis. However, in examining the results from this analysis it is necessary to consider the wide range of limitations present, notably the high proportion of Maori and Pacific peoples in more deprived areas, and the absence of private health care insurance from the hospital dataset.

PAPER TWO

Examination of Compression of Morbidity using Life Table Techniques

RESEARCH QUESTION 3

3. Taking hospital utilisation as a proxy for severe morbidity, does the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through the use of hospital utilisation expectancies (HUEs)?
    a. Does the lifetime expectation of hospital use at birth and at different ages differ by deprivation decile?
    b. How have these trends changed over time?

ABSTRACT

Life expectancy at birth increases over the time period for both males and females, and HUE at birth decreases, providing support for the compression of morbidity hypothesis. Considerable convergence between male and female HUE at birth is observed, but only slight convergence between male and female life expectancy at birth. Life expectancy is increasing most rapidly at the oldest ages (85+), the same ages at which HUE is decreasing most rapidly. This indicates longer *and* healthier lives. These trends are not uniform by deprivation: less deprived areas have higher life expectancy and lower HUEs than more deprived areas. There is evidence that female life expectancy in the least deprived areas may have reached a maximum limit, and further gains in life expectancy may not be possible. At birth less deprived areas have consistently higher life expectancy and lower HUE, but at age 80 there is no clear deprivation gradient. There is strong evidence for compression of morbidity despite limited evidence of rectangularisation of the survival curve.

Low HUEs in less deprived areas are *not* driven by private health insurance coverage. Despite a large proportion of New Zealanders having private health insurance policies, claims constitute a disproportionately small percentage of total health expenditure in New Zealand. Of greater concern is evidence of financial barriers to primary health care in more deprived areas, which may be artificially inflating public hospital use in these areas.

SECTION 1 – LITERATURE REVIEW: Compression and Expansion of Morbidity Hypotheses

1.1 Introduction

When we die, what we die from, and our health, are matters of great concern to us all. In the last 40 years or so researchers have attempted to understand and document macro trends in mortality and morbidity. The epidemiologic transition theory describes historical patterns and variations in different countries. The compression and expansion of morbidity and the dynamic equilibrium hypotheses were proposed in order to investigate future trends. It has been an era of transition theories: models that describe demographic processes have been formulated including the demographic, nutrition, the fertility and economic transitions. This paper examines life expectancy and life expectation of hospital use to understand the nature of the health transition (compression, expansion or dynamic equilibrium) in New Zealand. The paper begins with a literature review of health transition theories. This is followed by an examination of health transition theories in New Zealand using life table methods, and then a discussion in which private health care insurance is explored as a potential confounder of hospital utilisation in later life.

The key points of the epidemiologic transition are summarised in the literature review. Health transition theories (compression/expansion/dynamic equilibrium) are described and the evidence reviewed. This is followed by discussion about life span limitations, inequalities in health between sub-populations, and methodological concerns, all of which are relevant to the integrity and interpretation of the proposed morbidity trends.

1.2 Epidemiologic Transition

The epidemiologic transition was proposed by Abdel Omran in 1971, and describes the history of mortality patterns, the changes in life span, and the causes of death that result in the mortality experience of the present day (Omran 1971, p476; Omran 1983). The end result (declining mortality from infectious causes, increased mortality due to degenerative and man-made diseases, increases in life expectancy, and decreases in fertility) is

anticipated to be the same for all countries, but the tempo and mechanisms through which this is achieved varies. Different versions of this model were proposed to describe the ongoing process in various countries. In developing countries the epidemiological transition is characterised by a slower transition, and the importance of medical interventions in controlling infectious diseases, resulting in declining mortality without a corresponding drop in fertility. This is the 'classic' model. In countries where the transition occurred more quickly such as Japan, the 'accelerated' model denotes swift declines in fertility, in part enabling rapid declines in mortality.

Omran initially identified three stages to the epidemiologic transition. The first stage is an age of pestilence and famine, characterised by high and fluctuating death rates and high fertility rates. This gives way to an age of receding pandemics, where mortality rates fall and spikes in mortality become less frequent. This stage is accompanied by large increases in population size due to natural growth, as the mortality rate falls but the fertility rate remains high. In the third stage mortality continues to decline, eventually stabilising at a low level, and degenerative and man-made diseases become the principal causes of death (Omran 1971). In the years since this transition theory was proposed, researchers have hypothesised a fourth stage, the stage we are now entering. Some see it as a hubristic, aspirational stage, where our health and mortality fall under our control (for better or worse) and most causes of morbidity and death are either caused by us, or can be controlled by us. This is accompanied by an illusion of immortality: that one cannot be harmed or suffer (Rogers and Hackenberg 1987). However, while we can influence the progression (and in some cases the contraction) of various diseases in the hubristic stage, this necessitates behavioural changes at the individual level. Such changes will not occur if individuals become over-confident in the ability of medical advances in solving their health problems. It is interesting to note the secondary definition of hubris in the Collins dictionary:

> **Hubris, hybris (n)**
> 1. pride or arrogance
> 2. (in Greek tragedy) an excess of ambition, pride, etc., ultimately causing the transgressor's ruin
> (Collins English Dictionary 5th Edition 2000)

Other researchers, considering the high life expectancy we are now experiencing, see an age of delayed degenerative diseases: healthy life until late middle age, when

diseases related to the ageing process (stroke, cardiovascular disease and cancers) gradually set in, culminating in illness and death at a relatively old age (Olshansky and Ault 1986).

Omran disputed fourth stages such as the hubristic stage and the age of delayed degenerative diseases. He argued that they relied on the experience of a single country, and emphasised the diversity of the path of the epidemiologic transition. Instead he suggested a fourth stage characterised by declining cardiovascular mortality, ageing, lifestyle modification, emerging and resurgent diseases, followed by a fifth, futuristic stage. The fifth stage he defines as the age of aspired quality of life with paradoxical longevity and persistent inequalities. This stage involves increases in life expectancy to over 90 resulting from both an increase in the life span due to medical advances, and a relative increase in the longevity of disadvantaged groups (Omran 1998).

Although the epidemiologic transition theory is widely seen as a useful way to describe the changes in predominant causes of death and in life expectancy as populations advance through the demographic transition, it has drawn criticism. Rogers and Hackenberg criticised Omran's transition for not considering social pathologies, such as suicide, homicide, violence, drug and alcohol abuse and accidents (Rogers and Hackenberg 1987). Both direct (suicide) and indirect (alcohol abuse leading to liver cirrhosis) social pathologies are becoming increasingly important as causes of death and morbidity in developed countries.

While differences in the transition are apparent between countries and even between different population subgroups within countries, the same broad shifts seem to occur in cause of death. These shifts link in with several other established transition theories: the demographic, fertility, and socioeconomic transitions to name a few. These transitions act as a base from which to generate predictions about the nature of future demographic trends. The compression and opposing expansion of morbidity hypotheses are two such predictions. The dynamic equilibrium hypothesis fits somewhere in the middle of the two hypotheses.

1.3 Compression and Expansion of Morbidity Hypotheses

1.3.1 Compression of Morbidity Hypothesis

The compression of morbidity hypothesis was proposed by James Fries in 1980, and developed over many years until reaching paradigm status (Fries 1980; Fries and Crapo 1981; Fries 1983; 1984; 1989; 1990; 1992; 2000). It follows on from the epidemiologic transition by, with a few notable assumptions, applying the mortality patterns described by the epidemiologic transition to the future population structure. Assumptions are that the human life span is relatively fixed, life expectancy is very gradually increasing and will level out at about 85 years, and chronic, degenerative diseases associated with ageing are superseding acute diseases as the main cause of death. Hessler *et al.* summarised the compression of morbidity hypothesis in four premises:

1 - Human life span is finite
2 - Number of very old will not increase much
3 - Age at first (disabling) illness will increase
4 - Length of disabling illnesses will decrease from younger to older cohorts
(Hessler, Eriksson *et al.* 2003)

The second point refers to the number of centenarians and people living considerably longer than the current life expectancy. This is related to the argument that the human life span is finite, and that variation in the age at death will continue to decrease resulting in increased survival at older ages. Life expectancy is considered to have an upper limit, estimated at about 85, at which point it will stop increasing. The proportion of the population in the older age groups will grow due to increased survival at older ages (and, to a lesser extent, lower fertility and large historical birth cohorts), resulting in a bigger elderly population 'at risk' of getting ill. However, Fries argues that if the onset of chronic, non-fatal diseases can be postponed, this would leave a shorter period of illness prior to death, 'compressing' the duration of illness between the moveable age at onset of morbidity, and fixed life span. If the onset of morbidity is postponed enough, it could be postponed until after (the fixed event of) death, representing prevention. The mechanisms he advocates involve redirecting medical research funds to combat non-fatal diseases as well as fatal ones, and a greater responsibility on the individual for their own health. If people decreased unhealthy behaviours (smoking, drug abuse, alcohol, poor diet) and increased healthy ones (regular exercise, low-fat diet), he argues that the onset of chronic diseases causing morbidity could be delayed.

Research and funds in modern medicine have focussed largely on eliminating and postponing ‘killer’ diseases, and diseases that cause morbidity or disability but are not fatal have been somewhat overlooked. This has scornfully been described by Fries as a ‘heroic’ (Fries 1984): it is much more heroic to rescue someone from certain, impending death than to cure chronic incontinence, or gout.

1.3.2 Expansion of Morbidity Hypothesis

Gruenberg raised the question of whether extra years of life are healthy years of life in 1977, going against the widely held assumption (up until that time) that as mortality declines, so does morbidity (Gruenberg 1977). Schneider and Brody challenged Fries’ compression of morbidity theory, questioning the ongoing process of rectangularisation and suggesting that life expectancy is not approaching a limit, a pre-requisite of morbidity compression (Schneider and Brody 1983). Olshansky and others then picked up the baton and developed a counter theory to the compression of morbidity – the expansion of morbidity hypothesis (Johansson 1991; Olshansky, Rudberg *et al.* 1991).

The expansion of morbidity hypothesis argues that a larger elderly population at risk of morbidity will result in more morbidity, and that any gains in life expectancy will result in additional unhealthy years. Hessler *et al.* provide four premises for the expansion of morbidity hypothesis:

1 - Human life span is increasing, because of medical science
2 - Number of very old will increase a lot
3 - Disease is the main cause of death
4 - The longer the life span, the longer the period of disability prior to death
(Hessler, Eriksson *et al.* 2003)

Olshansky *et al.* indicate that the current maximum life expectancy is around 85 years given medical technology, but they expect medical advances and technologies to raise life expectancy further in the future. Due to this, they foresee not only more people surviving to the current life expectancy, but also living beyond. A bigger elderly population ‘at risk’ of getting ill, and increased survival at these ages will, according to the expansion of morbidity hypothesis, result in more people suffering from morbidity, and for longer. "Declining mortality would therefore be hypothesized to accelerate the growth in that segment of the population most likely to experience comorbidities, and extend the

lives of individuals with comorbidities" (Olshansky, Rudberg *et al.* 1991, p211). Medical advances are seen as the key method for controlling morbidity, with far less emphasis on behavioural factors. Contrary to Fries' belief that human behaviour is malleable, the expansion camp consider unhealthy habits to be very difficult to change at a population level. Although seen as the solution, medical advances can also exacerbate the morbidity burden, stopping the chronically ill dying from other (opportunistic) causes and prolonging the time spent in an ill state. There are now "...better techniques for thwarting killers which had been weeding out the chronically ill" (Gruenberg 1977, p18). Omran even considers medical technology to be potentially dangerous "Health hazards can be presented by health care itself, which is getting much more aggressive and more likely to take risks" (Omran 1998, p117). In a study of heart failure in New Zealand, Riddel gave a simple example of this paradox of decreasing mortality and increasing morbidity:

> "The aging of populations and enhanced survival of coronary heart disease patients in developed countries now means that heart failure is a significant public health concern" (Riddell 2005, p118).

Hessler's third premise for the expansion of morbidity hypothesis is somewhat debatable. The logic is that if deaths are primarily due to diseases, not a ticking ageing time bomb, and if these diseases can be prevented, life expectancy will increase. However, even if deaths *are* due to ageing processes, such as frailty and senesence, medical and technological advances will eventually be able to delay, or slow this process, a view shared by several advocates of the expansion hypothesis. Life span and death in the presence or absence of disease are discussed further later.

While not necessarily disagreeing with the logic of the potential for declines in morbidity at older ages if the onset of morbidity were to be delayed, Olshansky *et al.* disagree with the mechanisms that Fries identified for this delay, and bring into question the possibility of delaying the onset of morbidity at all. They argue that there is no scientific, medical, proven evidence that the onset of non-fatal diseases *can* be delayed (Olshansky, Rudberg *et al.* 1991). However, the time at onset of a disease is very difficult to determine, and this may be a factor in the lack of evidence for the postponing of chronic diseases. This is particularly the case for non-fatal diseases which tend to creep up gradually and asymptomatically for several years (Olshansky, Rudberg *et al.* 1991). If the

phenomenon itself cannot be measured accurately, empirical evidence supporting (or opposing) the phenomenon will not exist.

1.3.3 Dynamic Equilibrium

The dynamic equilibrium hypothesis is a mid-way point between the compression and expansion of morbidity hypotheses. It claims that as life expectancy increases, the age at onset of morbidity increases in tandem, thus overall morbidity in a population remains constant (Manton, Stallard *et al.* 1991). As with the compression of morbidity hypothesis, the dynamic equilibrium hypothesis foresees an increase in the onset of chronic morbidity, either due to medical advances or due to behavioural changes. But unlike the compression of morbidity hypothesis, it does not foresee a cap on life expectancy, a fixed maximum which average life expectancy cannot exceed. Very large increases in life expectancy however, are also incompatible. Gradual, modest increases in life expectancy are in line with the dynamic equilibrium hypothesis.

The dynamic equilibrium hypothesis lacks much of the theoretical fanfare of the compression and expansion of morbidity hypotheses. Rather than presenting an opposing theory, it suggests that neither extreme is correct, that the truth lies somewhere in the middle.

Which hypothesis policy decisions are based on has, to paraphrase Hessler *et al.*, 'rather important' implications. Demand for health services, and different types of health services need to be predicted to avoid over or under supply. Governments and pension planners need to know how long people will spend in an ill state after retiring. If people are more likely to be healthy into their 70s and 80s in the future, it may result in an increase of the retirement age. Whether medical research should continue to focus on fatal diseases or pay more attention to non-fatal diseases is also in question; Schneider and Brody warn against the dangers of being 'seduced' by Fries' views on the direction of medical research. These hypotheses should clearly not be taken at face value, on theory alone. A review of the evidence for the different theories is presented below, followed by discussion on life span, health inequalities and methodology, all of which influence the interpretation of these hypotheses.

1.3.4 Evidence for Compression/Expansion of Morbidity

While advocates of the compression and expansion hypotheses both consider the evidence to be overwhelmingly on their side, in reality research has found rather mixed results, but overall there appears to be slightly better evidence for compression of morbidity. Mor reviewed the literature and concluded that in the last few decades most researchers agree that there has been approximately a 1 percent reduction in mortality in older people, and a 2 percent reduction in disability (Mor 2005). While this sounds rather impressive, most studies have found only partial support for compression or expansion of morbidity. For instance, while Doblhammer and Kytir's 2001 Australian study found evidence for a reduction in morbidity, it found no evidence for a limit to life expectancy, instead finding increases in life expectancy that accelerated over the time period 1978 to 1998 (Doblhammer and Kytir 2001).

In 2003 Hessler *et al.* used Swedish data to examine days spent in hospital in the last year of life as a proxy for morbidity. They found that longer-lived individuals seemed to live healthier lives (in the last year of life). "[Individuals] who die between 70 and 85 years of age generally are very ill in the months and even years prior to death. Individuals who live more than 85 years seem to have what colleagues define as physical, social and mental 'vitality' or healthy aging" (Hessler, Eriksson *et al.* 2003, p220-221). This suggests that longer lives may also be healthier lives, although the window of one year prior to death is very limited and may not be capturing the full picture. If compression of morbidity is interpreted as the postponement of morbidity until later in life then these findings offer tentative support for the compression of morbidity hypothesis. The trend is towards increasing life expectancy, and deaths at older ages are associated with shorter periods of morbidity prior to death. Another study using hospital data in the United States (US) estimated the cost of acute care in the last two years of life, and found the same trend. As age at death increased, acute costs in the last two years of life decreased (Spillman and Lubitz 2000). Further support for the compression of morbidity hypotheses is offered by Khaw, who found morbidity in the UK to be decreasing: the proportion of men at any age unable to perform four activities related to daily living halved between 1976 and 1994 (Khaw 1997).

Support for the expansion hypothesis is equally fragmented, with studies offering support for certain aspects of the hypothesis. In a study investigating the work ability of the aged in the US, Feldman found increases in disabilities such as arthritis and other musculoskeletal diseases. Jagger found support for expansion of morbidity in all countries excluding France (Jagger 2000). However, when levels of disability are taken into account,

much of this increase appears to be of less severe disability. Jagger interprets this to suggest that the dynamic equilibrium may be a better fit. Other research has also found that while people are experiencing morbidity for longer periods, it is less severe morbidity than in the past (Gruenberg 1977; Johansson 1991). Differing definitions for morbidity and disability in health research is a serious issue in its own right, and is discussed in more detail later on. There have also been several studies that have found support for the dynamic equilibrium hypothesis (Manton, Stallard *et al.* 1991; Graham, Blakely *et al.* 2004).

The findings of research appear to depend heavily on both the interpretation of the compression and expansion debate, and on the indicators used to gauge morbidity. Spillman and Lubitz for example, found the cost of acute care in the last 2 years of life to decrease with age, supporting compression. They also found the cost of long-term care to increase substantially with age, supporting expansion. When the severity of morbidity is considered, many of the studies supporting expansion are brought into question. These mixed findings emphasise the point made by Hessler *et al.*, that the compression of morbidity hypothesis has not been adequately tested empirically.

1.3.5 The Human Life Span

Life expectancy, and life span, is a central part of the compression of morbidity hypothesis. It is argued that, given that life *span* is relatively fixed and the variance of ages at death is declining, we have reached the peak in longevity and the human life expectancy is not going to increase very much. This is best illustrated by the rectangularisation of survival curves over time. However expansionists argue that life expectancy is still increasing, that life span will be extended by technological advances in the future, and that variance in age at death is not decreasing. In short, that we are not reaching a maximum limit of life expectancy. The importance of trends in life expectancy and variance at age at death was emphasised by Manton. Manton found evidence to suggest life expectancy may be increasing, and to suggest variance of age at death is also increasing. Both findings are inconsistent with rectangularisation, and while current increases in life expectancy are relatively modest he argues that even 'small' increases in life expectancy have big implications for ageing and demand for health services (Manton 1982).

The mechanisms behind old age mortality are crucial to whether life span can be extended, but there are conflicting views about these mechanisms both within and between the compression and expansion advocates. It is clear that as people get older, they age

biologically. And with this ageing comes frailty and senescence, which leave the body vulnerable to opportunistic diseases. The question is whether age at death is cause specific. Particular diseases can be postponed or prevented, but (with current medical technology) ageing cannot. This has implications for the potential of extending the life span.

Fries envisaged 'natural death' to be death at an old age due to frailty and multiple organ failure, in the absence of disease. In this theory, the cause of death for elderly people becomes less and less important compared to the underlying cause of frailty and reduced capacity for organ repair (Fries 1980). Fries later moved away from this theory arguing that, with age, the proximate cause of frailty becomes less and less important compared to the opportunistic disease that results in death (Fries 1984). Hayflick is a strong supporter of natural death, considering ageing to be the main cause of death in the elderly, going so far as to claim that for people aged over 70 or so, the vast majority of cause of deaths recorded by the coroner will be incorrect (Hayflick 2000). While ageing is not a disease itself, it is ageing not the presence of disease that causes death in the elderly. Fries later considered ageing as a three-dimensional polygon, with each side being represented by an organ in the body. Decline in the ability of each of these organs to repair damage occurs at a linear rate, resulting in a 'shrinking' of that side of the polygon. Frailty is a symptom of this declining organ reserve. When the polygon has decayed to the centre, organ failure resulting in death occurs. During the decline the individual becomes progressively more frail and vulnerable to opportunistic infections that can result in death, such as pneumonia, stroke, and coronary heart disease (Fries 1984; Fries 2000).

While some researchers hypothesise about possible future medical advances for delaying ageing and thus mortality, most foresee only a modest increase in life span, to about 90 or so (Olshansky, Rudberg *et al.* 1991; Hayflick 1994; Omran 1998). Longevity goals have been pursued for as long as humans have experienced mortality. While remarkable scientific, medical, and technological breakthroughs have occurred in recent decades, accompanied by large gains in life expectancy, it has been argued that we have not succeeded in increasing human life span, or the percentage of centenarians in a population. Given that it has no precedent, it would be foolish and blindly optimistic to trust that medical advances will substantially extend life span. Such a claim is verging on science fiction. Even if such technology existed it would not necessarily result in a longer *healthy* life, if the expansion of morbidity (or even dynamic equilibrium) hypothesis were to be correct. Further, it would (initially at least) be very expensive and available to only the rich, who already have relatively high life expectancies (Hayflick 1994). However,

there is every possibility that life expectancy, if not life span, *can* be substantially increased, in the near future. High life expectancies are observed among some sub-populations, and life expectancy has been shown to vary by socioeconomic and area characteristics, with up to 10 years difference observed between population groups in some countries. Hefford *et al.* demonstrated a 9 year difference in life expectancy in New Zealand between the least and the most deprived deciles of the population, and between some ethnic groups (Hefford, Crampton *et al.* 2005). Increases in life expectancy are achievable through reducing inequalities (such as deprivation and inequitable access to health care). If everyone lived as long as sub-populations with high life expectancies, life expectancy would increase considerably, without an increase in life span. This goal is much more achievable, realistic, and equitable than aims to increase the human life span.

1.3.6 Rectangularisation of the Survival Curve

Rectangularisation of the survival curve occurs when age at death is compressed into a relatively small age range, close to theoretical maximum life expectancy. It is considered by Fries to be a necessary prerequisite for compression of morbidity to occur (Fries 1980): if life expectancy continues to increase without limit, increases in the age at onset of morbidity will not be able to outstrip increases in life expectancy. It could however be argued that compression of morbidity can occur in the *absence* of rectangularisation of the survival curve, at least in the short- to mid-term, as long as increases in the age of onset of morbidity outstrips increases in life expectancy. If rectangularisation is occurring, when plotted against age the survival curve (the l_x function in a life table) remains close to the radix until close to the age of life expectancy, when the number alive declines quickly. This suggests a biological limit to life expectancy, which can be attained but not exceeded. Rectangularisation of the survival curve is a focal part of the compression of morbidity hypothesis, but this can be argued to be far less important than reduction in variability at the age of death. Rectangularisation does not prove the existence of a fixed upper limit to life span, but if there *is* a fixed maximum life span, rectangularisation is a necessary consequence of sustained decline in mortality (Wilmoth and Horiuchi 1999). The discussion concerning continued rectangularisation is in itself flawed given that complete rectangularisation is hypothetical and unachievable. For complete rectangularisation to occur, *everyone* would have to survive until exact age x, and then die at the same time. What is the maximum achievable or ‘ideal’ rectangularisation? Further, individual differences in the 'biological endowment' of life span (differences in potential life span at

birth) confuse rectangularisation on an aggregate level. As such, alternative models *without rectangularisation* can be made to fit the data equally well due to individual differences in life span (Manton, Stallard *et al.* 1991). To confound matters even further, there is little agreement on how to measure rectangularisation, which is an intuitive concept, with no formal definition (Wilmoth and Horiuchi 1999). Rectangularisation is clearly not an ideal measure for the purposes of proving a limit to life span. "...it is unfortunate that the notion of rectangularization of the survival curve ever became a central part of the discussion about biological limits to human mortality" (Wilmoth and Horiuchi 1999, p476).

1.3.7 Inequalities in Health

Bringing everyone up to the same longevity standard, and inequality in general have not been considered much in the literature (Omran 1998). In an interview with the BBC, Michael Marmot argued that recent increases in health inequalities in the UK indicate that the magnitude of inequalities is flexible. If inequalities can increase in a short period of time, they can also reduce. He plays down the role in medical advances in increasing life expectancy, advocating reducing inequalities in health as a much more effective way of increasing life expectancy (BBC 2008). Pool suggests that epidemiologic polarisation may be occurring in some countries, there may be more than one health transition (compression/expansion/dynamic equilibrium) occurring in any given country (Pool 1994).

House *et al.* investigate cause-specific mortality and morbidity by individual measures of socio-economic status (House, Kessler *et al.* 1990). It is well known that low income, education, and living in a deprived area are associated with a higher risk of dying and a lower life expectancy. But how this translates into different types of morbidity has been relatively unexplored. House *et al.* used a combination of education and income to calculate a socio-economic indicator, and found a non-linear association between age and self-reported health, but only for people in the low socio-economic group. The young and the old tended to have low morbidity levels, while the middle-aged experienced higher morbidity. Other studies have found rectangularisation of mortality among high socioeconomic groups due to a lower distribution of death rates until age 70, when accelerating rates of disability and convergence with other socioeconomic groups are seen. This suggests compression of morbidity may be occurring, but primarily among high socioeconomic groups (Fries 1992). Reducing inequalities in health has huge potential for raising life expectancy and reducing morbidity, without going beyond the bounds of current medical technology.

1.3.8 Methodological Issues

The choice of morbidity indicator can have a big impact on which morbidity hypothesis fits the data. The choice of statistical technique can be equally important, and concepts regarding rectangularisation of the survival curve can influence interpretation of results. These three areas for ambiguity are discussed in this section.

What constitutes morbidity? The interchange of morbidity, disability, and functional limitation in the literature, with no agreed definitions makes comparing results and arguments difficult. Different diseases have different implications for quality of life, disability, functional limitation and hospital use. A blanket term 'morbidity' can be misleading (Mor 2005). Fries identified four types of morbidity: morbidity associated with fatal ‘killer’ diseases, morbidity associated with non-fatal diseases, morbidity associated with the senescent process and morbidity associated with terminal care. Morbidity associated with terminal care is morbidity immediately prior to death that is often extended by intensive medical care (Fries 1992). The type of morbidity with which a morbidity reduction is associated has important implications. Reductions in morbidity associated with fatal diseases extend life, increasing the period at risk of being ill. As such, reduction in these diseases may *not* represent a ‘true’ reduction in morbidity. Instead of dying from heart disease an individual remains alive for longer, during which time they may suffer from an illness or disability. In contrast reducing morbidity from non-fatal diseases does not directly extend life, and therefore does not increase the period at risk of being ill, representing a ‘true’ reduction in morbidity. It could be argued however that reducing morbidity from non-fatal causes may indirectly extend life by a reduction in insult accumulation over the lifecourse.

Apart from the type of disease with which morbidity is associated, the severity of morbidity is also of importance, but very difficult to quantify. As Gruenberg and others identified, an increase in morbidity may not necessarily be a bad thing if that morbidity is of a much less severe nature than previously. Is it better to have a month in chronic pain, or a year in mild discomfort? There are also inherent problems with comparing morbidity over time and across generations, due to different perceptions about health, and what constitutes good or poor health. One study aimed to overcome this by asking individuals to rate the health of their parent (of the same sex) when they were the same age (Fries 1992), although even this is fraught with difficulties. The individuals are likely to have not known or remembered their parent’s health at their age, or to deflate their parent’s previous health due to the parent having poor health at present, and fearing that they will go the same way.

The type of model used to test the hypothesis can alter the results. The observed prevalence life table is the most popular life table for assessing the compression/expansion of morbidity hypotheses, and is used in conjunction with Sullivan's method to generate health expectancies. However this method relies on prevalence rates, and does not identify entry or exit from a morbid state (Rogers, Rogers *et al.* 1990). Recovery from morbidity has been found to occur, studies suggest 19 to 25 percent recovery from dependence over a period of up to two years (Rogers, Rogers *et al.* 1990) yet is ignored by most of the research, which broadly considers morbidity to be unidirectional. In contrast, there is support for Sullivan's method (and associated period prevalence life tables). It has been shown that "…the Sullivan method provides a good estimate of trends in health expectancy in the long term if changes in the prevalence of good or ill health are smooth and relatively regular" (Doblhammer and Kytir 2001, p386). However, the period prevalence model is not the only model to draw criticism. Cause elimination models assume endogenous causes of death to be independent, but this is often not the case. By definition, cause elimination models also consider a cause to be 'eliminated' on contact, but this is an over-simplification. Someone who survived a heart attack, for example, would be eliminated from the risk of future heart attacks. However they would be more likely to die from a subsequent heart attack than the general population (Manton, Stallard *et al.* 1991). Important determinants of morbidity can also simply be missed from models. Old-age morbidity and mortality are affected by health behaviours, physical activity, nutrition and socioeconomic conditions, but these are rarely incorporated into models (Manton, Stallard *et al.* 1991).

1.4 Conclusion

The fact that we are facing an ever growing population at risk of morbidity regardless of life expectancy increases is agreed upon by all sides of the expansion/compression/dynamic equilibrium debate. The importance of conclusions drawn from the hypotheses presented is highlighted by Manton. Supporters of rectangularisation who believe a) we are near an upper limit of life expectancy, and b) the upper limit of life expectancy is determined by senescence (natural death), think that biomedical research should invest in preventing senescence rather than chronic degenerative diseases. Manton instead thinks mortality is more a process of chronic degenerative diseases precipitating

death and disagrees with 'natural death' (Manton 1982). If the ‘wrong’ research path is pursued, it would represent a diversion of funds away from research which would have the potential to prolong lives and reduce morbidity.

This research area has far reaching implications for health policy. Given the potential for increasing life expectancy and reducing morbidity by reducing inequalities, more research investigating cause-specific mortality and morbidity differentials by socioeconomic factors is needed. Of particular interest is the type of health transition occurring in different sub-populations within the same country. Is epidemiologic polarisation occurring? In this paper a variation of health expectancies, Hospital Utilisation Expectancies (HUEs) are used to examine population health trends over time, by area deprivation, cause and gender in New Zealand. HUEs incorporate both mortality and morbidity (hospital utilisation) information and are discussed further in the next section.

SECTION 2 – DATA AND METHODS: Calculation of Hospital Utilisation Expectancies

2.1 Introduction

In this paper, Hospital Utilisation Expectancies (HUEs), a variation of health expectancies, are used to examine hospital use in New Zealand across the lifecourse. This is compared with life expectancy to identify if compression or expansion of morbidity is occurring in New Zealand and if there are any differences by area deprivation. A HUE is "the number of days while still surviving, that a person of a particular age can expect to spend in hospital" (Cheung, Katzenellenbogen *et al.* 2001, p47). Methodologically these expectancies are fairly simple, and use readily available administrative data. The lack of complexity of the method and the simple data requirements makes HUEs readily applicable in many different countries and settings. They extend existing, widely used health expectancy methodologies, substituting self-reported health measures for the number of hospital bed days over a certain period of time.

The same mortality and hospital datasets and hospital filters are used as were used in Paper One, and do not require additional discussion here. Although HUEs directly combine mortality and hospital data to yield a single health measure, record linkage is not necessary. Instead of record linkage, period prevalence life tables are used to combine age-specific mortality and hospital bed day rates, using a technique developed by Daniel Sullivan (Sullivan 1971). In addition to HUEs, life expectancies are also presented in the next section. It is assumed that the reader is familiar with the calculation of standard life table functions such as life expectancies (Siegel and Swanson 2004).

This section discusses Sullivan's method, the basis behind of a range of life table based health measures. The place of HUEs within this family of health measures is discussed, as well as the relevance of using HUEs in a compression of morbidity context. The calculation of the HUEs presented in Section 3 is also examined.

2.2 Sullivan's Method

As mortality in developed countries declines and converges at relatively low levels, health research is increasingly concentrating on quality of life (Robine, Romieu *et al.* 1999). In 1971, Sullivan presented a life table method (refered to as period prevalence life tables), where life expectancy is split into two different states: the expectation of life free from disability, and the expectation of life with disability (Sullivan 1971). The former is now more commonly referred to as disability free life expectancy, or health expectancy. There are two data requirements needed to calculate period prevalence life tables and associated health expectancies. Age-specific mortality rates for the population of interest (which are used to calculate a period life table) and the prevalence of disability in the population, often obtained from a health survey and population estimates. This framework can be adapted to represent different types of morbidity, by using different types of morbidity data. Thus, there are as many possible health expectancies as there are concepts of health (Robine, Romieu *et al.* 1999). A recent publication of health expectancies in New Zealand used three variants: limitation-free life expectancy, independent life expectancy, and active life expectancy (Ministry of Health and Statistics New Zealand 2009). Health expectancies are an ideal measure for examining the distribution of hospital utilisation over the life course, a prerequisite for identifying the nature of the health transition (compression, expansion or dynamic equilibrium) occurring for different population groups. HUEs are an adaptation of traditional health expectancies. Traditional health expectancies use health survey data to obtain the period prevalence of a particular health complaint in the population. HUEs use administrative hospital data to obtain the period prevalence of hospitalisation in the population.

2.3 Previous Research using Hospital Utilisation Expectancies

HUEs are a method generated by Dr. Jit Cheung and colleagues at the Population Studies Centre, University of Waikato, New Zealand (Cheung 1999). In 2001 Cheung and colleagues published an article in the *Australian Health Review*, examining regional trends in HUEs in New Zealand from 1980 to 1998, and introducing HUEs as a measure of population health (Cheung, Katzenellenbogen *et al.* 2001). In 2006, a discussion paper elaborated on this research, examining regional trends in HUEs in New Zealand from 1986

to 2006, with accompanying health information such as regional smoking rates and sickness/benefit claim rates (Pool, Baxendine *et al.* 2006). The latest publication to use HUEs is a monograph on regional hospital utilisation in New Zealand over a 14 year period from 1988 to 2001, with some limited analyses between 2002 and 2006 (Pool, Baxendine *et al.* 2009). This monograph examines hospital and mortality trends in far greater depth than the previous publications using HUEs. Hospital use and mortality trends by ethnicity are examined in detail, and as such are not reported here. However, the analysis is confined to regional patterns and does not consider area deprivation. This paper builds on the analysis conducted in the monograph by covering a much longer time period (1974 to 2006) in the national analysis, and by examining trends by area deprivation from 1991 to 2006.

2.4 Hospital Utilisation Expectancies and the Compression/Expansion of Morbidity Debate

In the previous paper the relevance of (heavily filtered) hospital utilisation statistics as a proxy for morbidity was discussed. However it was considered ideal to use a measure of health that incorporates both morbidity and mortality aspects. HUEs are calculated from both mortality and morbidity information. A period life table is generated using age-specific death rates, and the period prevalence of hospital utilisation included to yield expected time spent in and outside of hospital over a lifetime. This makes HUEs a very powerful and versatile measure of population health.

HUEs can be calculated at birth (HUE_0) and at any subsequent age. It can therefore be seen if hospital use is concentrated into older ages or spread over the life course. Concentration of hospital use into older ages and/or a decline in HUE_0 suggests compression of morbidity. Expansion is suggested by hospital use spread over the life course and no reduction in HUE_0. Trends in the distribution of hospital use over the lifecourse can be examined over time, by gender, by cause and by area deprivation. This will address the question: is compression of morbidity occurring uniformly across deprivation deciles in New Zealand? Or is there 'epidemiologic polarisation', a combination of health transitions occurring within the same country?

Combined mortality and morbidity life table measures such as health expectancies and disability free life expectancy have been used extensively to address the

compression/expansion of morbidity hypothesis. HUEs are a new measure in this family, which have not previously been used in this debate. Life table based methods are perfect for the purpose of quantifying the compression/expansion of morbidity debate. Regardless of changes in life expectancy, life table methods present a quantitative measure of future morbidity for individuals at any given age. If the quantity of expected future morbidity (hospital bed days) at a given age declines over the time period, yet life expectancy stays the same or increases, compression is occurring. This is of course a simplistic example, relative increases and declines in HUEs will need to be contrasted empirically with changes in life expectancy.

2.5 Calculation of Hospital Utilisation Expectancies

The HUEs presented in this paper use the period prevalence life table method. There are three steps to calculating HUEs using this method:

Step one. Age-specific death rates (presented in Paper One) are used to calculate a period life table. It is assumed that the reader is familiar with the calculation of period life tables, which are not discussed in detail here. See Hinde, 1998, Sections 4 and 5 or Shryock and Siegel, 1976, Section 15 for life table methodology (Hinde 1998; Shryock, Siegel *et al.* 1976).

Step Two. Next, age-specific hospital bed day rates are divided by 365, to give the average fraction of the year individuals in that age group spend in hospital. This can be thought of as the proportion of the population in hospital at any given time and is the equivalent of the proportion of the population with a disability, an indicator used in disability free life expectancies.

Step Three. The proportion in hospital is then multiplied by the number of person years lived at each age (L_x) to give the number of person years lived *in hospital* at each age (L^h_x). The life expectancy *in hospital* can then be calculated (e^h_x), in the same way as life expectancy was calculated in the mortality life table (e_x). This figure is then divided by 365 to change the unit from years to days. This is done because people tend to spend only a small fraction of a year in hospital over their lifetime, so presenting the figure in days is easier to interpret.

Abridged life tables are used, with five year age groups. Population data at Census Area Unit (CAU) level is only available in five year age groups, due to confidentiality

concerns. The maximum age group available is 85 years and over, and the first age group is 0 to 4. Certain assumptions are made in the calculation of a period life table, including the assumption that deaths within an age group occur evenly across the time period. Thus in a 5 year age group, it is assumed that individuals that died in this age group lived (on average) 2.5 years. This assumption does not hold for infants, in abridged life tables infants are usually a separate age group due to the uneven distribution of deaths in the first year of life. However population data for infants was not available at a CAU level. It is commonly assumed that in developed countries, babies that die in the first year of life live on average one tenth of a year ($_1a_0 = 0.1$). Given that most deaths that occur at ages 0 to 4 occur in the first few months of life, the average number of years lived by those who died between ages 0 and 4 is assumed to be one year ($_5a_0 = 1$). This is only a rough approximation, but given the low infant mortality rate in New Zealand, the choice of a_x for the lowest age group will have little effect on either life expectancy or HUE.

HUEs are calculated at a national level from 1974 to 2006, and at an area deprivation level from 1991 to 2006. These calculations are disaggregated by year, gender, age group, cause of hospital use, and (for the deprivation analyses) deprivation decile.

2.6 Conclusion

HUEs are a natural progression of the health expectancy methodology, and apply a different type of morbidity data to this methodology. HUEs have only been calculated for New Zealand data in a handful of studies, and have never been calculated by area deprivation, or interpreted with reference to the compression of morbidity hypothesis. The next section presents the results from the HUE analysis, at both a national level and by area deprivation. These results are contrasted with both previous HUE and health expectancies research in the discussion of these results, which is found in Section 4.

SECTION 3 – RESULTS: Evidence for/against Compression of Morbidity from Hospital Utilisation Expectancy Analysis

3.1 Introduction

This section examines trends in Hospital Utilisation Expectancies (HUEs) and life expectancy in New Zealand. National results are presented first, describing trends from 1974 to 2006. These results are examined to highlight the long-term trends in life expectancy and HUEs, and to identify if there is evidence for compression of morbidity at the national level. Results by area deprivation are then presented, for the time period 1991 to 2006.

Due to fluctuations in both life expectancy and HUEs from year to year, three year moving averages are used instead of single years, to smooth out the data. Three year averages also boost the number of deaths for the deprivation analysis, which goes down to smaller levels of aggregation than the national analysis. The missing 2001 mortality data are interpolated from the 2000 and 2002 mortality data.

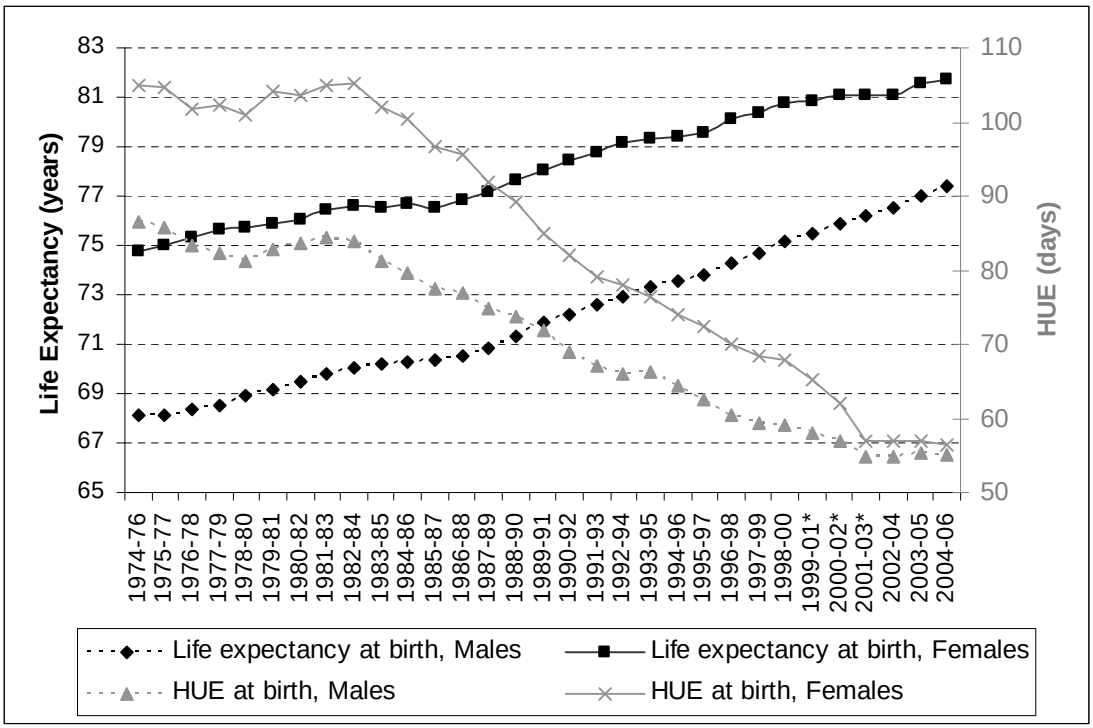
3.2 National Life Expectancy and HUEs (1974-2006)

As with the mortality and hospital bed day rates presented in Paper One, it is essential to examine national patterns in addition to patterns by area deprivation. This section presents trends in national life expectancy and HUEs from 1974 to 2006. Results are presented over time, by age, and by cause of hospital use.

3.2.1 National Results, Life Expectancy and HUEs Over Time, 1974-2006

Figure 3.2.1 presents life expectancy and HUE at birth for New Zealand from 1974 to 2006. For life expectancy (the black lines) refer to the left Y axis, for HUE (the grey lines) refer to the right Y axis. The point of intercept between life expectancy and HUE is arbitrary.

Figure 3.2.1: Life Expectancy and HUE at birth for Males and Females between 1974 and 2006, Three-Year Moving Averages.



* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

The main findings illustrated in Figure 3.2.1 are that life expectancy increased over the time period, HUEs decreased, and there was convergence between male and female expectancies. Life expectancy rose consistently over the time period, from 68.1 to 77.4 years for males, and from 74.8 to 81.8 years for females. Male life expectancy rose faster than female life expectancy, resulting in some convergence between male and female life expectancy. In 1974-76 female life expectancy was 6.7 years higher than male life expectancy, but by 2004-06 this had decreased to 4.4 years. With the exception of the early 1980s, HUE decreased over the time period. Females had higher HUEs than males, and female HUEs almost halved, from 104.8 to 56.4 days. Male HUEs declined from 86.6 to 55.1 days. Females experienced only slightly higher hospital bed day rates than males (as discussed in Paper One), but substantially higher lifetime hospital use (HUEs) due to higher life expectancy. Male and female HUEs converged suddenly in 2001-03, remaining stable until the end of the time period. The gap between male and female HUEs was less than a day in the mid 2000s, compared to a difference of nearly 20 days from the mid 1970s to the mid 1980s. The sharp convergence of male and female HUEs in 2001 occurred at the same time as a change in the ICD version used in the hospital data, from ICD9CM to ICD10. ICD10 was markedly different to the preceding ICD versions, and

while certain types of cases were excluded from the hospital dataset to improve consistency over time, the effects of coding changes are still present in the data.

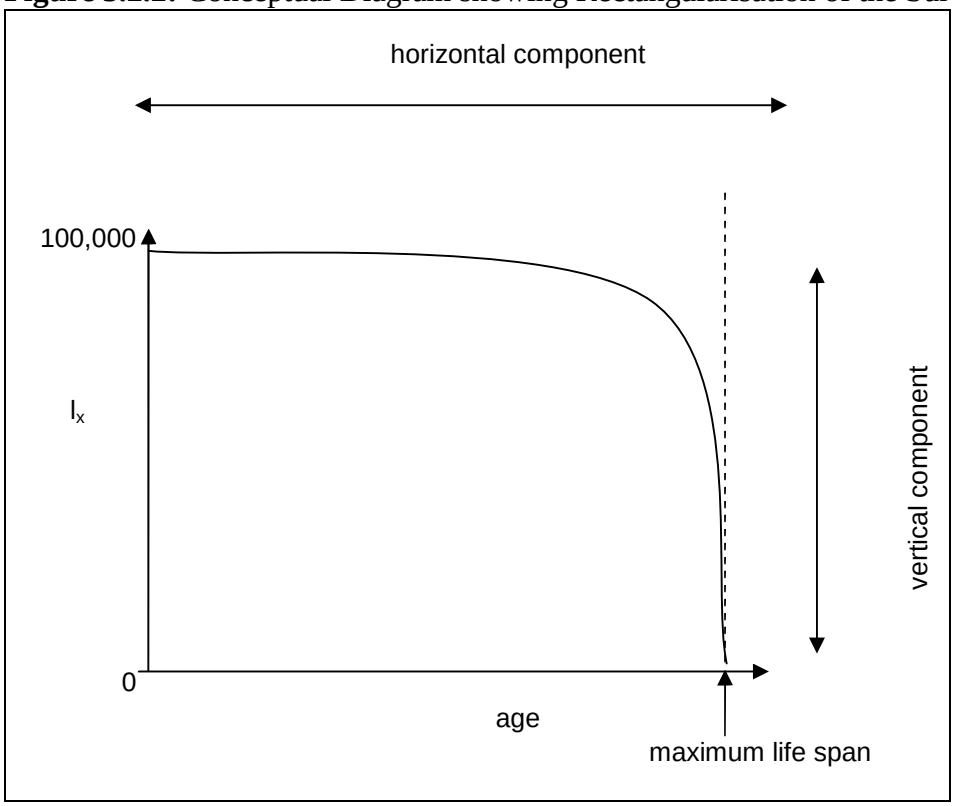
Life expectancy is increasing, yet lifetime hospital use (HUE_0) is decreasing. In other words the period 'at risk' of hospitalisation is increasing, yet the number of days spent in hospital over the lifecourse is decreasing. Figure 3.2.1 therefore provides evidence of compression of morbidity in New Zealand between 1974 and 2006, when time spent in hospital is used as a proxy for morbidity.

One prerequisite for compression of morbidity to occur is argued to be the rectangularisation of the survival curve (the l_x function in the life table) resulting from a reduction in premature mortality and a limit to human life span (Fries 1980; 1983). If there is a fixed limit to life span, then as life expectancy approaches this limit, further improvements in mortality will be the result of reductions in premature mortality rather than extensions in life span. This process implies a compression of mortality into a shorter age range. The result would be high survival levels until ages close to the theoretical maximum life span, followed by a rapid decline in the proportion surviving, hence the rectangularisation of the survival curve.

Figure 3.2.2 conceptualises the rectangularisation of the survival curve. The y axis on a survival curve represents the number of individuals surviving out of the radix[14] and the x axis represents age. The horizontal component represents the reduction of premature mortality, and the vertical component is the result of a fixed limit to life span.

[14] The radix is the number of people alive at exact age 0 in a period life table. This number is arbitrary, but is usually 100,000.

Figure 3.2.2: Conceptual Diagram showing Rectangularisation of the Survival Curve



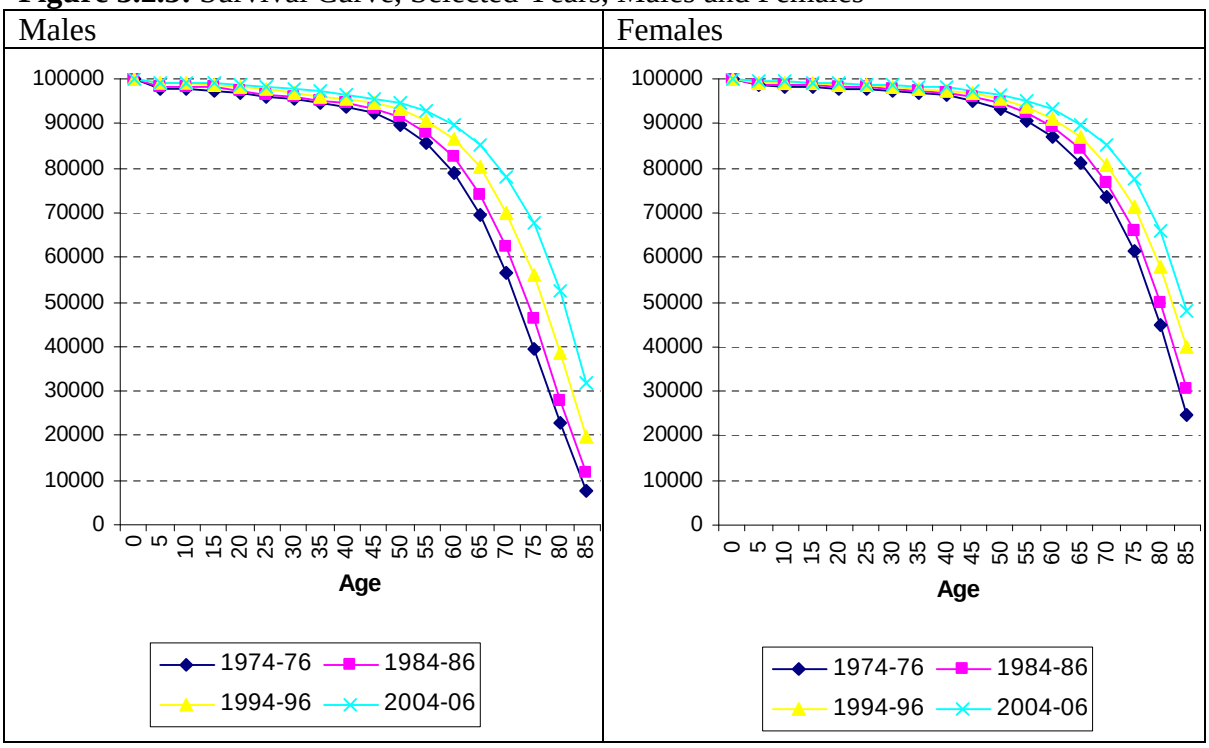
Source: concept of rectangularisation obtained from various sources

Figure 3.2.3 presents the survival curve from an abridged period life table for selected years. The radix of the life table is 100,000. As expected given the climate of declining mortality, in each subsequent time period there is a higher proportion surviving at every age group, for both males and females. This is particularly noticeable at older ages, at age 85 the percentage still alive increased from 8 to 32 for males, and from 25 to 48 percent for females between 1974-76 and 2004-06. Therein lies an inherent flaw in interpreting the survival curve for life tables with a maximum age group. 85+ is the highest age group for which population denominator data are available at a CAU level in New Zealand, so the life tables used in the area deprivation analysis are limited to a maximum age group of 85+. While the UN population estimates used for the national analysis go up to age 100+, the population is given in thousands. The population for age groups over 85 years in New Zealand is small[15], so it is unreliable to use population estimates that are rounded to the nearest thousand, which for several of the highest age groups are 0. Therefore the survival curve can only be plotted up to age 85, at which age nearly half of

[15] While the 2004-06 period life table estimates that nearly half of females and nearly a third of males survive to age 85+, this is the survivorship that would be observed *if* the age-specific mortality rates of 2004-06 remained the same for at least 85 years. Mortality rates in previous decades were much higher, so only a small proportion of previous cohorts survived to age 85. This has resulted in a small population aged 85+ in New Zealand.

females and a third of males in the 2004-06 period life table are still alive. This makes it difficult to assess if rectangularisation is occurring, particularly for females. The survival curve at ages over 85 years is unknown.

Figure 3.2.3: Survival Curve, Selected Years, Males and Females



However, it can be seen from Figure 3.2.3 that premature mortality is reducing. That is, the horizontal component of rectangularisation seems to be occurring. Whether the vertical component of rectangularisation is occurring (denoting the presence of a fixed limit to life span) is harder to ascertain due to the maximum age of 85. However increases in life expectancy at birth do not appear to be slowing down (refer back to Figure 3.2.1) suggesting that rectangularisation of the survival curve is *not* occurring. This analysis provides mixed findings, leading to tentative but incomplete support for rectangularisation of the survival curve, considered by some a necessary prerequisite for compression of morbidity (Fries 1980; 1983).

3.2.2 National Results, Life Expectancy and HUEs by Age, 1974-2006

Age-specific trends in national life expectancy and HUEs followed much the same pattern as for life expectancy and HUEs at birth (shown in Figure 3.2.1), and as such are not presented graphically here. Life expectancy at all ages saw a steady increase over the time period, and HUEs at all ages saw a brief rise in the early 1980s, followed by a sustained

decrease. There was however a difference in the relative increases/decreases by age group, which are shown as ratios in Table 3.2.1. These ratios were calculated by dividing the 2004-06 figure by the 1974-76 figure, for life expectancy and for HUEs. Thus a ratio below 1 indicates a decline over the time period, and a ratio above 1 indicates an increase. A ratio of 0.5 represents a reduction of 50 percent and a ratio of 1.5 represents an increase of 50 percent.

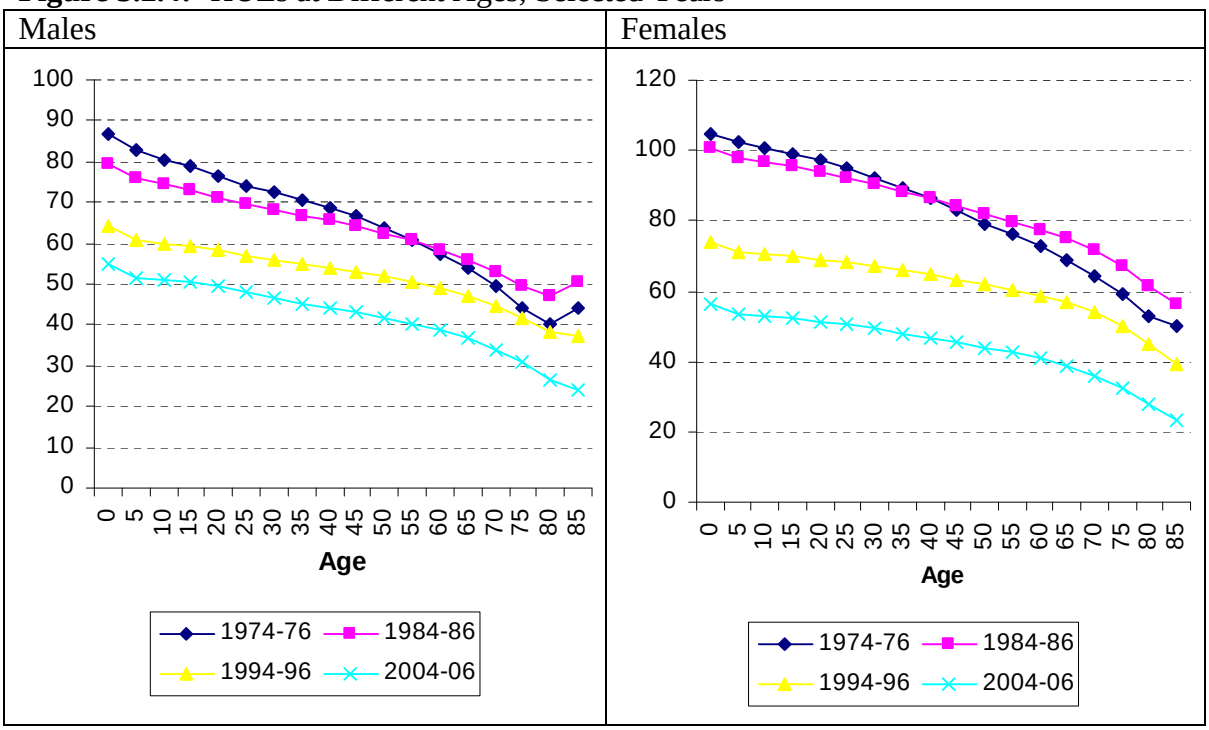
Ratios of life expectancy and HUEs in 2004-06 to 1974-76, show that the relative decrease and increase for HUEs and life expectancy respectively, was greatest at older ages (see Table 3.2.1). Males experienced larger relative increases in life expectancy than females, and females experienced larger relative decreases in HUEs than males (possibly due to the smaller increases in life expectancy for females) The largest relative increase in life expectancy was for males at age 85 (an increase of 82 percent), and the largest relative decrease in HUEs was for females at age 85 (a decrease of 51 percent). However, both life expectancy and HUEs are smallest at older ages: as age increases, both life expectancy and expected future hospital use decrease. So an increase of 82 percent in life expectancy for males at age 85 over the time period translates to an increase of only 2.8 years.

Table 3.2.1: Ratio of 2004-06 to 1974-75 for Life Expectancy and HUE, Selected Ages

		Life Expectancy		HUEs	
		Male	Female	Male	Female
Age	0	1.14	1.09	0.61	0.52
	20	1.16	1.11	0.61	0.51
	40	1.25	1.16	0.60	0.52
	60	1.43	1.25	0.62	0.54
	70	1.56	1.34	0.61	0.51
	80	1.75	1.44	0.56	0.49
	85	1.82	1.43	0.55	0.46

Figure 3.2.4 shows the HUEs at different ages and at four points in time. HUEs did not decrease much between 1974-76 and 1984-86, and even increased at ages over 60 for males and 45 for females. This matches the observed increase in HUEs in the mid 1980s reported previously. In subsequent years HUEs fell consistently at all ages. The different relative declines for males and females that were reported in Table 3.2.1 are easily observable. The smallest relative decrease was for male HUE_0, which fell from 95.7 to 58.3 days between 1974-76 and 2004-06, a decrease of 39 percent, whereas female HUE_{85} fell from 50.0 to 23.0 days, a decrease of 54 percent.

Figure 3.2.4: HUEs at Different Ages, Selected Years

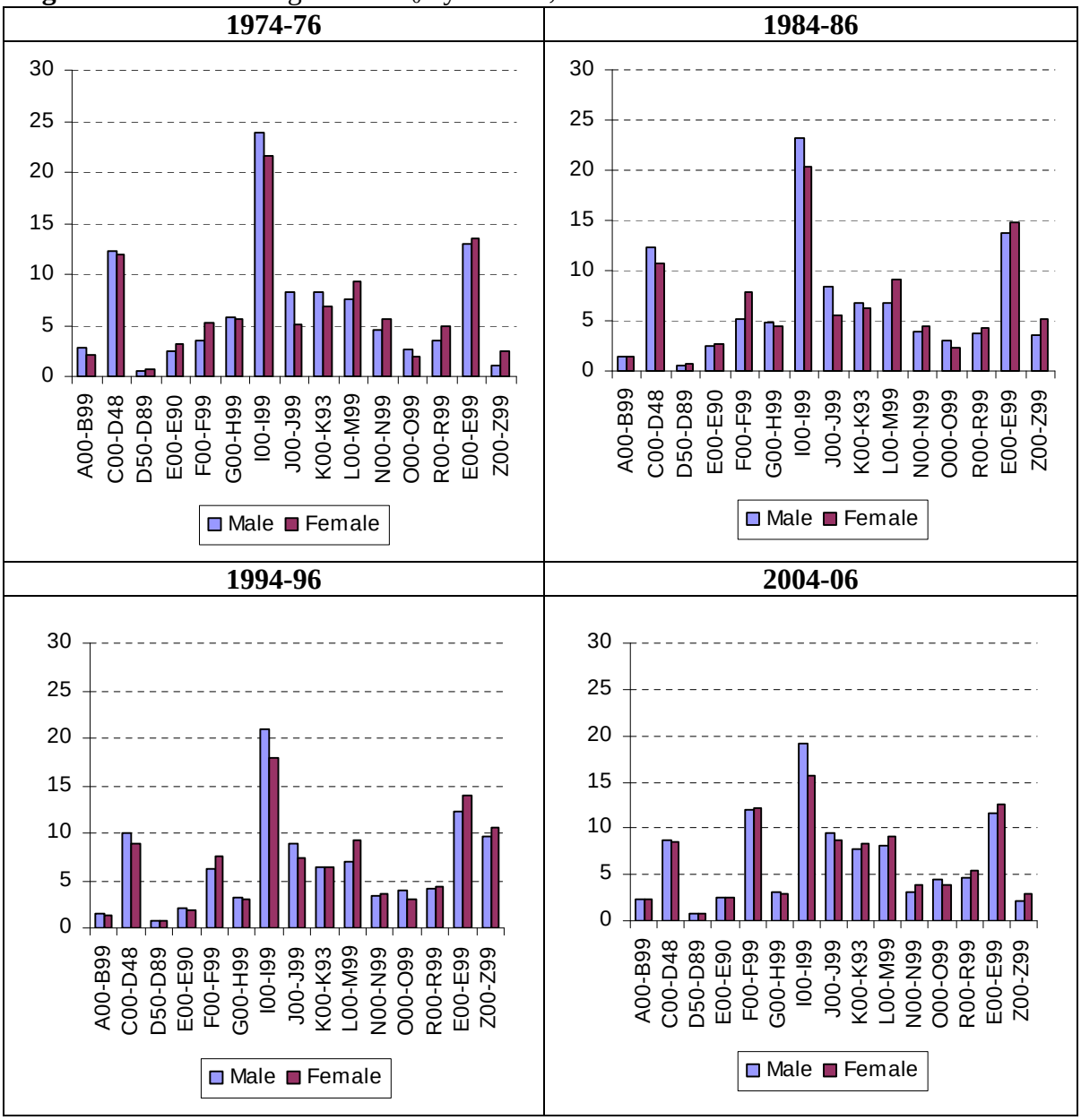


The evidence that life expectancy is increasing fastest at the oldest ages indicates that the vertical aspect of the rectangularisation of the survival curve may not be occurring. However, the evidence that hospital use is decreasing fastest at the oldest ages *despite* life expectancy increasing fastest at these ages, suggests that compression of morbidity *is* occurring. The overall reduction in HUEs at all ages despite increases in life expectancy at birth also denotes compression of morbidity (but not concentration of morbidity into the period prior to death).

3.2.3 National Results, Life Expectancy and HUEs by Cause, 1974-2006

Figure 3.2.5 presents the percentage of HUE_0, which can be attributed to different diagnostic groups. Results are shown for 1974-76, 1984-86, 1994-96 and 2004-06. The bars in each graph add up to 100 percent for males and 100 percent for females.

Figure 3.2.5: Percentage of HUE_0 by Cause*, Selected Years



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

The most noticeable finding from Figure 3.2.5 is the lack of any compositional change in HUE0 by cause and sex over the time period, with two exceptions. The first

exception is the increase and decline of Z00-Z99 (factors influencing health status and contact with health services). The HUEs for this diagnostic group (not shown here) increase from 3.1 and 3.7 to 8.7 and 12.8 days for males and females respectively between 1992-94 and 1998-2000, before dropping to less than 2 days in 2001-2003. The decrease in 2001-2003 corresponds with the change to ICD version 10, but there is no corresponding ICD version change at the start of the increase in the early 1990s. This diagnostic group represents (amongst other things) medical examinations, and screening for various diseases. A large proportion of this diagnostic group, such as codes relating to obstetrics, well babies, and boarders, are excluded from the analysis. The high Z00-Z99 HUEs in the 1990s are almost certainly artefactual and the result of inconsistencies in the data. It would be unrealistic for the *demand* for Z00-Z99 related hospital use to alter so rapidly. However this needs further investigation.

The second main change is the increasing importance over time of mental and behavioural disorders (F00-F99) as a component of HUE_0. Mental and behavioural disorders make up 3.5 and 5.2 percent of HUE_0 for males and females respectively in 1974-76, but this increases to about 12 percent for males and females in 2004-06.

Apart from Z00-Z99 and F00-F99, there is very little change in the causal composition of HUE_0 over the time period. The main cause of hospital use, circulatory conditions (I00-I99) decrease slightly from 21.7 to 15.8 percent of HUE_0 between 1974-76 and 2004-06 for males, with a similar decrease for females. Neoplasms (C00-D48) do not change in importance over the time period. Respiratory (J00-J99), digestive (K00-K93) and diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissues (L00-M99) all show moderate increases in importance, especially for women.

3.2.4 Conclusions from National Analysis of Life Expectancy and HUEs, 1974-2006

Over the period 1974-2006, life expectancy steadily increased and was higher for females than males, although there was considerable convergence. Over the same period, hospital utilisation decreased (with a blip in the early 1980s) and was higher for females than males, although there was almost complete convergence by 2001-03. The greatest relative increases in life expectancy were seen at the oldest ages (especially for males) and the greatest relative decreases in hospital utilisation were also seen at the oldest ages (especially for females). Survival curves for the period indicate that premature mortality is being reduced, but are insufficient to determine if there is evidence of a fixed limit to life

span due to a maximum age group of 85+. Increasing life expectancy over the time period however suggests that rectangularisation is not occurring.

These national trends are now used as a basis to examine variation in life expectancy and HUEs by area deprivation, from 1991 to 2006.

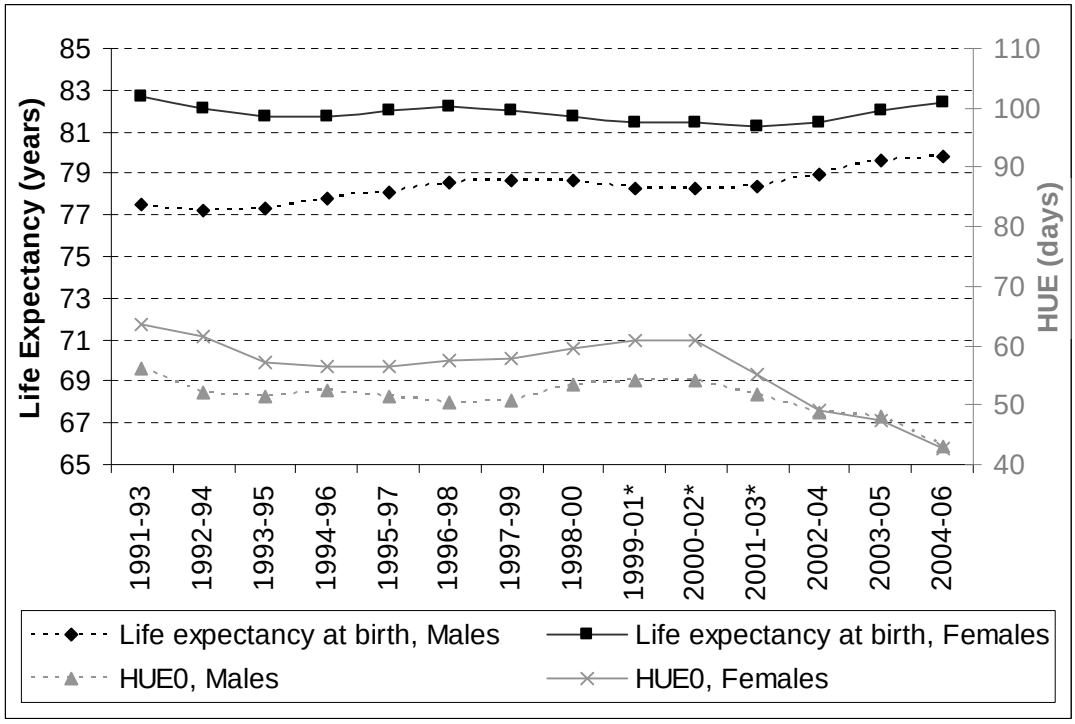
3.3 Life Expectancy and HUEs by Area Deprivation (1991-2006)

Life expectancy and HUEs are examined by area deprivation over a shorter period than for the national analysis. Again, results are presented in three ways, although there is naturally some overlap. Life expectancy and HUEs are presented over time, by age, and by cause of hospital use.

3.3.1 Area Deprivation Results, Life Expectancy and HUEs Over Time, 1991-2006

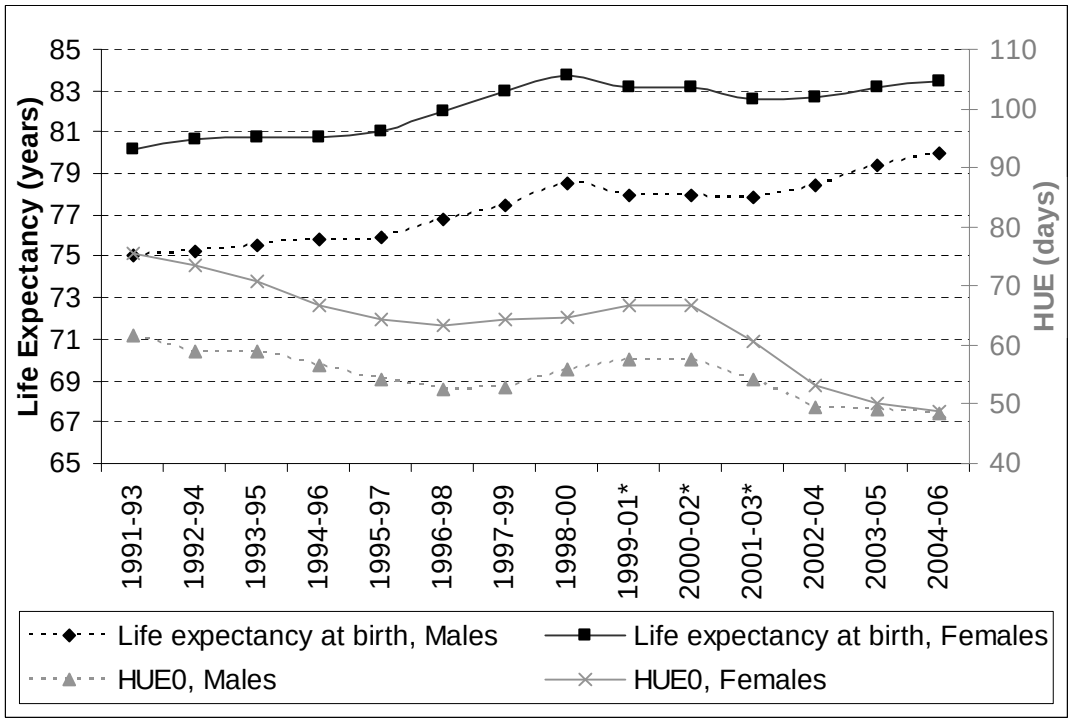
Figures 3.3.1 to 3.3.4 show the patterns in life expectancy and HUE at birth from 1991 to 2006, for deprivation deciles 1, 2, 9 and 10. Three-year moving averages are used to boost numbers and reduce year-to-year fluctuation. The same scales for life expectancy and HUEs are used for all four graphs, to make levels of life expectancy and hospital use directly comparable between the deciles. Decile 1 represents the least deprived tenth of areas, and decile 10 represents the most deprived tenth of areas, as measured by NZ*dep*91, 96 and 01.

Figure 3.3.1: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 1 between 1991 and 2006, Three-Year Moving Averages



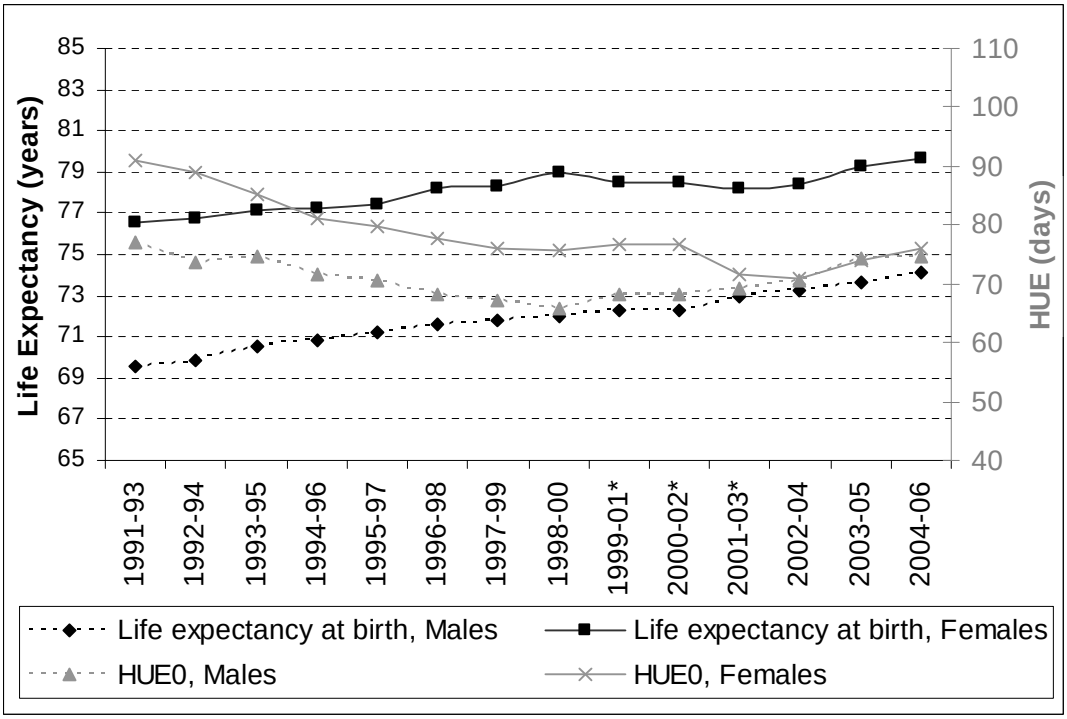
* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

Figure 3.3.2: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 2 between 1991 and 2006, Three-Year Moving Averages



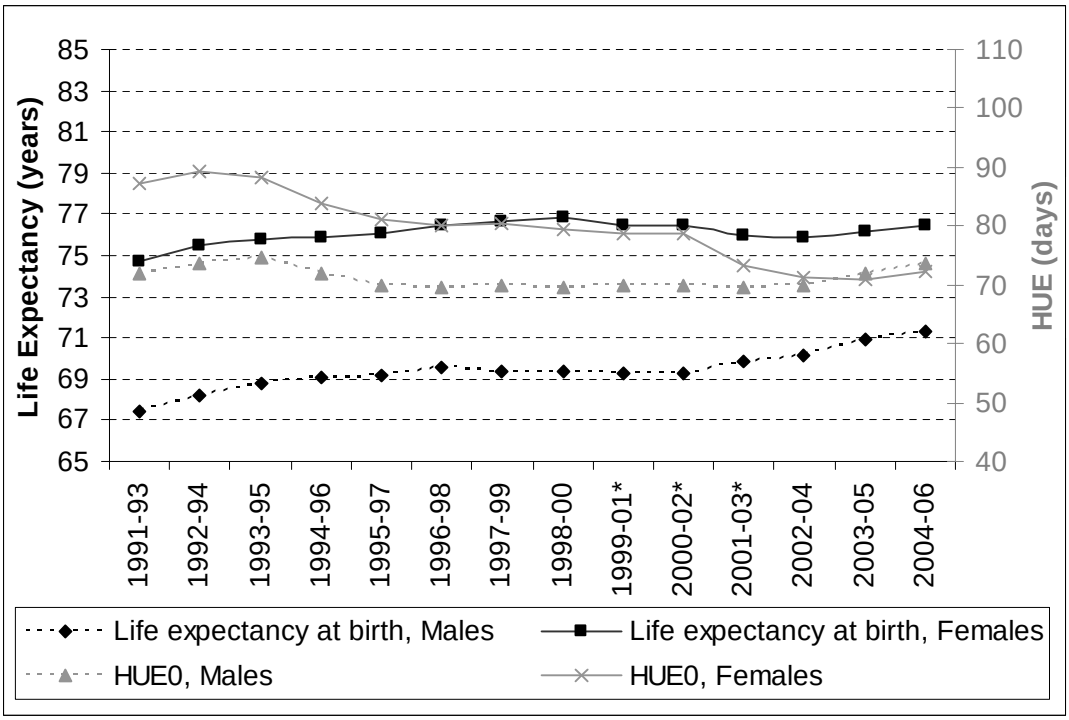
* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

Figure 3.3.3: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 9 between 1991 and 2006, Three-Year Moving Averages



* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

Figure 3.3.4: Life Expectancy and HUE at birth for Males and Females in Deprivation Decile 10 between 1991 and 2006, Three-Year Moving Averages



* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

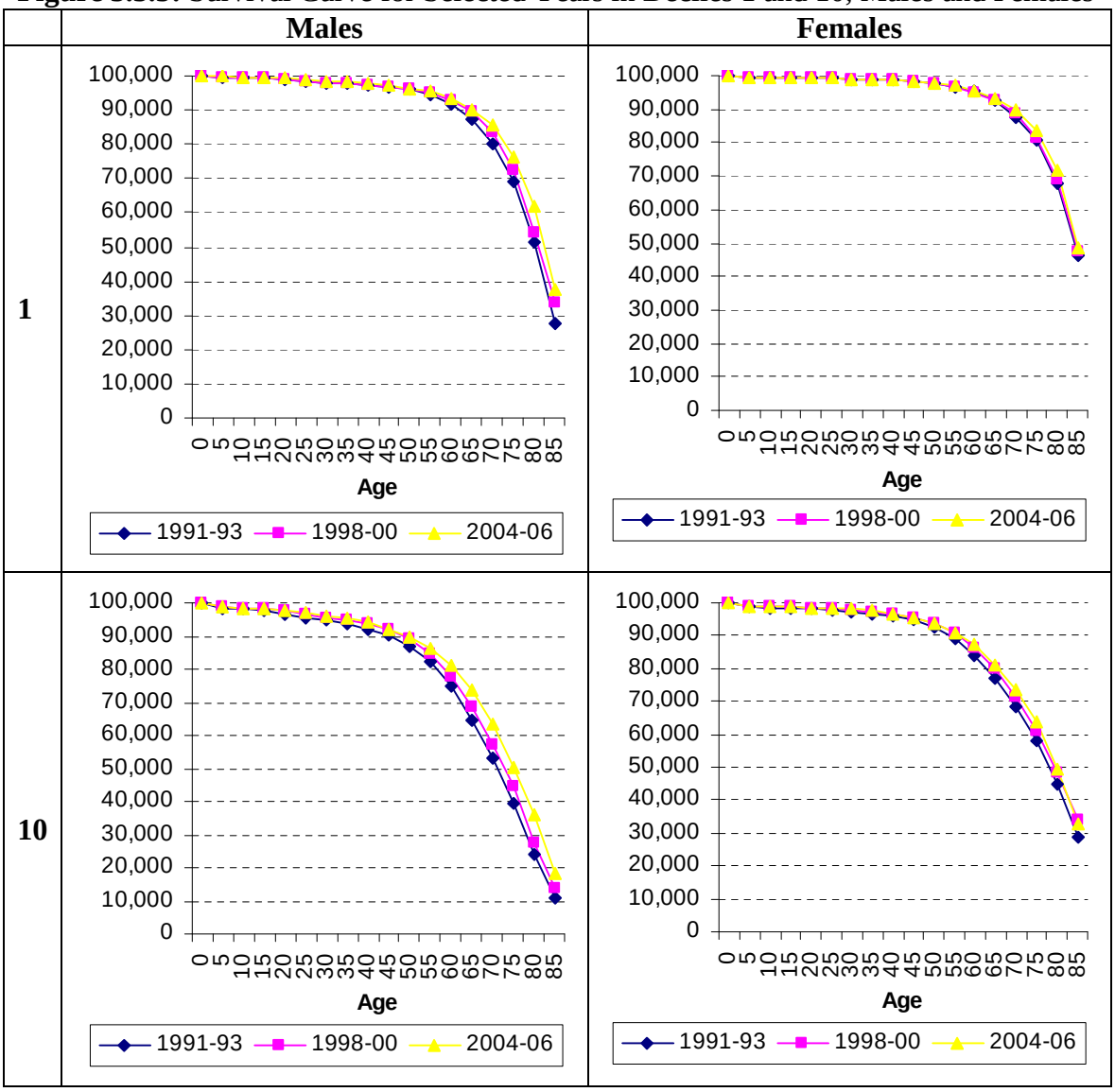
Life expectancy is highest in deciles 1 and 2, and lowest in decile 10. In contrast, HUEs are lowest in decile 1, and highest in deciles 9 and 10. This is remarkable when it is considered that HUEs represent *lifetime* hospital utilisation. People in the least deprived deciles live for longer, yet spend less time in hospital than people in more deprived deciles. Even though people in more deprived deciles have shorter lives, they spend more time in hospital. The pattern of decreasing life expectancy and increasing HUEs with increasing deprivation is also observed in deciles 3 to 8 (not shown here). Despite the different levels of life expectancy and HUEs in different deciles, the same patterns of increasing life expectancy and decreasing HUEs over time is seen in all deciles. The gap between male and female life expectancy in the least deprived deciles is smaller than in the most deprived deciles, although convergence between male and female life expectancy is seen over time in all deciles. Convergence in HUEs for males and females, to almost the same level in 2001-03 is also seen in all deciles.

Figure 3.3.5 examines the survival curve for males and females in deciles 1 and 10. The proportion of females surviving in decile 1 at all time periods remains very high until well into middle-old age, and only drops below 98 percent at age 55. In contrast, the proportion of females surviving in decile 10 begins to decline at a much younger age, dropping below 98 percent at age 35 in 2004-06. By age 85, only 33 percent of females in decile 10 were still alive in the 2004-06 life table, compared to 49 percent of females in decile 1. There is a similar difference between decile 1 and decile 10 for males. The number of males surviving in decile 1 stays very close to 100,000 until well into middle age, whereas the number of males surviving in decile 10 drops off at a much younger age. Males in decile 10 saw the most improvement in the survival curve over the time period, and females in decile 1 the least. However this may be in part self-selecting because males in decile 10 have the most *potential* for reductions in premature mortality. Given the high levels of survival into middle-old age, females in the least deprived deciles may be approaching a limit to mortality reductions. This is supported by the stagnation of female life expectancy in decile 1, shown in Figure 3.3.1. Males in decile 1, and both males and females in all other deciles saw an increase in life expectancy over the time period.

In assessing rectangularisation of the survival curves it is apparent that survival curves are more ‘rectangular’ for less deprived areas and for females, and less ‘rectangular’ for more deprived areas and for males. The vertical component of rectangularisation requires a fixed limit to life span, and as average life expectancy approaches this limit, the majority of deaths occur in a short age range, close to maximum life span. While the

hypothetical fixed limit to life span may vary by sex, it can be assumed that maximum life span for each sex is the same for different deprivation deciles. However, life expectancy varies substantially by deprivation decile, female life expectancy in 2004-06 ranged from 82.4 years in decile 1 to 76.4 years in decile 10. Life expectancy in more deprived deciles is further from the fixed limit to life span, so rectangularisation must be further from completion in more deprived deciles.

Figure 3.3.5: Survival Curve for Selected Years in Deciles 1 and 10, Males and Females

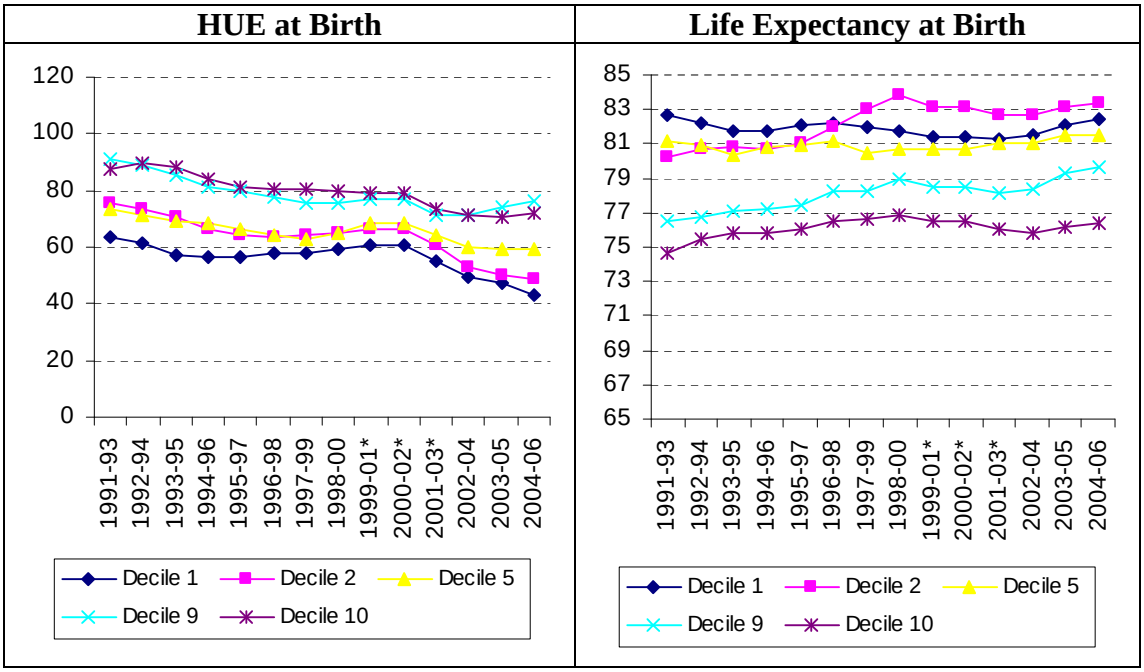


3.3.2 Area Deprivation Results, Life Expectancy and HUEs by Age, 1991-2006

This section examines trends in HUE by age. Figure 3.3.6 shows that at birth, there is a clear distinction in HUE_0 for females between the least and the most deprived deciles. HUE_0 for deciles 1 and 2 is considerably lower than HUE_0 for deciles 9 and 10. There is also a relatively consistent pattern for female life expectancy at birth, with more deprived

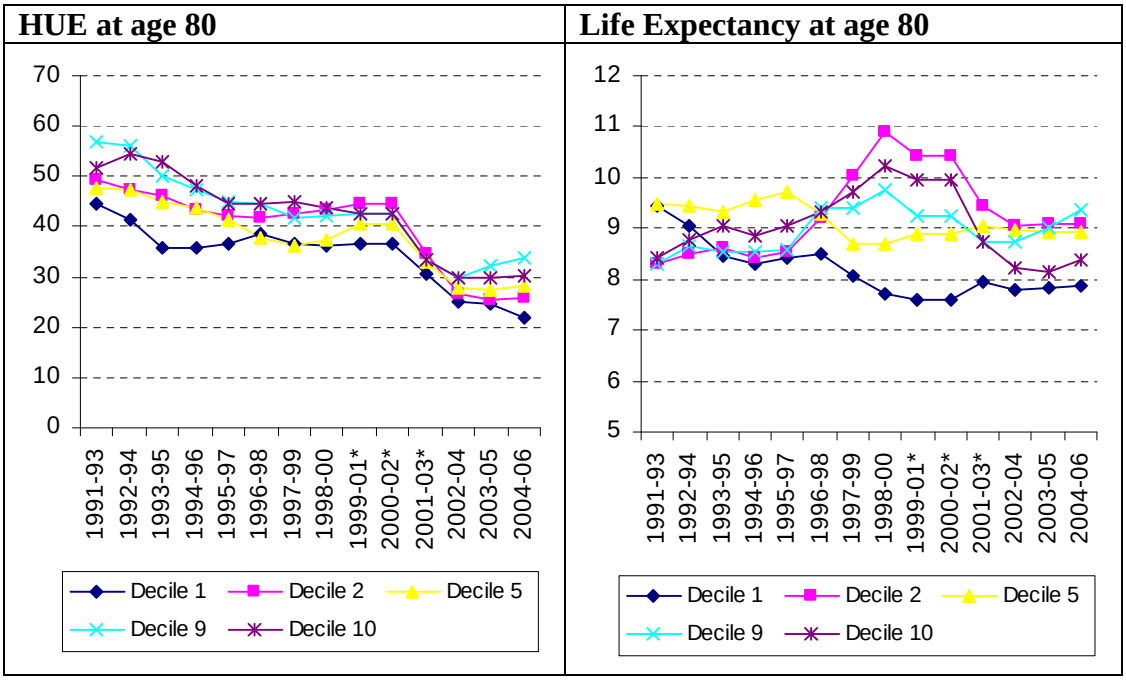
deciles having lower life expectancy and less deprived deciles having higher life expectancy. However, when the same graphs are examined for age 80, the pattern is not so clear (see Figure 3.3.7). Decile 2 (the second least deprived decile) has one of the *highest* HUEs, and decile 1 (the least deprived decile) has the *lowest* life expectancy for several years for age 80. Similar trends were also seen for males. HUEs and life expectancy for different deciles converge at older ages, the same trend seen for mortality and hospital bed day rates in Paper One. Figure 3.3.8 illustrates this clearly. At younger ages there are large differences in HUEs (and, to a lesser degree, life expectancy) by decile, with less deprived deciles having lower HUEs and higher life expectancy. But as age increases, the difference in HUEs between the deciles narrows. The same trend is true of males.

Figure 3.3.6: HUE_0 and Life Expectancy at Birth for Females, Selected Deprivation Deciles



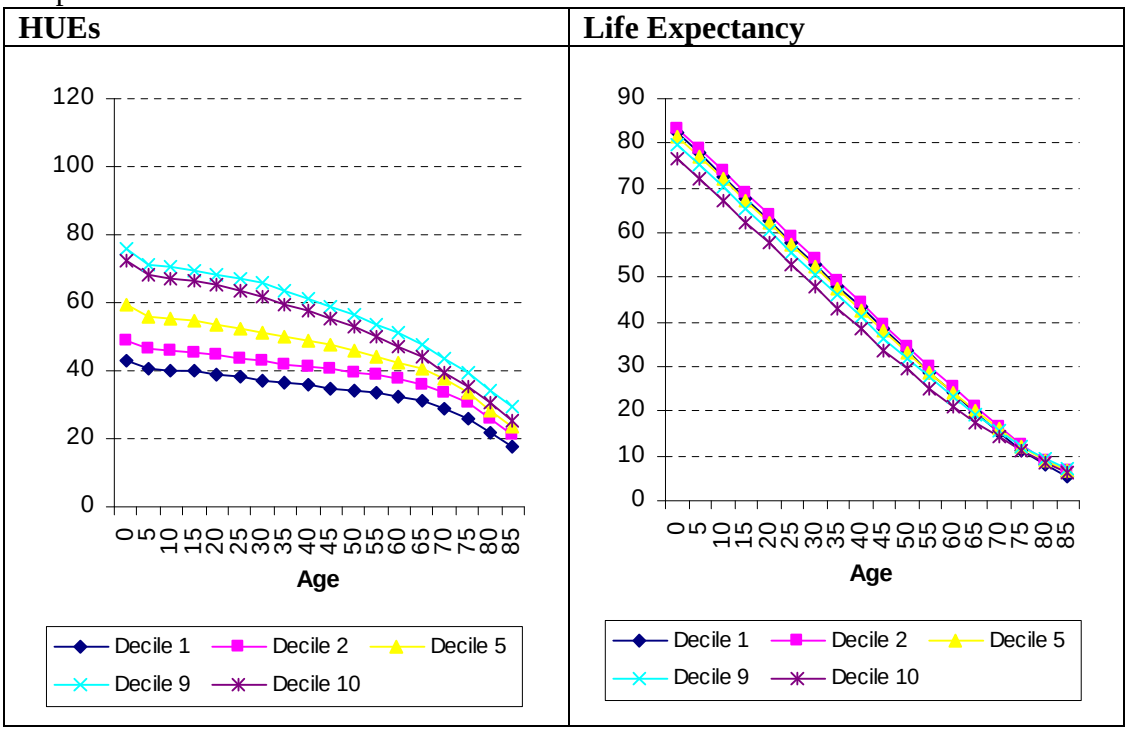
* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

Figure 3.3.7: HUE_{80} and Life Expectancy at age 80 for Females, Selected Deprivation Deciles



* 2001 mortality data are missing. 2001 mortality data for the three-year moving averages is linearly interpolated from 2000 and 2002 mortality data.

Figure 3.3.8: Female HUEs and Life Expectancy at Different Ages in 2004-06, by Deprivation Decile



It is also useful to examine ratios for HUEs and life expectancy, to see for which deciles and age groups change is occurring most rapidly between 1991-93 and 2004-06. The national analysis of ratios between 1974-76 and 2004-06 found that HUEs were

decreasing most rapidly for females in the oldest age group, and life expectancy was increasing most rapidly for males in the oldest age group (Table 3.2.1). See Table 3.3.1 for ratios of 2004-06 to 1991-93 for life expectancy and HUEs, by deprivation.

HUEs are decreasing more rapidly in less deprived deciles than in more deprived deciles. The largest relative decline is seen for females aged 85 in decile 1, with a ratio of 0.42 (indicating a decline in the HUE of 58 percent). The only groups not to see a decline in HUE over the time period were young males in decile 10, for whom there was a slight increase in the HUE (ratio of 1.03 and 1.02 for ages 0 and 20 respectively).

The pattern is less clear for life expectancy. The national analysis found relative life expectancy increases to be greatest for males in the oldest age group, and found life expectancy to increase for all ages. However, for the oldest age group in several deciles, life expectancy *decreased* over the period, and life expectancy over the period for females in decile 1 decreased slightly at every age group from age 25 up, and experienced no change in life expectancy for age groups below 25. The largest relative decrease in life expectancy was for females aged 85 in decile 1, this group had a ratio of 0.71 due to a decline in life expectancy from 8 years to 5 over the time period. This is a decline in life expectancy of 29 percent. These decreases have been highlighted in bold in Table 3.3.1. For both males and females aged 85 in deciles 1, 5 and 10, life expectancy also decreased; there is clearly no deprivation gradient to this finding, although the decline in life expectancy does appear to be most severe for females in decile 1. This matches the pattern found when examining the survival curves in Figure 3.3.5. There appeared to be almost no change in the survival curve for females in decile 1 over the time period, and it was considered that mortality rates for females in decile 1 may have reached their minimum level, that further declines in mortality were not possible. This however does not explain the declines and stagnation in life expectancy seen at the oldest ages in some of the more deprived deciles.

Table 3.3.1: Ratio of 2004-06 to 1991-93 Life Expectancy and HUE by Deprivation Decile, Selected Ages

		HUEs									
	Age	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Male	**0**	0.77	0.78	0.81	0.77	0.95	0.75	0.94	0.79	0.97	**1.03**
	20	0.74	0.77	0.79	0.77	0.92	0.74	0.92	0.78	0.97	**1.02**
	40	0.72	0.75	0.76	0.75	0.88	0.71	0.89	0.75	0.92	0.96
	60	0.70	0.75	0.73	0.74	0.84	0.68	0.86	0.73	0.84	0.87
	70	0.68	0.74	0.70	0.74	0.77	0.66	0.82	0.69	0.77	0.80
	80	0.63	0.67	0.59	0.77	0.66	0.56	0.72	0.58	0.67	0.65
	85	0.54	0.57	0.48	0.62	0.59	0.48	0.65	0.56	0.65	0.58
Female	**0**	0.67	0.65	0.76	0.69	0.81	0.67	0.85	0.67	0.84	0.83
	20	0.65	0.63	0.74	0.67	0.78	0.66	0.84	0.66	0.81	0.80
	40	0.63	0.61	0.73	0.66	0.75	0.64	0.81	0.64	0.79	0.78
	60	0.60	0.60	0.70	0.64	0.71	0.61	0.77	0.61	0.73	0.73
	70	0.56	0.58	0.67	0.64	0.68	0.58	0.74	0.58	0.67	0.66
	80	0.49	0.52	0.62	0.61	0.59	0.54	0.70	0.51	0.60	0.59
	85	0.42	0.48	0.56	0.58	0.54	0.47	0.63	0.47	0.53	0.54
		Life Expectancy									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Male	**0**	1.03	1.07	1.06	1.08	1.03	1.06	1.04	1.06	1.07	1.06
	20	1.03	1.09	1.08	1.10	1.04	1.07	1.05	1.08	1.08	1.07
	40	1.05	1.12	1.10	1.15	1.05	1.10	1.06	1.11	1.11	1.09
	60	1.09	1.22	1.17	1.27	1.09	1.17	1.12	1.19	1.20	1.14
	70	1.09	1.28	1.19	1.37	1.08	1.18	1.16	1.24	1.25	1.13
	80	1.03	1.32	1.14	1.54	1.01	1.13	1.13	1.21	1.30	1.04
	85	**0.90**	1.26	1.05	1.34	**0.94**	1.02	**0.98**	1.22	1.36	**0.94**
Female	**0**	1.00	1.04	1.04	1.05	1.01	1.03	1.02	1.03	1.04	1.02
	20	1.00	1.05	1.05	1.07	1.00	1.04	1.03	1.04	1.05	1.03
	40	**0.99**	1.06	1.07	1.10	1.00	1.06	1.04	1.05	1.07	1.04
	60	**0.98**	1.10	1.09	1.16	1.00	1.08	1.06	1.08	1.10	1.04
	70	**0.95**	1.12	1.09	1.21	1.00	1.08	1.09	1.08	1.11	1.01
	80	**0.83**	1.10	1.07	1.24	**0.94**	1.08	1.09	1.04	1.13	**0.99**
	85	**0.71**	1.08	1.01	1.16	**0.88**	1.01	1.03	**0.98**	1.06	**0.96**

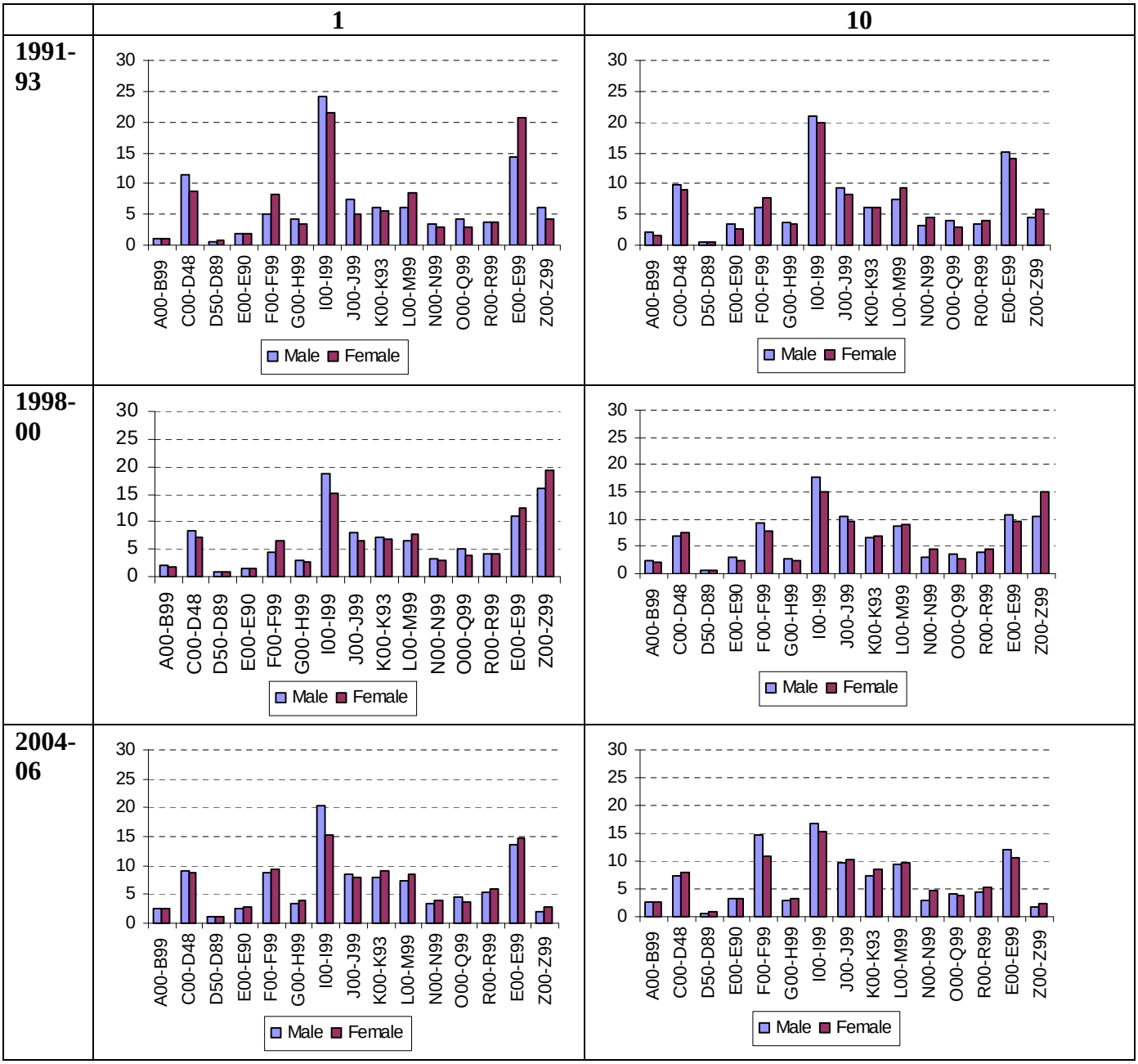
Not all deciles saw a decline or a stagnation in life expectancy at the oldest ages. The fastest increase in life expectancy was seen for males aged 80 in decile 4, a ratio of 1.54, an increase of 54 percent. However, there does not appear to be any consistent pattern across deprivation deciles. In conclusion, males saw larger relative increases in life expectancy, especially at older ages, but there were also some declines in life expectancy at age 85. Females saw smaller relative increases in life expectancy, and more declines in life expectancy at older ages, especially in decile 1. There is considerable variation in life expectancy patterns at older ages by deprivation decile, with no consistent pattern. This is due to a convergence in life expectancy (and HUEs) at the oldest ages, as illustrated by Figures 3.3.7 and 3.3.8.

3.3.3 Area Deprivation Results, Life Expectancy and HUEs by Cause, 1991-2006

This section examines the cause-specific composition of HUE_0 for deciles 1 and 10, at different points in time. This is shown in Figure 3.3.9.

Figure 3.3.9 breaks HUE_0 down into broad diagnostic groups. The bars in each graph add up to 100 percent for males and 100 percent for females. The increase and decrease in factors influencing health status and contact with health services (Z00-Z99) and the increase in mental and behavioural disorders (F00-F99) documented in the national analysis are also apparent in Figure 3.3.9. However, factors influencing health status make up a much larger proportion of HUE_0 in decile 1 than in decile 10 in 1998-00, and mental and behavioural disorders make up a larger proportion of HUE_0 in decile 10 than decile 1 in later years. External causes comprise a larger percent of female HUE_0 than male HUE_0 in decile 1, yet a larger percent of male HUE_0 than female HUE_0 in decile 10. Circulatory conditions are a more important component of HUE_0 in decile 1 than decile 10 in all time periods, especially for males, while diseases of the respiratory system (J00-J99) and diseases of the skin, subcutaneous tissue, musculoskeletal system and connective tissue (L00-M99) make up a slightly higher proportion of HUE_0 for decile 10 than decile 1.

Figure 3.3.9: Percentage of HUE_0 by Cause*, Deprivation Deciles 1 and 10, Selected Years



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality ('ecodes', separate category to E00-E90)
Z00-Z99 Factors influencing health status and contact with health services (not used for mortality data)

3.3.4 Conclusions from Area Deprivation Analysis of Life Expectancy and HUEs, 1991-2006

Life expectancy is highest in the least deprived areas, and lowest in the most deprived areas. In contrast, HUEs are lowest in the least deprived areas and highest in the most deprived areas. So even though people in the least deprived areas live longer, they spend less time in hospital over their lifetimes, and although people in the most deprived areas do not live as long, they spend more time in hospital over their lifetimes. At the oldest age groups (80+) both HUEs and life expectancy in different deciles converge. Life expectancy did not increase uniformly across deprivation deciles, over the time period, especially for females. Females in the least deprived decile saw stagnation or declines in life expectancy at all ages. However, this may be due to the very low mortality rates experienced by females in the least deprived deciles leaving little potential for improvement. This could be considered indicative of rectangularisation: reaching a limit beyond which further improvements in mortality are not possible. It is difficult to assess if rectangularisation is occurring in different deciles, due to different life expectancies. The causal composition of HUE_0 only varies slightly by deprivation decile.

3.4 Conclusion

A wide range of trends have emerged from the analysis of life expectancy and HUEs presented in this section. These trends are not uniform across deprivation deciles or across age groups. At a basic level, increasing life expectancy and decreasing HUEs nationally indicates compression of morbidity must be occurring. People are living longer, yet spending less time in hospital over their lifetimes. This pattern of increasing life expectancy and decreasing HUEs is consistent across almost all deciles (with the exclusion of stagnating and reducing life expectancy for females in decile 1). However, the combination of higher life expectancy and lower HUEs in less deprived deciles and lower life expectancy and higher HUEs in more deprived deciles indicates that individuals in less deprived areas experience less morbidity over their lifetimes than individuals in more deprived deciles. This is despite people in less deprived areas living longer. This finding appears to support the theory that as mortality decreases, morbidity also decreases.

Rectangularisation of the survival curve has been considered a prerequisite of the compression of morbidity. The findings in this section present two drawbacks to

identifying the vertical componet of rectangularisation, which results from a fixed limit to life span. In small populations such as New Zealand, previously high mortality rates result in a small population in the oldest age groups (80+). Due to this, population data may be aggregated, and a low maximum age group used for mortality data, to protect anonymity. This results in a cut-off point on the survival curve, at an age when a large proportion of the individuals in the period life table are still alive. It is therefore difficult to ascertain whether life expectancy is approaching a fixed limit to life span.

The second drawback with identifying the vertical component of rectangularisation is related to inequalities in mortality. Life expectancy varies considerably by deprivation decile. Assuming that a fixed limit to life span is the same in all deciles, this means that deciles with high life expectancy are already closer to the maximum life span. The rectangularisation of the survival curves for different deciles cannot therefore be directly compared. The findings however provide evidence of the compression of morbidity *in the absence* of rectangularisation of the survival curve.

There are considerable limitations to this study, including the relevance of hospital use as a proxy for morbidity, and the higher prevalence of private health insurance in less deprived areas, which may artificially deflate public hospital use in these areas. The next section explores these results and their limitations in greater detail, with reference to previous research using HUEs in New Zealand.

SECTION 4 – DISCUSSION

4.1 Introduction

The life expectancy and Hospital Utilisation Expectancy (HUE) results presented in Section 3 are discussed in more detail in this section. The results from the national analysis and the area deprivation analysis are interpreted with reference to the compression of morbidity hypothesis, and to findings from other studies. Private health insurance has been identified as a potential confounder of public hospital use by deprivation. There is little information on private health insurance coverage by deprivation, so national levels are discussed to identify the potential impact of private health insurance. Different types of private health care are discussed, followed by a breakdown of total health expenditure in New Zealand by type (public/private), and an examination of the population coverage of private health insurance. Finally the relevance of the results to the research questions, and the limitations of the research are discussed.

4.2 National Analysis, Discussion of Results

Rectangularisation of the survival curve is considered by many to be an essential prerequisite to compression of morbidity (Fries 1980). This can however be contested: it is argued that as long as increases in healthy life expectancy exceed increases in unhealthy life expectancy, compression of morbidity can occur *in the absence* of rectangularisation of the survival curve. That is, if decreases in lifetime hospital utilisation exceed increases in life expectancy. The vertical component of rectangularisation (the presence of a maximum life span) could not be satisfactorily examined in the analyses, due to the maximum age group of 85 and above. It was impossible to assess if the majority of deaths were occurring within a narrow age range due to high proportions of individuals in the period life tables surviving to age 85. The horizontal component of rectangularisation (reduction in premature mortality) did appear to be occurring over time. Premature mortality was reduced for both males and females and a higher proportion of individuals in the period life table survived into middle and old age, with very little drop off at younger ages. Despite limited evidence for rectangularisation of the survival curve, the national analysis provided several compelling indications of the compression of morbidity. It was demonstrated that

lifetime hospital use (HUE_0) decreased by nearly 50 percent, while life expectancy at birth increased by over 9 years for males and by 7 years for females, in the space of just three decades. Decreasing lifetime hospital use (when hospital use is taken as a proxy for morbidity) denotes compression of morbidity, yet these decreases occurred against a backdrop of rising life expectancy, resulting in longer exposure to the risk of hospital use. Life expectancy increased most rapidly at ages 80 and 85, which suggests that age 85 is not very close to the hypothetical fixed limit to life span: the force behind the vertical component of rectangularisation. If age 85 was close to the maximum life span, there would be only small relative increases in life expectancy at age 85. However, that HUE also decreases most rapidly at age 85 is further evidence for compression of morbidity, despite the apparent absence of an upper limit to life expectancy. Life expectancy is increasing fastest at the oldest ages, yet hospital use is decreasing fastest at these ages, indicating longer, but healthier lives.

This research mirrors the findings of other studies of HUEs and life expectancy in New Zealand (Cheung, Katzenellenbogen *et al.* 2001; Pool, Baxendine *et al.* 2006; Pool, Baxendine *et al.* 2009). The filtering process used varies from previous research, resulting in slightly higher female than male age-standardised bed day rates in this study, compared to higher male than female age-standardised bed day rates in previous studies. However, when HUEs are considered, the longer female life expectancy and higher period 'at risk' of hospitalisation results in substantially higher female than male HUEs, in line with previous research. This suggests that filtering differences do not affect sex-specific HUEs to the same extent as bed day rates due to the fact that HUEs are mediated by life expectancy. However, the gap between male and female HUEs is greater in this study than in previous studies, Pool *et al.* 2009 found a difference in HUE_0 less than 10 days between males and females in 1980, compared to a difference of nearly 20 days in this study. The patterns of HUEs over time also match previous studies, despite the filtering differences, a plateauing of HUE0 in the early- to mid-1980s, followed by sustained decline, levelling off slightly in the mid and late 1990s.

Different results, supporting expansion of morbidity were found by other New Zealand studies that used period prevalence life tables, but a different measure of health: self reported health data from surveys. A study using independent life expectancy (a variant of health expectancies) found total life expectancy to be increasing faster than independent life expectancy between 1996 and 2006, resulting in an expansion of morbidity (Ministry of Health and Statistics New Zealand 2009). Two further studies using

health expectancies also found tentative support for expansion of morbidity (Davis, Graham *et al.* 1999; Davis, Mathers *et al.* 2003). Thus while the results obtained agree with previous research using HUEs, they deviate from research using self reported health data from surveys. Hospital bed days have many limitations, some of which are examined in section 4.4. However, it is argued that when potential confounders are taken into consideration and the data is filtered to represent ill health and improve consistency over time, hospital bed days provide a far more objective and reliable measure of health than self reported health measures.

4.3 Area Deprivation Analysis, Discussion of Results

The area deprivation analysis found that while life expectancy was increasing and HUEs decreasing in all deciles, there were differences by decile in the pace of both these changes, and the levels of life expectancy and HUEs. HUE_0 in 2004-06 for males and females was about 33 days higher in the most deprived decile than in the least deprived decile, a large differential considering that the national HUE_0 at this time was less than 60 for both males and females. Life expectancy in the most deprived decile was 7 and 10 years lower than in the least deprived decile for males and females respectively in 1991-93. However, this gap narrowed over time to 5 and 9 years for males and females in 2004-06. Females in decile 1 saw no improvement in life expectancy over the time period, and at many ages female life expectancy in decile 1 decreased.

The stagnation of mortality declines for females in decile 1 was also apparent in the survival curve, with almost no change over the time period despite having the most evidence of rectangularisation. The largest improvements in the survival curve were seen for males in the most deprived deciles, the group which also had the least evidence of rectangularisation. This suggests that females in the least deprived deciles have reached the minimum attainable mortality rates, and that the greatest improvements in mortality occur for males in more deprived areas because there is the greatest *scope* for improvements. At older ages, as was seen for mortality (but not for hospital bed day rates), convergence is seen in life expectancy and HUEs by deprivation decile.

Other studies have found rectangularisation among high socioeconomic groups due to a lower distribution of death rates until age 70, after which accelerating rates of disability and convergence with other socioeconomic groups is seen. This matches the

findings from this research and suggests compression of morbidity may be occurring among high socioeconomic groups (Fries 1992). This research also supports the observation by Pool of epidemiologic polarisation within countries (Pool 1994). However, given the differences in life expectancy by deprivation in New Zealand, this would appear to be a self fulfilling prophecy: it is not necessarily the case that rectangularisation of the survival curve and compression of morbidity are *not* occurring in more deprived areas, only that more deprived areas are further behind in this process due to life expectancy being lower and thus further from the hypothetical fixed limit to life span.

No direct comparison can be made with previous research, given that this is the first study to examine HUEs by area deprivation, in New Zealand or elsewhere. However the deprivation analysis used the same methods as the national analysis, which was found to be consistent with previous research using the same data and methods.

4.4 Private Health Insurance in New Zealand, Implications for Public Hospital Use by Area Deprivation

One shortcoming of the hospital data used in this research is the exclusion of private hospital data. If private health care is common in New Zealand it could distort public hospital figures for areas with high uptake of private services. This section examines the prevalence of private health insurance in New Zealand and the proportion of health expenditure that comes from public, private and other sources. While the majority of health expenditure in New Zealand comes from public sources, not all public health care is free and there is a thriving private health insurance industry in New Zealand. Between 1995 and 2006, the percent of total health expenditure from public sources increased slightly from 77.2 to 77.8 percent, with privately sourced health expenditure decreasing accordingly from 22.8 to 22.2 percent (World Health Organization 2010).

There are three types of private health expenditure: private health insurance, out-of-pocket expenditure, and not for profit organisations. Private health insurance policies can be either comprehensive, or cover only major medical and surgical care (Health Funds Association of New Zealand 2001). Out-of-pocket expenditure is expenditure that is not reimbursed by either public or private health insurance (Hopkins and Cumming 2001). This includes, in order of decreasing importance, prescription medications, dental care, general practitioner care, and surgical and medical care (Ministry of Health 2005). There

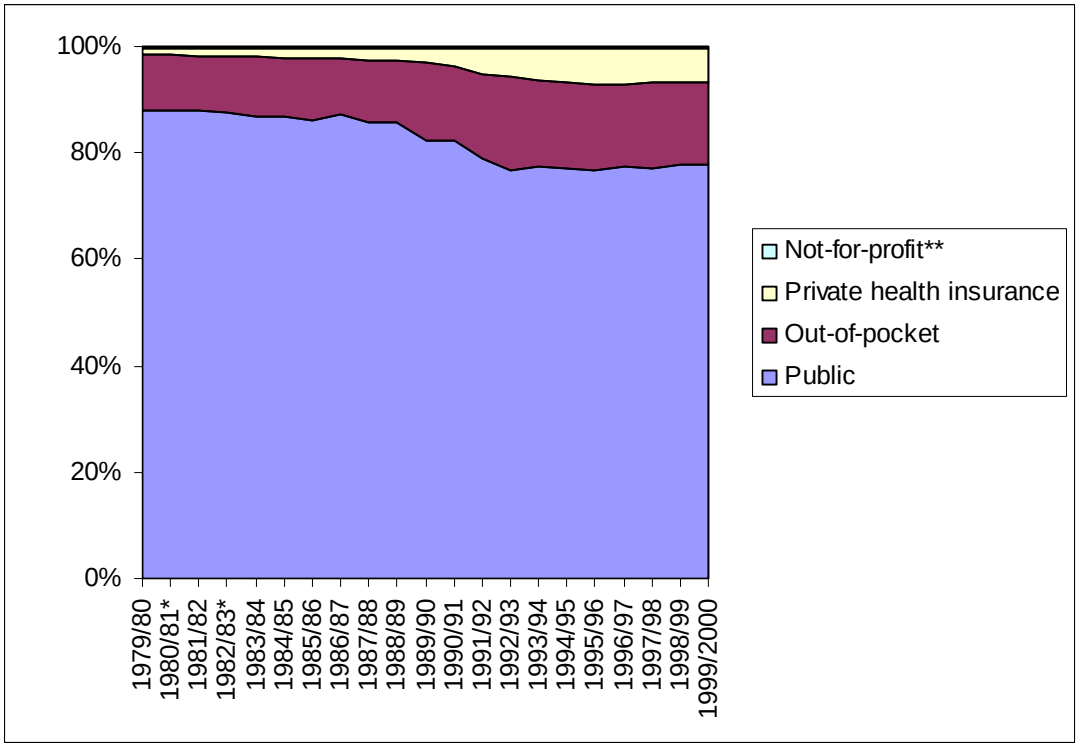
are also not-for-profit and voluntary organisations that provide health care, which receive funding from governmental and non-governmental sources. Non-governmental funding for these organisations is also categorised as private health expenditure (Ministry of Health 2005). Figure 4.4.1 shows the contribution to total health expenditure made by public and private sources. Expenditure due to private health insurance is expenditure on health by the private health insurance industry i.e. the cost of claims, not the cost of the insurance premium itself.

The percentage of total health expenditure coming from private sources has remained constant since 1992/93, following an increase in the late 1980s and early 1990s. The largest private source of health expenditure was out-of-pocket expenditure, which increased from 10.4 to 15.4 percent of total health expenditure between 1979/80 and 1999/2000. Private health insurance only made up 6.3 percent of total health expenditure in 1999/2000, but had increased nearly six fold since 1979/80 when it made up only 1.3 percent. A decline has occurred in more recent years (not shown in Figure 4.4.1): in 2004/05 and 2005/06 private health insurance accounted for only 4.7 percent of total health expenditure (Ministry of Health 2008). Of private sources, not for profit organisations constitute consistently less than 0.5 percent of total health expenditure over the time period, and as such are not legible in Figure 4.4.1.

The increase in out-of-pocket costs of health care is of importance for this research. Increases in costs for prescription medications and general practitioner care may deter those on low incomes from seeking primary health care for conditions amenable to primary interventions, and so inflate hospital use in more deprived areas. There are no out-of-pocket costs incurred from public hospital use in New Zealand[16]. The surgical and medical component of out-of-pocket costs refers to private treatment, for those who do not have private health insurance, or have policies that do not cover such treatment.

[16] With the exception of inpatient charges between 1991 and April 1993, and outpatient charges between 1991 and July 1997, for individuals who did not hold Community Services Cards. Individuals that met certain income criteria were eligible for Community Service Cards (Hopkins and Cumming 2001).

Figure 4.4.1: Percent of Total New Zealand Health Expenditure by Source, 1970/80 to 1999/2000, Adjusted for Inflation



* Data missing for 1980/81 and 1982/83, linearly interpolated from adjacent years.
** Not-for-profit sources constitute consistently less than 0.5% of total health expenditure, and are therefore not legible.
Source of data: (Ministry of Health 2002)

The percentage of total health expenditure accounted for by private health insurance increased over the time period studied. However, the percent of New Zealanders with private health insurance decreased. Reliable data is only available from 1990, at which time 50 percent of New Zealanders were covered by private health insurance (Hopkins and Cumming 2001). Since 1990, the percent of the New Zealand population covered by private health insurance has decreased to 33 percent, see Table 4.4.1.

Table 4.4.1: Percent of New Zealanders with Private Health Insurance Policies, 1994/95 to 2002/03

Year	%
1994/95	41
1996/97	40
1998/99	37
2000/01*	34
2002/03*	33

* Data for 2000/01 and 2002/03 provided by Health Funds Association, data for earlier years provided by Statistics New Zealand.
Source: (Ministry of Health 2002; Ministry of Health 2005)

Despite declining prevalence, a third of the population had private health insurance in 2002/03. If the third of New Zealanders with private health insurance all live in the top three deprivation deciles, it would be expected that public hospital use in these deciles would be severely deflated. However even if this were the case, the data presented here suggests that deflation of public hospital use in less deprived deciles may be only minimal. Approximately half of the people with private health insurance policies in New Zealand in the early 2000s had major medical and surgical policies and half had comprehensive policies, the proportion with major medical and surgical policies is increasing over time (Health Funds Association of New Zealand 2006). Those with major medical and surgical private health insurance policies would continue to use public health services for their day to day health care needs. Furthermore, consider the disparity between the percent of the population with private health insurance, and the percent of total health expenditure that comes from private health insurance claims. Despite over a third of New Zealanders having private health insurance policies at the turn of the century, private health insurance claims made up only 6.2 percent of total health expenditure, dropping to 4.7 percent in 2004-06. Either people with private health insurance have superb health and therefore low demand for health services, or they continue to use public health services to a large extent.

Private health insurance cushions against out-of-pocket health costs. Hopkins and Cumming examined the proportion of household income that was spent on out-of-pocket costs and private insurance policies by income decile in New Zealand between 1984 and 1998. They found that households with low incomes (particularly the lowest tenth) spend a higher proportion of their income on out-of-pocket health care costs than households with higher incomes (particularly the highest tenth) (Hopkins and Cumming 2001). In 1998, out-of-pocket expenditure for the households in the lowest income group was nearly 7 percent of income, compared to less than 2 percent for households in the highest income group. The proportion of household income spent on out-of-pocket health services is increasing over time, especially for households with low incomes, despite those on low incomes being eligible for subsidised primary health care through the Community Service Card scheme, which was replaced with area-based funding in 2002 (Hefford, Crampton *et al.* 2005). Both uptake and knowledge of the Community Service Card scheme were poor. Only 72 percent of those eligible held a Community Services Card (Hopkins and Cumming 2001) and 64 percent of those who did *not* know what Community Services Card was, currently held one (Hefford, Crampton *et al.* 2005).

Rather than raising concerns over the deflation of public hospital use in less deprived areas due to private health insurance, this examination has emphasised the failures of the New Zealand health system to reduce barriers to health care for those on low incomes. Private health insurance expenditure has increased sharply since 1979 despite a reduction in population coverage, but still constitutes only 6.2 percent of total health expenditure in New Zealand, which is disproportionately small compared to the percent of the population with health insurance. The increase in out-of-pocket health expenditure, from 10.4 to 15.4 percent of total expenditure, coupled with the increasing proportion of household income spent on out-of-pocket health expenditure for households with low incomes has serious implications for access to primary health services (and resultant inflated public hospital use) in more deprived areas.

4.5 Relevance of Findings to Research Questions

The research question outlined for this paper involved the examination of mortality and morbidity using life table methodology to identify the health transition (compression, expansion, dynamic equilibrium) that is occurring in New Zealand. The research question is given below:

3. Taking hospital utilisation as a proxy for severe morbidity, do the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through the use of hospital utilisation expectancies (HUEs)?
    d. Does the lifetime expectation of hospital use at birth and at different ages differ by deprivation decile?
    e. How have these trends changed over time?

The evidence provided in Section 3 provides strong support for compression of morbidity when lifetime hospital use is used as a proxy for morbidity. No evidence was found for a fixed limit to life span (or expectancy), which is considered by many to be a prerequisite to compression of morbidity. Indeed, life expectancy was found to increase fastest at the oldest ages. However lifetime morbidity declined *despite* increasing life expectancy, and declined fastest at the oldest ages which also saw the fastest increases in life expectancy. This suggests that a limit to life expectancy and the resultant vertical component of rectangularisation are *not* essential for compression of morbidity to occur. It also suggests that longer lives may be healthier lives, and that increases in the numbers of

very old (85 and over) will not result in either increased morbidity or increased demand for hospital services at these ages.

The lifetime expectation of hospital use by deprivation decile was found to differ by deprivation decile. These differences were greatest at birth, when, in 2004-06 there was a difference of approximately 33 days in HUE_0 between deciles 1 and 10, and converged at older ages. The difference in HUE_{85} between deciles 1 and 10 for females was 12 days. HUEs decreased over time for both sexs, at all ages and in all deciles, despite rising life expectancy for both sexs at nearly all ages and in nearly all deciles.

4.6 Limitations of this Research

The limitations for this research are similar to the limitations outlined for Paper One. These include the higher proportion of Maori and Pacific peoples in more deprived areas; heterogeneity of socioeconomic conditions within small areas; the absence of private hospital data; shortcomings of public hospital use as a proxy for morbidity; and coding and recording inconsistencies in the hospital data. However it would seem that any sex-specific effects of the filtering process used are less influential for HUEs than for hospital bed day rates, due to the weighting of hospital use by life expectancy. Residential care homes are not included in the public hospital data. This may confound public hospital statistics for those in the oldest ages (85+), for whom declining HUEs were found despite increasing life expectancy. If a large proportion of the population aged 85+ are residents in nursing or retirement homes, palliative care may be provided in these settings rather than in a hospital setting. This potential confounder is examined in more detail in Paper Three.

It must be reiterated that results presented in this section are drawn from period life table analysis, and therefore represent life expectancy and hospital use patterns that *would* occur to a hypothetical cohort *if* the age-specific rates of mortality and hospital bed days for a given time period remained constant over a lifetime. These results therefore are not predictions, but snapshots of the mortality and hospital use experience of the population at different points in time.

4.7 Conclusion

The results from this study provide compelling, but not comprehensive evidence for compression of morbidity, particularly in less deprived areas. It is proposed that rather than more than one health transition occurring in New Zealand, more deprived areas are simply further behind in the compression of morbidity process.

This section has also examined the potential confounding effect of private health insurance on public hospital trends by deprivation in New Zealand. Private health insurance undoubtedly has an effect on public hospital use, especially in less deprived areas, and one in three people in New Zealand are covered by private health insurance policies. However it must be remembered that approximately half of these policies apply to only major medical and surgical conditions, and that the percent of total health expenditure that comes from private health insurance claims is disproportionately small in comparison to the percentage of the population with private health insurance. Of greater concern is the rising importance of out-of-pocket expenses in total health expenditure, coupled with the high proportion of household income spent on out-of-pocket health expenditure for households with low income, which is rising over time. Despite subsidies, there may be substantial barriers to accessing primary health care for individuals on low incomes, which could artificially inflate hospital use for these groups.

Given the decreasing HUEs at the oldest ages, increasing life expectancy in New Zealand is not of great health policy concern. What *is* of concern is the differences in hospital use by area deprivation. In more deprived areas individuals not only die younger, but also spend much more time in hospital over their lifetimes than their longer lived counterparts in less deprived areas. This difference in public hospital use by deprivation may be due in part to persistent financial barriers to primary health care and a complex, little understood subsidy system.

PAPER THREE

Examination of Compression of Morbidity using Longitudinal Techniques

RESEARCH QUESTION 4

4. Taking hospital utilisation as a proxy for severe morbidity, do the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through analysing time spent in hospital in the last months of life?
    a. Does hospital use in the last months of life differ by deprivation decile?
    b. Does hospital use in the last months of life differ by age at death?
    c. How have these trends changed over time?

ABSTRACT

Hospital use in the last few months of life decreased over the time period 1990-2006, the largest declines were seen at older ages. There has been considerable convergence between more and less deprived deciles in hospital use in the last few months of life for individuals who died in New Zealand between 1993 and 2006, particularly at older ages (where the differences by deprivation were greatest). Declining hospital use in the last few months of life over the time period (particularly at older ages) may suggest that people are being missed by the public hospital system in the months and years prior to death. This could be due to substitution of healthcare from public hospital to other settings such as private hospitals, nursing/retirement homes, or palliative care in the home. These explanations are examined and, while there is some limited evidence for substitution of end of life care from public hospitals to other settings, it cannot fully explain the decreases in hospital utilisation prior to death over the time period. Other explanations include changes in health seeking behaviour, improved primary health services enabling effective preventive treatment, earlier diagnosis, and technological advances resulting in less invasive hospital procedures and faster recovery time.

SECTION 1 - LITERATURE REVIEW: Ethics of Using Medical Records in Research and Individual Record Linkage

1.1 Introduction

In this paper record linkage is used to examine hospital utilisation in the months prior to death, with the aim of inferring if compression of morbidity is occurring in New Zealand. The research question in this paper is similar to the research question in Paper Two: both explore the compression/expansion of morbidity debates using New Zealand hospital and mortality data, and examine differentials by deprivation decile and changes over time. However unlike Paper Two, which used cross-sectional data to compute period prevalence life tables, this paper uses record linkage, which facilitates longitudinal techniques. This enables the examination of hospital utilisation (a proxy for morbidity) in the last months of life for decedents. If hospital utilisation in the months leading up to death is declining this indicates compression of morbidity; increasing hospital utilisation in the months leading up to death indicates expansion. The use of record linkage also allows us to explore differentials by age at death, something not possible using period prevalence life table methods. The difference between this paper and Paper Two therefore hinges around the type of methods used (cross-sectional or longitudinal). Different methods have different strengths, and limitations. A range of methods is essential to gain a full understanding of the subject and improve the validity of results.

Discussion of the literature regarding the use of hospital records as a proxy for morbidity, and discussion of the literature concerning the compression and expansion of morbidity theories is found elsewhere (in Papers One and Two respectively). In this, third paper in the series, the literature review focuses on considerations unique to this paper, which are primarily methodological in nature. Areas discussed include the ethical issues surrounding use of individual administrative hospital records in health research, and how ethical considerations change when record linkage is conducted. There is relatively little precedent for the use of longitudinal methods in compression/expansion of morbidity research, the few studies that do exist are also discussed.

Using record linkage in health research has immense benefits, but also entails considerable responsibility in ensuring that the privacy and confidentiality of the data are

conserved. Formal data linkage systems, where datasets for a population are automatically updated and strict protocols are followed, is by far the best way of maximising the benefits of linked health data while minimising risks to privacy and confidentiality (Kelman, Bass *et al.* 2002). However, due largely to political reasons, there are few such data linkage systems worldwide and the use of individually linked health data in research is largely conducted on an *ad hoc*[17] basis. Public knowledge of medical research and the safeguards in place when conducting medical record linkage is poor (Sibthorpe, Kliewer *et al.* 1995; Robling, Hood *et al.* 2004); the concept of a population-level linkage system that links an individual's medical records from different sources is initially rather alarming. Epidemiologists need to become salespeople and sell their research to the public, emphasising the benefits to public health. To safeguard the future of medical research it is necessary to raise public awareness of its procedures, findings, and benefits.

This literature review first explores the ethical issues of consent, privacy, and confidentiality regarding the use of (unlinked) medical records in health research. The benefits and limitations of record linkage to medical research are then discussed, along with associated ethical considerations. The Western Australia Data Linkage System (WADLS) is used as a case study of a population-level data linkage system, and the benefits of such a system over *ad hoc* record linkage are explored. Existing studies using linked data to examine hospital utilisation in the months prior to death are then discussed, and the key points from the literature review are revisited in the conclusion.

1.2 Use of Medical Records in Research: Privacy and Consent

The data used in these papers are administrative public health and mortality data. In the previous two papers this data has been used in a non person specific, events based way. The researcher was not interested in to whom the events occurred, but how many events occurred within the population at risk. While an individual identifier (encrypted National Health Index (NHI) number) was included in the datasets, this was of no interest to the researcher and may as well have not been present. In this paper the presence of the individual identifier is crucial to the analysis, and the ethics of using potentially

[17] *Ad hoc* record linkage refers to one-off data linkage conducted by researchers as needed specifically for a particular study, rather than a data linkage system that routinely links and provides data to researchers.

individually identifiable administrative public health records in research is therefore discussed.

In recent decades research involving medical records and other health data has been subject to "an increasingly complex, and constraining web of [privacy] legislation" (Magnusson 2002, p8). Technological advances, modern computers and the internet have increased the possibilities for medical research, but also increased the risk (real and perceived) of loss of privacy and confidentiality of the individuals to whom the data relates. Increasing the privacy legislation in developed countries regarding the use of health data has been appropriate for commercial research, but less so for medical research (Trutwein, Holman *et al.* 2006) and there is evidence that changes in privacy laws have had a detrimental impact on health and medical research (Whiteman 2005). The need for action in order to protect the ability to conduct medical research was recognised by Gordis *et al.* as far back as 1977:

> "The time is long past when epidemiologists and other investigators can retire into secluded academic corners and ignore the need for political and social attention" (Gordis, Cold *et al.* 1977, p168).

There are two issues regarding the use of individual level health data in research: consent of the patient to involvement in research, and the risks to privacy posed by modern health information systems. Together, privacy and consent protect individual dignity and autonomy (Magnusson 2002) and must be vigorously protected at all stages of data collection and utilisation. Issues surrounding the privacy of administrative health records are discussed first.

There are several different types of research that use administrative health data, however the primary use for which administrative health data is collected is clinical. Legislation regarding clinical use of data is fairly relaxed, while increasingly strict privacy legislation has targeted secondary uses of administrative health data. Secondary users of administrative health data include medical (or health) researchers, commercial researchers, and law enforcement administrators. To health researchers, the identity of the individuals in the data is irrelevant (Sibthorpe, Kliewer *et al.* 1995), and therefore the potential for protecting the privacy of individuals in the data is high. The aim is to conduct unbiased research which will ultimately benefit public health and therefore the community (which the individuals in the data hail from). While the results of health research may not directly

benefit the individuals who contribute to the data (rather obviously in the case of mortality data), it is in the best interests of the community as a whole.

In contrast, commercial researchers have a primarily financial aim which is of benefit to the organisation conducting the research and not necessarily the community. The identity of individuals in the data is often of interest to commercial researchers, who may wish to contact individuals, resulting in risks to privacy. Law enforcement administrators use health data to uncover fraud and alleged wrongdoing committed by individuals (Magnusson 2002). The identity of the individuals in the dataset is of great importance to law enforcement administrators, and while there may be benefits for the community as a whole (less tax evasion and fraud leading to lower taxes and insurance premiums) there is a distinct disadvantage for the individual concerned.

Much of the privacy legislation that has been implemented in developed countries in recent decades has made little distinction between these types of research. And in the public eye, non-clinical research using administrative data tends to be associated with commercial or law enforcement purposes. There is limited knowledge of the benefits of health research; in order to change public opinion researchers need to sell research to the public and make the public more aware of the types of health research being conducted and of the safeguards (ethics committee approval) and benefits of this research to public health (Magnusson 2002).

The second issue regarding the use of administrative health data is the consent of the patient for their information to be used in research. The research community is united in opposition of informed consent, an opinion that is initially rather shocking. However, while asking patients for consent for their records to be used in research is desirable, it is unfortunately not practical on many levels. Asking patients for consent for their information to be used in medical research would defeat the purpose of such data. Unique benefits of administrative health data such as lack of selection bias would be lost (Kelman, Bass *et al.* 2002). Asking patients for consent would generate extra paperwork and administration for both the patient and the doctor (Magnusson 2002), and in many situations it is simply not possible to contact patients for consent (decedents) (Wald, Law *et al.* 1994). Patients are by definition not well, and are often not able to give informed consent. Lack of informed consent for medical records to be used in public health research is not necessarily objected to by the public. While there has been little research on this, when attempts have been made to ask the public for their views on the lack of informed consent in health research, results have been favourable. In an Australian survey, 88

percent of people agreed with the lack of consent in research using patient records (Sibthorpe, Kliewer *et al.* 1995), and in a series of focus groups held in South Wales in the UK there was recognition by participants of sample bias and of the need for individuals to forego consent for the benefit of the community (Robling, Hood *et al.* 2004).

Ethical considerations become rather more complex when linking records together for each individual. The reasons for such linkage and associated ethical considerations are discussed next.

1.3 Benefits and Limitations of Medical Record linkage

Record linkage is arguably as old as records themselves (Acheson 1967), and has been conducted on a small, *ad hoc* basis for various reasons as long as records have been collected. As early as 1946, well before technology made large scale medical record linkage possible, Halbert Dunn recognised the potential of linking records and described the process as compiling a 'Book of Life'.

> "Each person in the world creates a Book of Life. This Book starts with birth and ends with death. Its pages are made up of the records of the principal events in life. Record linkage is the name given to the process of assembling the pages of this Book into a volume." (Dunn 1946, p1412)

Dunn was aware of the impracticability of physically 'binding' an individual's records into a book, suggesting instead that an index was created indicating *where* the different pages (records) were located. This has proved surprisingly prophetic, although the reasons for keeping records separate are now related to privacy rather than practicality[18].

In light of the ethical considerations when using *un*linked medical records in health research, linking records seems foolhardy. However, medical record linkage has considerable benefits. Internationally record linkage at a population-level is well established in other (non-health related) fields, such as taxation, but remains controversial in health research. Record linkage enables the researcher to test the reliability of data. Different sources of data often include the same information (such as date of birth on

[18] Even with modern computers and the potential to collate an individual's records in one place, data linkage systems such as the Western Australia Data Linkage System (WADLS) opt to keep records separate and retain a 'chain' of links in order to connect records when required for research (this is discussed in more detail later in section 1.4).

hospital and mortality records), comparing this information in one source with another can highlight errors in the data, such as duplications and typing errors (Acheson 1967; Johansson and Westerling 2000). Several factors have made record linkage necessary for medical research in recent decades. The increasing mobility of populations in developed countries and the increasing complexity of health systems mean that health events may occur in different facilities and different places, requiring integration of data from different sources (Acheson 1967). There has also been an increase in the length of the observation period required for medical research, brought about by the increased importance of chronic diseases as a cause of death in developing countries. "Today time acts on a different scale. The predominant diseases creep upon us over years or decades, the onset may be insidious and the course prolonged" (Acheson 1967, p8). Longitudinal research that examines the habits of a lifetime, genetic factors, and exposures to risk factors is required to capture morbidity and mortality patterns in countries that are at an advanced stage of the demographic transition. Countries where chronic and degenerative diseases are the leading causes of sickness and death.

While it is *possible* to conduct longitudinal research by following individuals over time, and conducting interviews and questionnaires, this approach is expensive, very large samples are required for rare outcomes, and there are long delays between collection and completion of the research due to delays between exposure and outcome in disease aetiology, making the results unsuitable for timely public health action (Sibthorpe, Kliewer *et al.* 1995). In comparison record linkage is cheap and timely to public health intervention: following individuals in real time is simply unnecessary (Magnusson 2002).

Record linkage is therefore an increasingly relevant and important tool in public health research in settings where disease onset is unclear and disease progression is slow. These are characteristics of chronic and degenerative diseases, which are accounting for a larger share of morbidity and death in many developed countries, and which will become increasingly important in developing countries as they progress through the stages of the demographic transition.

Record linkage requires either a unique identifier or key that is the same for an individual across all data sources, such as a National Health Service number,[19] or a set of personally identifiable information that is included on each data source, such as name,

[19] It is necessary for this key to be present on *all* datasets to be linked. So a National Health Number would be suitable for linking primary and secondary public health services in the UK, but not private health services.

address, date of birth and gender. While no one piece of personally identifiable information is enough to link records for an individual, when used collectively it is possible to identify individuals across data sources (Newcombe and Kennedy 1962). Evidently, a unique identifier is preferable when conducting record linkage. It reduces the possibility for incorrect matches, and eliminates the need for the researcher to see personally identifiable information such as names and addresses, improving both privacy and confidentiality. However public opinion is fiercely against a universal identity (ID) number, despite benefits to privacy and confidentiality. "Ironically, having a unique identifying number…would provide the most confidential basis for linkage" (Sibthorpe, Kliewer *et al.* 1995, p253). In the UK this opposition dates back at least to the Second World War, when a universal ID number was associated with conscription, identity cards (which were required to be carried at all times) and wartime security (Acheson 1967). In more recent years in developed countries a universal ID number has been associated with a 'big brother' society and fears of a central 'dossier' for each individual, containing sensitive, personal information. However, as Acheson argued the question is not *should* we have a unique number for each individual, but how *many* should we have. At the time of writing in 1967 Acheson counted over 35 different types of number that could be assigned to an individual for health and welfare purposes alone (Acheson 1967). Acheson also pointed out that birth and death records were public documents which could be viewed by anyone for a fee, and contain highly sensitive information such as cause of death (on death certificates) and enough information to deduce illegitimacy on birth certificates. In a system of electronically linked records there would be no reason for this information to be publicly available (Acheson 1967).

Record linkage is not however perfect. In an examination of the accuracy of record linkage between American hospital discharge and death records, Zingmond *et al.* discovered that unidentified linkages were slightly greater for vulnerable populations, including ethnic minorities, the elderly, and women (Zingmond, Ye *et al.* 2004). This would result in under representation of these groups in research using record linkage. It is worth bearing in mind however that the hospital data used in this study had a high proportion of hospital records with no unique identifying number (800,000 out of 2.8 million records) and therefore a high proportion of unidentified linkages. Despite this where achieved, record linkage was found to be highly accurate.

Poorly conducted *ad hoc* record linkage, without safeguards such as approval from an ethics committee and storage of the data in a secure facility, does pose risks to privacy

and confidentiality of the data. Large scale health data linkage systems have been established around the world. These systems follow strict protocols and procedures when linking datasets and making when linked data available to researchers. One such system is the Western Australia Data Linkage System (WADLS). The procedure WADLS uses for linking data and providing data to researchers and associated benefits over *ad hoc* methods are described below.

1.4 Western Australia Data Linkage System: A Case Study

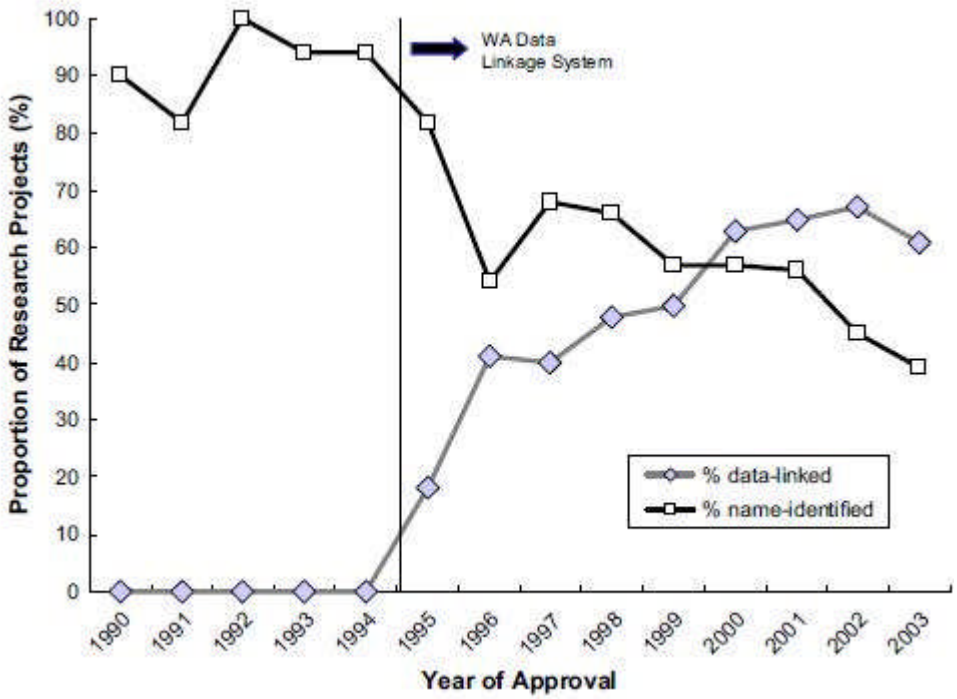
Kelman *et al.* stress the benefits to privacy and confidentiality of using a formal data linkage system such as WADLS rather than *ad hoc* record linkage (Kelman, Bass *et al.* 2002). In a formal data linkage system, privacy and confidentiality are protected at every stage. A good example of this is the WADLS. WADLS was established in 1995, since which some 400 studies have been conducted using WADLS data, from which there have been over 250 journal publications and 35 graduate degrees (Holman, Bass *et al.* 2008).

There is no unique ID across all datasets used by WADLS. It is therefore necessary to conduct probabilistic linkage using personally identifiable demographic data including name, address, sex, and date of birth. This is conducted by an independent third party, after which the personally identifiable demographic data is deleted. Thus the use of personally identifiable data is transitory: once a unique ID number has been assigned to each individual following the linkage, identifiable information such as names and addresses are deleted. The third party who perform this initial linkage does not see any health data, and has no communication with anyone who *does* have access to the health data. Their role is simply to generate a unique person identifier for each individual which is the same across the datasets. There is no central repository of data or ‘health dossier’ for each individual, instead chains of links are stored and the datasets are held by separate data custodians. One data custodian holds the de-identified dataset for hospital discharges, another the de-identified dataset for mortality, and so on. On applying for data, the researcher has to go through a formal process, including clearance by an ethics committee, before being approved. The researchers are only supplied with data needed for their particular project. Each data custodian extracts the necessary records, applies a separate, project-specific ID number for each individual, and after removing the person identifier sends the dataset to the researcher. The researcher then links the datasets together using the project-specific ID

number. In this way *only* the end point researcher sees the final linked file containing a project-specific ID, and after the research is completed this linked file is deleted (Kelman, Bass *et al.* 2002; Trutwein, Holman *et al.* 2006).

For researchers, the alternative to using a data linkage system such a WADLS is often to apply for name-identified data on an *ad hoc* basis. In this way the researcher conducts the linkage of personally identifiable data themselves with considerably fewer safeguards to privacy and confidentiality than are provided by WADLS. The instigators of the Western Australia Data Linkage System have a particularly impressive claim. Since conception of WADLS in 1995 to 2003, the proportion of ethics-approved research projects in Western Australia that were approved utilising name-identified administrative health information fell from nearly 90 percent, to less than 40. In comparison, the percent of approved projects utilising de-identified, data-linked administrative health information rose from 10 percent to over 60 (see Figure 1.4.1).

Figure 1.4.1: Proportions of Ethics-Approved Research Projects using Name-Identified and Data-Linked Administrative Health Information in Western Australia 1990-2003



Note : N=408
Source: (Trutwein, Holman *et al.* 2006, p280)

On examination, a formal data linkage system such as WADLS provides superior privacy and confidentiality than the alternative: *ad hoc* record linkage. And benefits to

privacy and confidentiality would be far higher still if a unique id number could be used for data collection from all administrative health sources in Western Australia. This would bypass the need for anyone to see personally identifiable identification such as names and addresses, however transitory this may be. Persuading the public that such a system is in their best interests however, is not easy.

Due to the absence of such a system, it is necessary to conduct *ad hoc* record linkage for this study. Extensive safeguards are applied, as described in Section 2. Where available however, formal record linkage systems are far preferable to *ad hoc* record linkage.

1.5 Previous Research using Linked Health Data to Examine Hospital Utilisation Prior to Death

Due to data quality and access issues, and the rarity of formal data linkage systems, relatively few studies have examined hospital utilisation prior to death. Most of these studies have been conducted on an *ad hoc* basis, with a focus on cost of healthcare, rather than morbidity. These studies are discussed below.

Research has repeatedly found that a disproportionate amount of health service utilisation (and health care expenditure) is concentrated in the last few years and months of life (Roos, Montgomery *et al.* 1987). This has led to debate over the relevant importance of ageing and time to death, and (given increasing longevity) implications for population health expenditure in future. Much research on hospital utilisation prior to death has been conducted from a health economics perspective and has been concerned with health care expenditure and the *cost* of death (Brameld, Holman *et al.* 1988; Zeweifel, Felder *et al.* 1999; Felder, Meier *et al.* 2000; Seshamani and Gray 2004; Stearns and Norton 2004; Moorin and Holman 2008). Stearns and Norton used American data and found that due to increasing longevity and concentration of health care expenditure prior to death, predictions that don't account for time to death overestimated costs by 9 percent in 2003. For 2020 projections, the overestimation of costs was 15 percent.

Zweifel *et al.* used a two-step Heckman model and Swedish private health care data and found that age was not related to health care expenditure once time to death was accounted for. The increase in overall health care expenditure with age was argued to be due to increased longevity resulting in more people being in older age groups, where death

is more common. The methods used in this study have however been vigorously disputed (Salas and Raftery 2001; Seshamani and Gray 2004), Seshamani and Gray replicated the Zweifel *et al.* the study using data from the Oxford Record Linkage Study and found the same results. However when a different methodological approach was applied, both age *and* proximity to death were found to be significantly associated with health care expenditure. Seshamani and Gray also found a levelling off of health care costs close to death at older ages. This is a finding replicated by others (Brameld, Holman *et al.* 1988; Felder, Meier *et al.* 2000; Moorin and Holman 2008), but brought into question by Roos *et al.*, who examined hospital, nursing home and physician visits in the last few years of life (Roos, Montgomery *et al.* 1987). They found that while *hospital* bed days (and expenditure) were lower prior to death for elderly decedents, nursing home utilisation was much higher than for younger decedents, which pushed up the cost of death at older ages. The vast majority of studies cited use only data from hospital episodes. The findings by Ross *et al.* emphasise the importance of other types of health care in contributing to total health expenditure prior to death, particularly for older decedents.

Relatively few studies have focussed on the health rather than the economic aspect of hospital use prior to death, and attempted to ascertain how hospital use prior to death relates to the compression/expansion of morbidity debate (Roos, Montgomery *et al.* 1987; Leibson, Ballard *et al.* 1992; Hessler, Eriksson *et al.* 2003; Lynch, Holman *et al.* 2007). Of these studies that do examine compression of morbidity, there is little consistency in the definition of morbidity used. Lynch *et al.* and Leibson *et al.* take the first hospital admission for a given condition: Lynch *et al.* use data from Western Australia to consider the onset of morbidity to be the first hospital admission for a chronic disabling, or activity limiting diagnosis, while Leibson *at al.* examine American data for trends in specific diseases (coronary heart disease, stroke, and cancer). Lynch *et al.* concluded that expansion of morbidity and medicalisation of more serious forms of morbidity was occurring. Leibson *et al.* found no compression occurring for coronary heart disease, mixed results for stroke, and support for compression of morbidity for cancers. However, a major disadvantage of using first time hospitalisation for a given condition (s) to represent onset of morbidity is that there is no potential for recovery.

Hessler *et al.* use a different measure to represent morbidity prior to death: the number of days spent in hospital in the last year of life (Hessler, Eriksson *et al.* 2003). This is examined for three different cohorts, and for different ages at death. Their findings concurred with the studies of health care expenditure prior to death: individuals who

survived to ages 85 and above were relatively healthy and spent little time in hospital in the last year of life, whereas individuals who died at younger ages, between 70 and 85, spent far more time in hospital in the last year of life. However, as highlighted by Roos *et al.*, hospital use fails to capture all health service utilisation, especially for older decedents.

Roos *et al.* are the only study to explore *all* health service utilisation, rather than just hospital utilisation, in a health system with universal health insurance (Roos, Montgomery *et al.* 1987). They examined hospital utilisation, nursing home utilisation, and number of physician visits in the years leading up to death. Their results suggested a shorter period of increased hospital use prior to death for the oldest decedents. This is compatible with the compression of morbidity hypothesis. Hospital bed days in 1973 were compared with the chance of dying in the eight subsequent years, using Least Squares Regression. For deaths at ages 45-74, deaths in all eight years after 1973 were significantly associated with bed days in 1973. At ages 75-84 this had reduced to the 6 years following 1973, while at ages 85+ only deaths in the year following 1973 were significantly associated with hospital bed days in 1973. This association was stronger the closer to the observation year (1973). A strong positive relationship was found between nursing home utilisation and age. Nursing home utilisation was found to be more spread out in the 4 years prior to death than hospital use. Females had considerably (80 to 100 percent) higher hospital utilisation than males at ages 65+, possibly due to increased widowhood of females compared to males resulting from higher female life expectancy. Regarding physician visits, the third indicator considered by Roos *et al.*, an inverse relationship was found with age: the very elderly had fewer physician visits. As has been mentioned, despite limited hospital bed days and physician visits for elderly decedents in the years prior to death, health care expenditure is still much higher than for other ages at death due to nursing home use at older ages. However, there was also evidence that a small number of decedents were consuming a disproportionate amount of care. The combined findings led Roos *et al.* to state:

> "Clearly, a substantial number of elderly individuals die without a prolonged period of illness and disability, or at least their deaths do not place major demands on a health care system which has few barriers to utilization by the elderly" (p245)

The findings from research in the field of hospital use prior to death highlight the importance of choice of method in the validity of results. Researchers from health

economics and public health backgrounds have different approaches to examining hospital utilisation prior to death. The importance of health service use *other* than hospitalisation in contributing to total health service utilisation prior to death has been brought into focus by Roos *et al.* While it is not possible to examine utilisation in other sections of the New Zealand health system given the data available, attempts will be made to take account of other forms of health service use in the discussion section of this paper (Section 4).

1.6 Conclusion

The use of administrative health data in medical research is an area poorly understood by the public in many developed countries. When appropriate safeguards and procedures are followed and the data treated with due respect, the privacy, confidentiality and indeed dignity of the individuals in the dataset can be maintained. The potential to health research of linking administrative health records is immense. The changing profile of cause of death, increasing health system complexity, and increasing population mobility, is making record linkage increasingly necessary. There is however relatively little existing research using longitudinal methods to examine the compression and expansion of morbidity theories. This paper attempts to rectify this deficit, and to improve the validity of the findings from the cross-sectional analyses in Paper Two. The next Section outlines the process of conducting record linkage of New Zealand hospital and mortality records, and the methods used to analyse hospital use in the months prior to death.

SECTION 2 – DATA AND METHODS: Conducting Record Linkage and Description of Two Longitudinal Techniques

2.1 Introduction

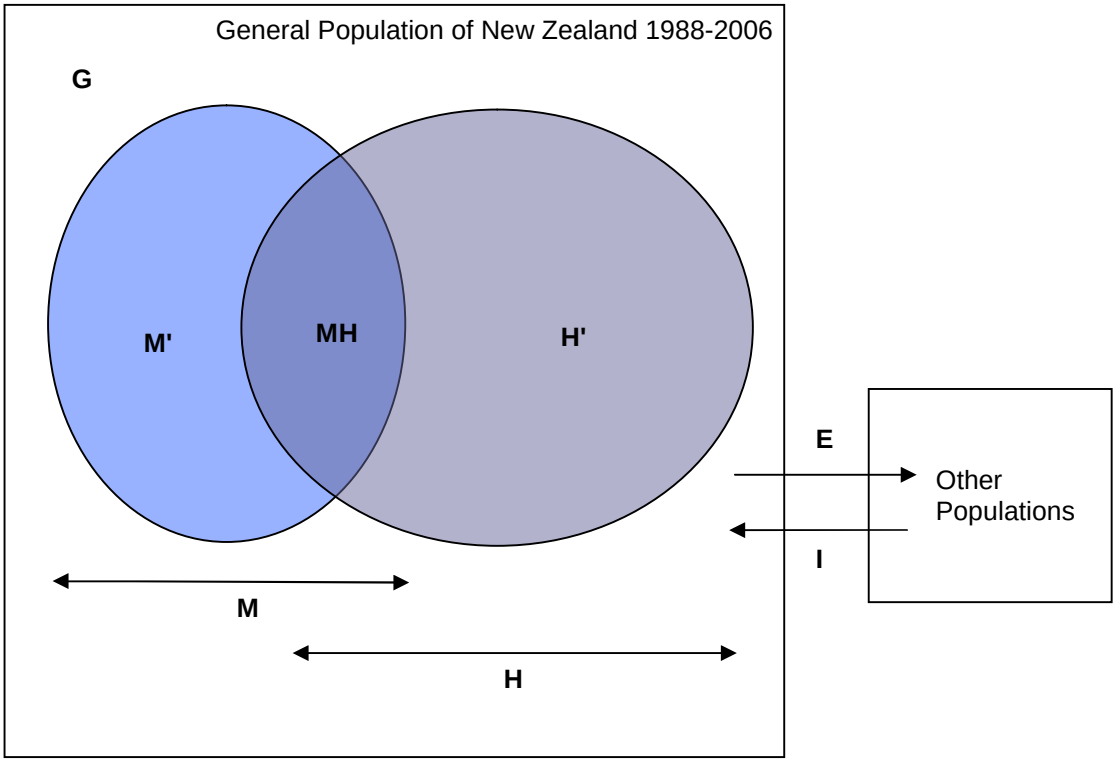
In the last two papers, cross-sectional analyses have been conducted on hospital and mortality data: in the first paper age-specific and age-standardised rates were presented, and in the second paper life table techniques were employed to calculate Hospital Utilisation Expectancies (HUEs) and life expectancies. This paper uses the same hospital and mortality data, but takes a rather different approach, using record linkage and longitudinal analyses. The datasets used include an encrypted National Health Index (NHI) number. While this could not be used with any other datasets due to anonymity and ethical reasons, it is possible to identify individuals both within the hospital dataset (several hospital events relating to the same person) and between the hospital and the morbidity datasets (hospital and mortality events relating to the same person). In the resultant linked datasets, individual information remains anonymous.

Longitudinal approaches are an invaluable, but are often unavailable tool for examining the compression of morbidity hypothesis. Rather than using cross-sectional data to *imply* the morbidity experience of individuals in the last few years of life, it is possible to extract the *actual* morbidity experience of individuals who have died. Two different techniques are used in this paper, reference point modelling, and person-month based logistic regression. Reference point modelling looks back from the date of death to examine hospital use in the months leading up to death. Person-month logistic regression explores the relationship between the probability of dying in the current month and hospital use in a previous month. This section discusses the application of these techniques, and then briefly describes the steps necessary to conduct record linkage (deterministic and probabilistic matching) and to prepare the datasets for analyses. More detailed explanations can be found in Appendix B.

2.2 Conceptual Diagram

Given the complex use of the datasets in this paper it is useful at this stage to present a conceptual diagram of the datasets used (see Figure 2.2.1).

Figure 2.2.1: Conceptual Diagram of the Datasets Used in the Analysis



Where:

G is the general population of New Zealand between 1988 and 2006.
M represents individuals with a mortality record during the time period (excluding 2001).
H represents individuals with a hospital record during the time period.
MH represents individuals with both mortality and hospital records.
M' represents individuals with a mortality record, but no hospital record.
H' represents individuals with a hospital record, but no mortality record.
E represents individuals who emigrated out of the New Zealand population during the time period.
I represents individuals who immigrated into the New Zealand population during the time period.

The analyses in Papers One and Two used the datasets **M** and **H**. In this paper the analyses will use a combination of **MH** (linked hospital and mortality records), **M'** and **H'**. Ideally the aim is to be able to generalise the findings at the population level (in this case **G**, the general population of New Zealand between 1988 and 2006). However there are

many people in the New Zealand population who might not have experienced a hospital admission or mortality event during the time period and so are missed from the datasets. Individuals may emigrate part way through the time period, with subsequent hospital or mortality events occurring in a different country, and therefore not included in the datasets used. Likewise individuals may immigrate into the New Zealand population part way through the time period with previous hospital events occurring elsewhere. A necessary assumption is that emigration from, and immigration to New Zealand occurred randomly over the time period.

It is useful to refer back to Figure 2.2.1 later on in this section when considering the data preparation and the types of analysis conducted in this paper.

2.3 Reference Point Modelling

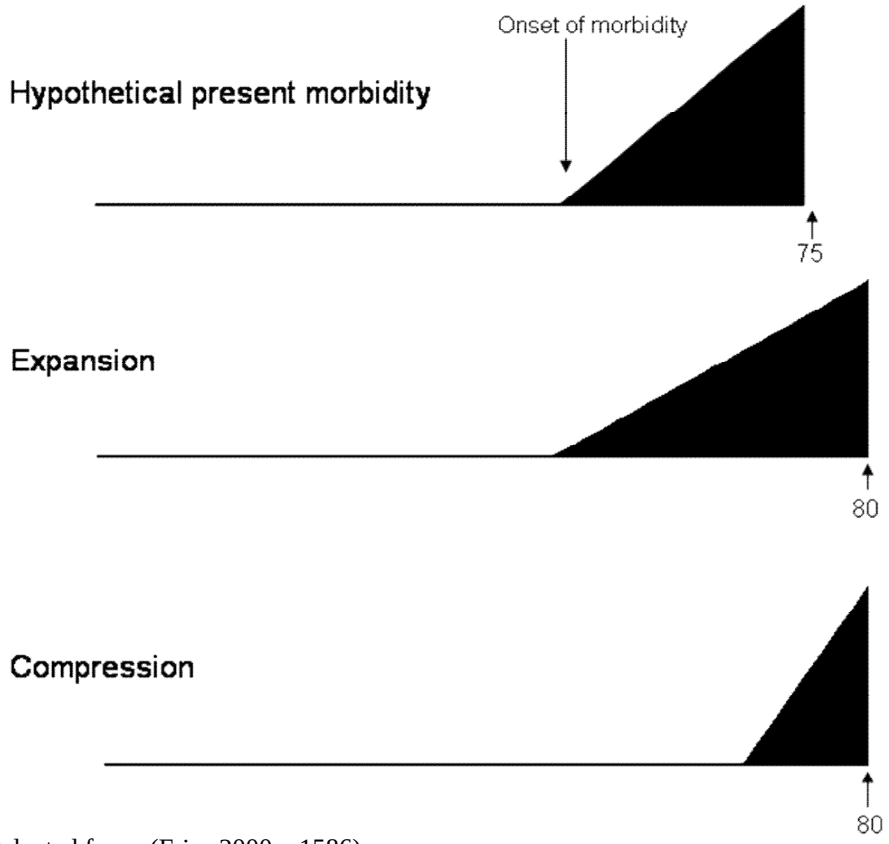
Reference point modelling uses record linkage to track public hospital bed days in the last months of life, for individuals who died during the observation period. Record linkage requires an encrypted NHI number, which is unique to each individual across the mortality and hospital datasets, and is available from 1988 onwards. Permission to link records was sought from the data providers, the Population Studies Centre at the University of Waikato, and a proposal for this research underwent formal approval at the University of Southampton, School of Social Sciences Ethics Committee. The data are held in a secure facility, as detailed in Paper One.

The rationale behind a linkage study is to identify both the time lag between the onset of morbidity and subsequent mortality, and the distribution of morbidity over that period (between onset of morbidity and death). Figure 2.3.1 illustrates the compression/expansion of morbidity debate in this context.

The average age at onset of morbidity is identified (the arrow above the horizontal line), followed several years later by the age at death (the arrow below the horizontal line), see Figure 2.3.1. The compression/expansion of morbidity relates to the difference between these average ages, at onset of morbidity and at death. However the level of morbidity is not constant between onset and mortality. Taking the y axis to represent severity of morbidity, the rising gradient between onset of morbidity and death implies rising severity and prevalence of morbidity at a population level. It is useful to further disaggregate death

by cause: different causes of death are associated with different patterns of morbidity in the period prior to death.

Figure 2.3.1: Conceptual Diagram of the Compression/Expansion of Morbidity Debate



Source: Adapted from (Fries 2000, p1586)

Reference point modelling aims to replicate this conceptual diagram. Taking only individuals who died in a given observation window, the average number of hospital bed days in each month prior to death is calculated, going back two years from the date of death[20]. This is then plotted on a line graph, and it can be seen in which month prior to death hospital use begins to rise (the point of onset of morbidity). This analysis is conducted separately for males and females and for different ages at death, deprivation deciles, and causes of death. Comparisons are also made over time. Reference point modelling takes the average number of bed days spent in hospital by *all* individuals who died, not just those with matched hospital records. The data used for reference point

[20] A month is defined as a period of 28 days. A period of two years equates to 24 'months' or 674 days, and so is slightly shorter than two calendar years (730 days).

modelling comes from **MH** and **M',** as represented in Figure 2.2.1, the conceptual diagram of the datasets.

We know that a large proportion of ill health and use of health services is concentrated into the last few months, and years of life: the compression/expansion of morbidity debates rely heavily on this assumption. However, most research examines health indicators at different ages, regardless of age at death. In comparison, this research examines the distribution of hospital use in the last few months of life, the tempo and timing of which is unique to each individual. There is very little precedent for using hospital and mortality data in this way. Previous research examining the compression/expansion of morbidity hypotheses using record-linked hospital and mortality data were discussed in the previous section. Only two studies used hospital bed days as a proxy for morbidity (Roos, Montgomery *et al.* 1987; Hessler, Eriksson *et al.* 2003) and these examined the average number of bed days in each year prior to death. In contrast, this study breaks down hospital use prior to death into months.

Due to record linkage between the mortality and hospital datasets, there is no need for a population denominator. New Zealand population data is limited at ages over 85 years and (when it is presented) is often aggregated to thousands. This is due to New Zealand's small population, which results in anonymity concerns arising from small numbers of deaths in some age groups. This study is not bound by such limitations. Results can be presented for ages over 85 years, as long as there are enough individuals at these ages to minimise random variation. However, it is also at these ages that the highest proportion of individuals are in nursing and care homes, and may be missed from public hospital data. A secondary examination of the proportion of the population in nursing and care homes is conducted in the discussion section of this paper (Section 4).

Reference point modelling only uses hospital data for individuals who have died. An individual's hospital utilisation is likely to increase as they age, so some modest increase in hospital utilisation over a two year period is to be expected. To resolve this problem and identify how much of the increase in hospital use over a two year period prior to death is due to ageing and how much is due to proximity to death, a baseline level of hospital use is calculated. This is achieved by taking the distribution in age at death (in days) from the **M** dataset, and randomly applying it to the **M'**, **MH** and **H'** datasets. This is *all* the data available (which in this case acts as a proxy for **G**, the general population), including the linked records. This gives every individual present in the datasets a 'fake' date of death, with the same age distribution as the observed deaths. After checking that

the individual has not already died by their fake date of death, reference point modelling is conducted on the fake date of death to give the baseline level of hospital use in each month prior to death, in the absence of proximity to death. This baseline level of hospital use is calculated separately for males and females, and for different deprivation deciles.

While the encrypted NHI is available from 1988, deprivation information can only be calculated from 1991. With a hospital utilisation window of two years prior to death, this national analysis can be conducted for individuals who died between 1990 and 2006 (excluding 2001)[21] and the deprivation analysis can be conducted for individuals who died between 1993 and 2006 (excluding 2001). The length of the window was determined by exploratory analysis of hospital use in the years prior to death. Prior to two years before death there is very little deviation from average hospital use at any given age. Record linkage is necessary for the identification of individuals both within the hospital dataset, and between the hospital and the mortality datasets. This process, and other data manipulation is discussed in more detail later in this section and in Appendix B.

Reference point modelling, while a good visual aid to examining the patterns of morbidity in the months prior to death, is hard to quantify. It is difficult to accurately pinpoint in which month before death hospital utilisation begins to rise, that is, in which month before death hospital utilisation starts to be significantly associated with death. While an attempt is made to incorporate individuals who did not die (through the calculation of baseline hospital use in each month prior to death) this is still a relatively crude approach.

It is beneficial to approach the data from a different angle, using different techniques that utilise **H'**, hospital events for individuals who do *not* die over the time period. This is achieved through logistic regression modelling of person-month data.

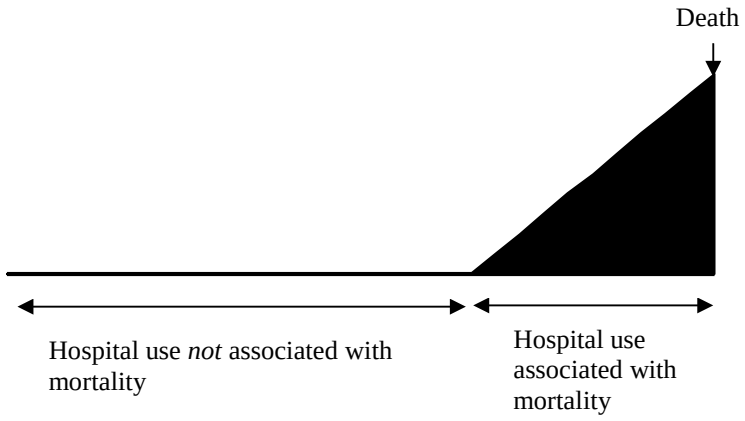
2.4 Person-Month Logistic Regression Modelling

It can be fairly confidently assumed that the probability of dying in the current month is significantly associated with the number of days spent in hospital last month. In the months prior to death, hospital use increases. But is there a significant relationship between the probability of dying in the current month and hospital use in the month one year previously? When does this relationship cease to be significant? At which month in the

[21] Mortality data for 2001 was missing from the original dataset.

past does hospital use fail to be a significant predictor of death in the current month? The point at which hospital use in a previous month is no longer significantly related to the probability of death in the current month is a statistical representation of the point where morbidity prior to death begins to increase, the 'onset of morbidity' denoted by Fries in Figure 2.2.1. This interpretation of the Fries diagram is shown in Figure 2.4.1.

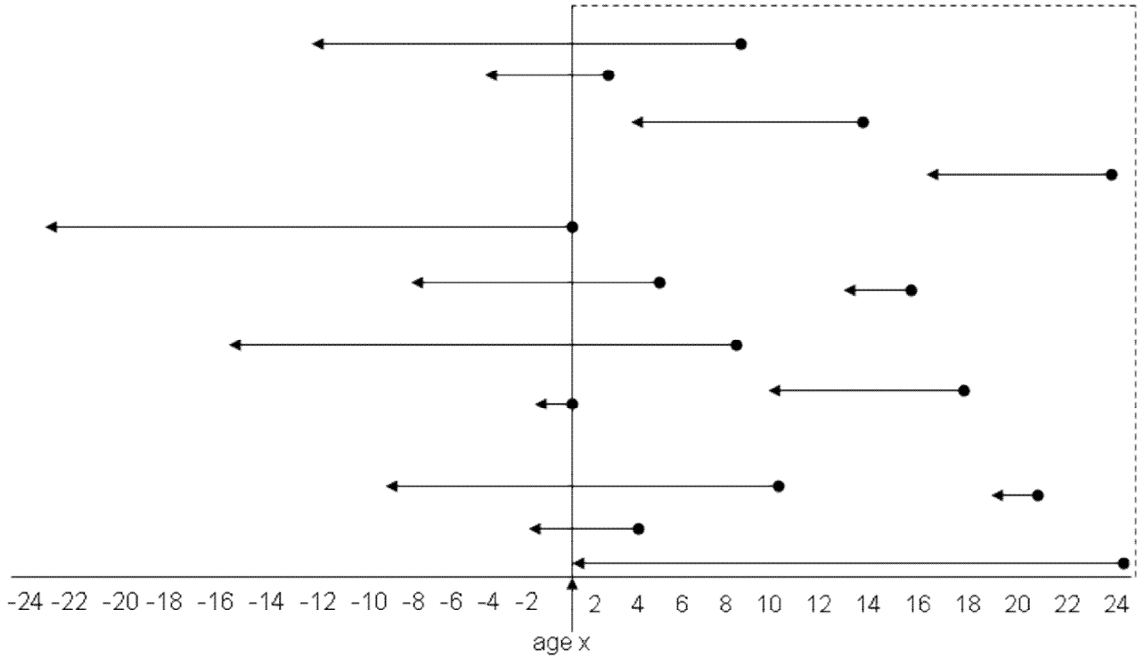
Figure 2.4.1: Interpretation of Compression of Morbidity for Person-Month Logistic Regression



Logistic regression modelling of data that has been converted to a person-month format is used to test whether hospital use in the past is significantly associated with death in the current month. All available datasets are used for person-month logistic regression, mortality (**M'**), hospital (**H'**) and the linked mortality and hospital data (**MH**). The mortality and hospital datasets used in this research are in person-event format: the mortality dataset contains one line per individual, denoting a mortality event and the hospital dataset contains one line per hospital event (of which there may be multiple events for each individual). In the merged dataset, every line contains either a hospital event, or a mortality event, or both. In preparation for person-month logistic regression the data need to be converted from person-event to person-month format, where each line represents a period of 28 days for an individual and each individual has the same number of lines, regardless of the number of mortality or hospital events experienced over this time period. Aggregated hospital stay variables were created for each person-month, giving the number of days spent in hospital by that individual in each month, and also a dichotomous variable denoting a death (or not) in each person-month.

As with the reference point modelling, a window of 2 years (or 24 'months') is used for the analysis. Figure 2.4.2 shows how this works.

Figure 2.4.2: Person-Month logistic Regression, Conceptual Diagram



The dots represent the current month, and the arrow heads represent a previous month in which hospital use is observed, with months along the x axis. Individuals are followed for a period of 24 months from a given age (x) and it is observed whether they die or are still alive in each of these months. It is necessary to 'look back' 24 months from age x (the left side of Figure 2.4.2), but as can be seen the current month is always in the 24 months following age x (the right side of Figure 2.4.2)[22]. Once dead, an individual is no longer included in the analysis (if an individual died in the 12th month after age x, they are excluded from the analysis for subsequent months). The outcome variable for the logistic regression is the dichotomous variable: 0 indicating no death in current month (*and still alive*), and 1 indicating death in current month. There are two predictor variables. The first and most important predictor variable is the difference between observed and expected hospital bed days[23] in a given previous month, going back a maximum of 24 months.

The second predictor variable is the 'number' of the current month: the order of the current month in the 24 months following age x. This directly relates to age, an individual

[22] If the analysis was conducted using records from only the 24 months following age x, there would be different potential for how far back the previous month could be. If the current month was the first month following age x, the previous month could only be one month ago. While if the current month was the 24th month following age x, the previous month could be anywhere between one and 24 months ago. This would introduce bias by age, given that the number of the current month is directly related to age. In the first month following age x an individual is younger than in the 24th month following age x.

[23] Expected bed days are the gender- and year-specific average of bed days in a given person-month. This difference between observed and expected bed days is hereafter referred to as excess bed days.

in the first month after age x is younger than in the 24th month after age x. This logistic regression is conducted separately for each previous month (1 to 24).

The equation for the person-month logistic regression is given in Equation 2.4.1.

Equation 2.4.1: Person-Month Logistic Regression Model

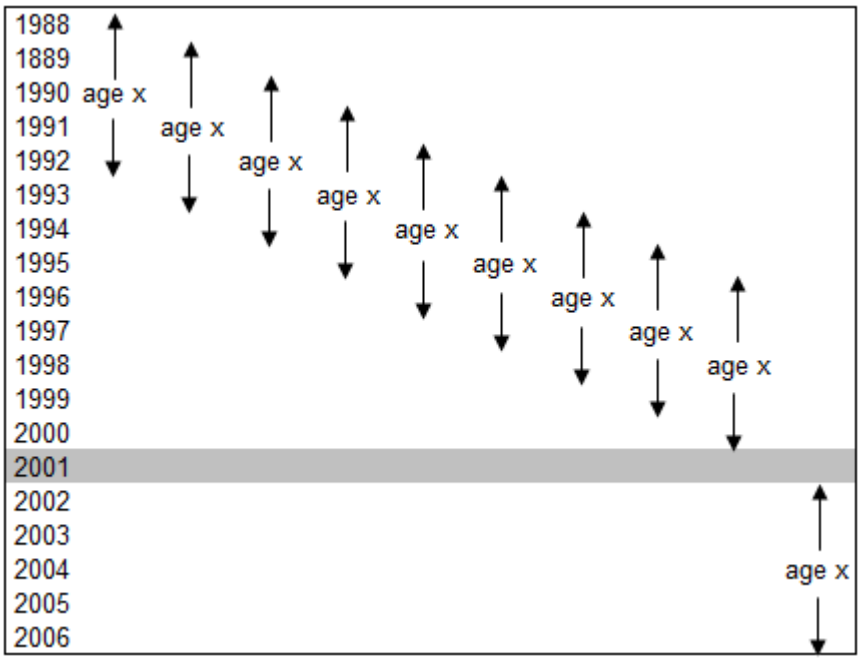
$$\begin{aligned}\text{Logit (death in current month)} &= \text{constant} \\ &+ \text{B1 excess bed days in a previous month} \\ &+ \text{B2 month 'number'}\end{aligned}$$

While the coefficient for month 'number' was rarely significant in the model, it was necessary to include it to control for age in the current month regardless of significance. Removing the month number from the model did not substantially change the results found. Thus both age in the previous month *and* age in the current month are controlled for, through the use of 'excess' bed days in a given previous month[24] and through the inclusion of the current month 'number' in the model.

The analysis is conducted for individuals attaining age x in a given year. Due to the need to examine the 24 months prior to age x as well as the 24 months following age x, a five year period of data is required. The long window of observation required and the missing mortality data in 2001 limit the years for which person-month logistic regression can be conducted. See Figure 2.4.3 for clarification. A period of 24 months[25] is slightly shorter than a period of 2 calendar years, thus Figure 2.4.3 is a simplification. Analysis of individuals turning 80 in 1990 does not include records going back to the first of January 1988 and forward to the 31st of December 1992. Instead it includes records going back to the 29th of February 1988 and forward to the 2nd November 1992 (1st January 1990 minus 672 days and 31st December 1990 plus 672 days).

[24] Excess bed days take into account the average number of bed days in a given month for an individual of a given sex and age in a given year.
[25] A month consists of 28 days.

Figure 2.4.3: Available Date Range for Person-Month Logistic Regression*



*Mortality data for 2001 missing

Death in the current month is a very rare event. It is unlikely for a person to die in a given month, even at older ages. Person-month logistic regression is a novel approach that utilises the strengths of population-level data: this analysis could not be conducted using sample data due to the rarity of the outcome variable. However despite using population-level data it was not possible to conduct analyses by deprivation or cause. Even splitting deprivation deciles into two groups (1 to 5 and 6 to 10) and only examining the main causes of death (cardiovascular and neoplasms) yielded too few outcomes (deaths) to produce reliable results. Person-month logistic regression was therefore only conducted at the national level, for males and females, attaining age x in 1990-1998 or 2004. Age x was limited to older ages where the most deaths occur: the analysis was repeated for a range of ages between 70 and 85 years.

While reference point modelling provides a valuable illustration of the patterns of hospital use prior to death over time, person-month logistic regression enables the relationship to be quantified and tests whether hospital use in the past is significantly associated with death in the current month. It is however limited by the rare nature of the outcome event (death in a given person-month) and can only be conducted at the national, all-cause level.

2.5 Data Preparation

Record linkage was necessary for both the reference point modelling and the person-month logistic regression. For the national analysis, 366,792 deaths that were registered in New Zealand between 1990 and 2006 (excluding 2001) could be linked to filtered public hospital records. This is 82.3 percent of all deaths registered over this period. For the deprivation analysis, 297,225 deaths registered between 1993 and 2006 (excluding 2001) could be linked. This is 82.5 percent of all deaths registered over this period. Evidently a high proportion of deaths in New Zealand were preceded by hospital stays in the years leading up to death.

Individual record linkage was conducted both between the mortality and hospital datasets, and within the hospital dataset. The main type of linkage was deterministic, using the encrypted NHI number. However this resulted in 14.4 percent of individuals with merged hospital and mortality records being excluded from the analysis. To reduce the number of mismatches, a combination of deterministic and probabilistic matching was carried out, reducing the percent of individuals excluded to 2.5 percent. Deterministic matching excluded individuals for whom different sexes were reported (either on different hospital events, or between hospital and mortality events). Probabilistic matching was conducted based on two criteria: date of birth and hospital discharges occurring after the date of death. The matching process is discussed in more detail in Appendix B. These individually linked data were used for both the reference point modelling and the person-month logistic regression analyses, but both methods required further processing prior to implementation.

Reference point modelling required the data to be reduced to one line per individual, regardless of the number of hospital stays. Individual hospital stays were aggregated to create length of stay variables for each month prior to death (rather than each hospital event). The aggregated length of stay variables for overlapping hospital stays were adjusted to include the number of days between the admission date of the first hospital stay and the discharge date of the last hospital stay in the overlap sequence. Duplicate hospital stays were identified and removed. The calculation of aggregated length of stay variables involved using SPSS syntax to turn the dataset ‘upside down’ and creating sequence variables for each individual. Date and repeat functions in SPSS were also used extensively to identify in which month prior to death a hospital stay fell and, where a hospital event

straddles two or more months prior to death, splitting the length of stay into the appropriate months prior to death. In order to calculate the baseline level of hospital use in the absence of mortality, a ‘fake’ date of death was applied to all individuals present in the datasets (regardless of whether they actually died in the observation window). A random number generator was used to apply the distribution of age at death in days from the individuals who died, to all individuals, thus giving a ‘fake’ death date. This was conducted separately for males, females, and deprivation deciles. This was then used to calculate hospital use in the months prior to the ‘fake’ death in the same way as for the actual, observed deaths, but only for individuals who had not already died by their ‘fake’ death date.

For the person-month logistic regression the dataset needed to be converted to 48 lines per individual, one for each of the 24 person-months preceding and following age x. To achieve this, length of stay variables for each of these months were calculated, and the dataset was reduced to one line per individual in the same way as for the reference point modelling. The data were then transposed using the SPSS variables to cases function, converting the data from one line per individual to one line per person-month. Average length of stay was calculated for each person month by taking the mean number of hospital bed days in each person month, by gender and year. Deducting the actual number of bed days in each person month from the average hospital bed days gave the excess number of bed days in each given person month. Lagged variables were computed to give the excess number of days spent in hospital in each previous month, the main predictor variable in the logistic regression model. Comparison of the date of death (where available) with the dates of each person-month prior and after age x was used to create a dichotomous variable, death in the current month, the outcome variable in the logistic regression model. An alive/dead variable was also created to indicate the alive/dead status of the individual in each person month. This was used to exclude individuals from the analysis once dead. This dataset was constructed for a given age x. The process was repeated to construct separate datasets for other ages.

Discussion of the data preparation for the reference point and the person-month logistic regression analyses is continued in greater detail in Appendix B, where the syntax for this data preparation is also presented.

2.6 Conclusion

The cross-sectional analyses conducted in Paper One and Paper Two provided a valuable insight into the interplay between hospital bed day rates and mortality rates in New Zealand. Cross-sectional analyses are not however ideal when examining the compression of morbidity hypothesis. Longitudinal analyses that follow individuals over time and track their health experiences prior to death are more suited to this use. This paper will present two very different approaches that utilise the strengths of the hospital and mortality datasets for compression of morbidity analyses. Both methods address the same research question: is compression of morbidity occurring in New Zealand? The reference point modelling answers this question subjectively, illustrating the intensity of morbidity (hospital use) in the months prior to death for different population groups. The person-month logistic regression takes a more objective approach, and attempts to identify the point at onset of morbidity prior to death. The methods used not only apply individual record linkage, thus enabling longitudinal analyses, but also take advantage of the large scale, population-level nature of the data through person-month logistic regression of a rare outcome variable. Using a range of methods enables the researcher to compare the findings of methods with different strengths and weaknesses, and in the process construct more accurate and reliable results than would be achieved through one method alone.

The next section will present the results from the analyses outlined above. These results will be examined with reference to the findings from Papers One and Two, and a comprehensive picture of hospital use in later life in New Zealand will be constructed.

SECTION 3 – RESULTS: Evidence for/against Compression of Morbidity from Reference Point Modelling and Person-Month Logistic Regression Analyses

3.1 Introduction

Both methods in this paper make use of individual record linkage to assess the evidence for compression of morbidity in New Zealand in the 1990s and early 2000s. Each method (reference point modelling and person-month logistic regression) brings separate strengths to the analyses, and examines the hypothesis from a different angle. Diverse methods are intended to strengthen the findings of this thesis. In this section the results of the reference point modelling are presented first, both nationally and by area deprivation. These results are disaggregated by sex, and the national results are further disaggregated by selected causes of death. This was not possible for the deprivation analysis due to numerical constraints. Conclusions are drawn from the reference point modelling before moving on to the person-month logistic regression. The person-month based logistic regression analysis is limited to national level all-cause disaggregated by sex due to the demanding data requirements of the method. Any further disaggregation resulted in too few observations to make the analysis viable. Following the conclusions from the person-month logistic regression, the results comparing both methods are presented. All analyses in this section use filtered hospital data (national or area deprivation filters, where appropriate), as detailed in Paper One.

3.2 Reference Point Modelling

The Fries framework on the compression of morbidity forms the analytical basis for the reference point modelling. The number of days spent in hospital in the months prior to death is calculated for individuals who died in the time period 1990 to 2006, going back 24 months (672 days) before the date of death. Hospital use of decedents is however heavily dependent on the disease circumstances (cause of death) and proximity to death. It is useful for comparison to also include the baseline level of hospital use that an individual would be expected to spend in hospital in the *absence* of proximity to death. Thus in Figures

3.2.1, 3.2.4 and 3.2.5 the lines labelled 'Male' and 'Female' are the average number of days spent in hospital in each month prior to death *for individuals who died*. These people are referred to as decedents. The lines labelled 'Male baseline' and 'Female baseline' show the actual number of days (on average) spent in hospital in each corresponding month for individuals who did *not* die in month 0. These people are referred to as survivors, and were allocated a *fake* date of death which was used to calculate the average hospital use in each month previously.[26] This allows the disaggregation of hospital utilisation associated with proximity to death, and hospital utilisation that would be expected in the *absence* of proximity to death. Where the actual and the baseline lines deviate represents the point at onset of morbidity associated with subsequent death.

Reference point modelling results are presented first at the national all-cause level, then at the national cause-specific level and finally at the area deprivation level. All results are disaggregated by sex, but only all-cause results are presented for the deprivation analysis due to small numbers. Finally, the results from the reference point modelling are discussed in the conclusion.

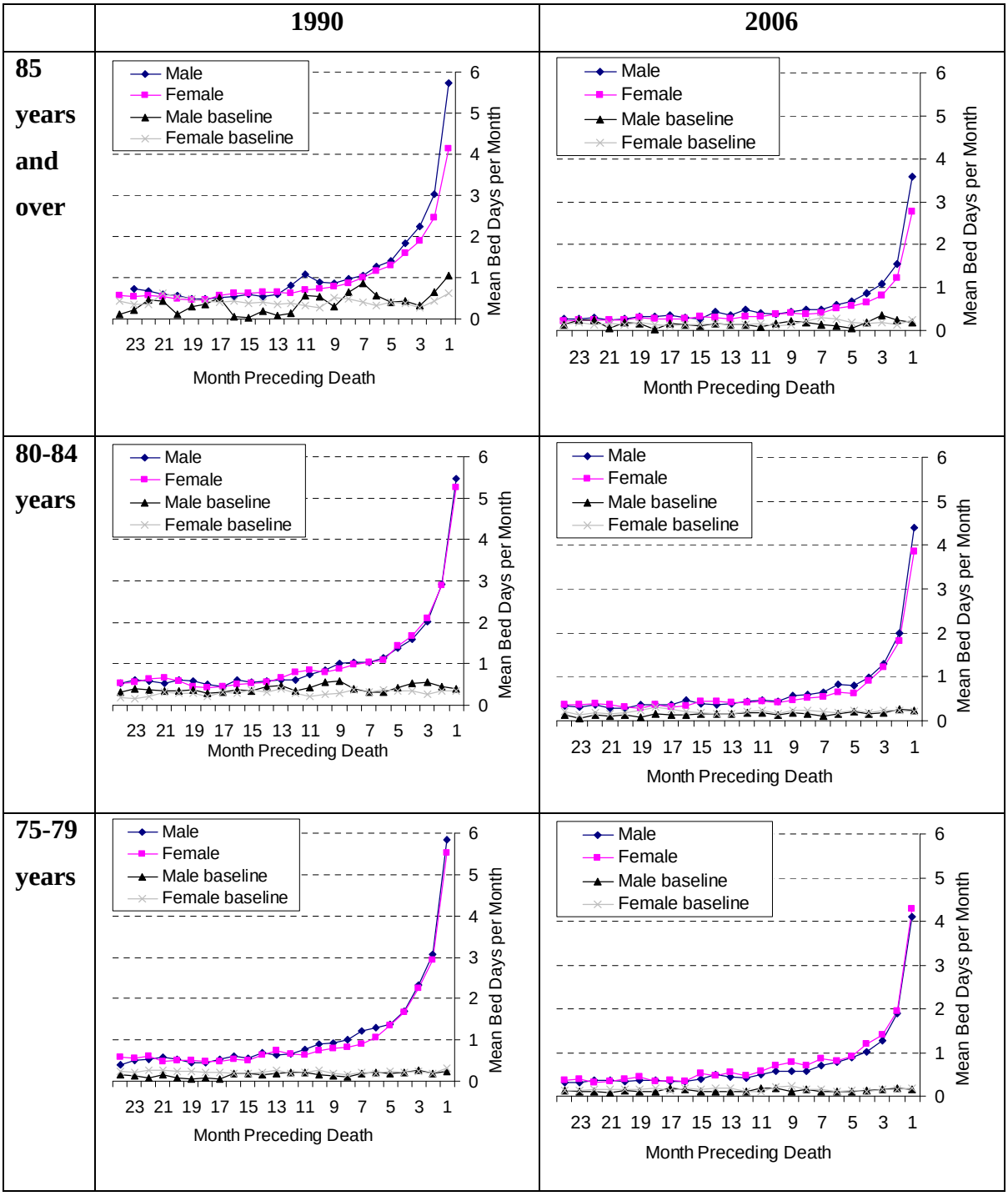
3.2.1 National Results, Hospital Bed Days in Months Prior to Death, 1990-2006

Paper One reported that hospital bed day rates decreased over time for both sexes and all ages except for the very young (0 to 4 years old). A similar trend is established in the reference point modelling analysis (see Figure 3.2.1), which shows decreasing hospital bed days in the months prior to death for death at different ages. Male decedents aged 85 years and over in 1990 could expect to spend on average 5.7 days in hospital in the last month of life. By 2006 this had dropped by more than a third to 3.6 days. Similar declines in hospital bed days in the last month of life are seen in other, younger age groups, but are generally of smaller magnitude. This decrease was not *just* seen in the few months immediately preceding death: females decedents aged 75 to 79 years in 1990 could expect to spend 0.6 days in hospital in the 24th month prior to death. By 2006 this had dropped to 0.4 days. Declines of greater magnitude are again found in the oldest age group, 85 years and over, where hospital use for females in the 24th month prior to death fell by almost two-thirds between 1990 and 2006, from 0.6 to 0.2 days. Although only the first and last year in the time period are shown in Figure 3.2.1, trends were relatively consistent across the time period. Consistent with the findings from Paper One, baseline hospital use in the absence

[26] These individuals may live for many years after their 'fake' date of death, or die the very next day. All that matters is that they are still alive on their randomly allocated 'fake' date of death.

of proximity to death also decreases over time, and becomes more stable with fewer random fluctuations. Sex-specific differences were also found. At the oldest ages, 85 years and over, males spend more time in hospital in the last month of life than females, while at younger ages there is little difference.

Figure 3.2.1: Mean Number of Bed Days in Months Prior to Death, Actual and Baseline*, 1990 and 2006, Selected Ages, by Sex



*Baseline bed days are the expected number of bed days in the *absence* of proximity to death

A large proportion of hospital use in the last 12 months of life is concentrated in the last 6 months, and particularly in the last month of life. Hospital bed days in the second month prior to death were approximately half of hospital bed days in the last month of life. For the oldest decedents (85 years and over and 80 to 84 years) the difference between hospital bed days in the first and second month prior to death was even higher, with a larger proportion of hospital use occurring in the last month of life compared to the second to last month of life.

One aim of the reference point modelling is to determine the point at onset of morbidity[27]. However it is difficult to determine in which month observed hospital bed days in the months prior to death diverge from the baseline level of hospital use in the absence of mortality. This is interpreted as the point at onset of morbidity. At younger ages at death (75 to 79 years) observed and baseline bed day rates appear to draw parallel at approximately 12 months prior to death, but observed bed days remain slightly higher than baseline bed days, and this does not appear to have changed much over time. At older ages the decrease in random fluctuation in both baseline and observed hospital bed days prior to death make it difficult to draw any conclusions about the point at onset of morbidity, although there does appear to be considerable convergence by about 7 months prior to death in both 1990 and 2006. This analysis provides little evidence for a change in the point of onset of morbidity (although estimating by eye is not a particularly objective measure). There is however evidence for decreasing *severity* of morbidity at older ages, suggesting compression of morbidity at older ages, but not through a shift in the point at onset of morbidity. This is easier to see in tabular form.

Average hospital bed days in each of the last 24 months of life for decedents in age groups 70-74 years and above decreased between 1990 and 2006 (see Table 3.2.1). This uniform decline in hospital bed days in the months prior to death was observed for both males and females, with just one exception: females dying aged 70 to 74 saw a 7 percent increase in hospital bed days in the 18th month prior to death over the time period (shown in bold in Table 3.2.1). A less consistent pattern was seen for decedents aged less than 70, however small numbers of deaths at many younger ages inhibit analysis.

The greatest declines in hospital use in the months prior to death were seen at the oldest ages. An examination of Table 3.2.1 reveals that over the time period 1990 to 2006

[27] Identifying the point at onset of morbidity is the rationale behind the inclusion of 'baseline' hospital bed days for survivors. Survivors are defined as individuals in the hospital and mortality datasets who are assigned fake dates of death based on the distribution of age at death of decedents. The point at onset of morbidity is the point at shich the gradients for survivors and decedents diverge.

deaths at ages 85 years and over had consistently larger declines in hospital bed days in the months prior to death than deaths at ages 70 to 74. This has only one exception: males in the 17th month prior to death.

There is evidence of smaller declines in mean hospital bed days in the months immediately prior to death compared to the months furthest from death, which suggests compression of morbidity into the months immediately prior to death. However this is not very consistent when compared across all months prior to death.

Table 3.2.1: Percentage change in Mean Hospital Bed Days in the Months Prior to Death between 1990 and 2006, for Deaths at Ages 70 Years and Above, by Sex

		Females				Males			
		70-74	75-79	80-84	85+	70-74	75-79	80-84	85+
Month Prior to Death	1	-17	-22	-27	-33	-19	-30	-20	-38
	2	-21	-33	-37	-50	-12	-38	-32	-49
	3	-17	-38	-42	-56	-19	-45	-37	-52
	4	-26	-28	-45	-59	-11	-40	-38	-53
	5	-27	-33	-56	-56	-25	-36	-42	-51
	6	-33	-22	-40	-56	-28	-40	-27	-53
	7	-20	-5	-47	-58	-48	-42	-38	-52
	8	-37	-14	-47	-57	-44	-44	-44	-50
	9	-31	-4	-47	-50	-38	-39	-44	-51
	10	-14	-6	-47	-48	-45	-37	-47	-58
	11	-15	-11	-50	-52	-36	-35	-37	-62
	12	-13	-28	-47	-46	-19	-35	-28	-42
	13	-31	-27	-39	-58	-13	-32	-37	-40
	14	-31	-28	-22	-55	-14	-28	-35	-20
	15	-34	-1	-17	-50	-20	-28	-30	-54
	16	-26	-37	-33	-56	-28	-44	-24	-48
	17	-6	-21	-30	-53	*-36*	-33	-19	-34
	18	**7**	-31	-16	-43	-17	-15	-29	-34
	19	-3	-12	-40	-38	-14	-21	-40	-33
	20	-13	-22	-48	-51	-18	-35	-53	-53
	21	-26	-32	-45	-56	-34	-39	-48	-59
	22	-31	-46	-40	-57	-24	-32	-38	-54
	23	-51	-29	-35	-58	-35	-37	-49	-63
	24	-42	-39	-34	-62	-43	-25	-37	-56

3.2.2 National Results by Cause, Hospital Bed Days in Months Prior to Death, 1990-2006

Over the time period 1990-2006, the cause-specific distribution of deaths changed. As was discussed in Paper One, the trend was for a more diverse pattern of causes of death: over the time period the two main causes of death, circulatory and neoplasms, saw a decline in

importance, and a larger proportion of deaths were due to a wide range of other causes (refer back to Section 3 in Paper One).

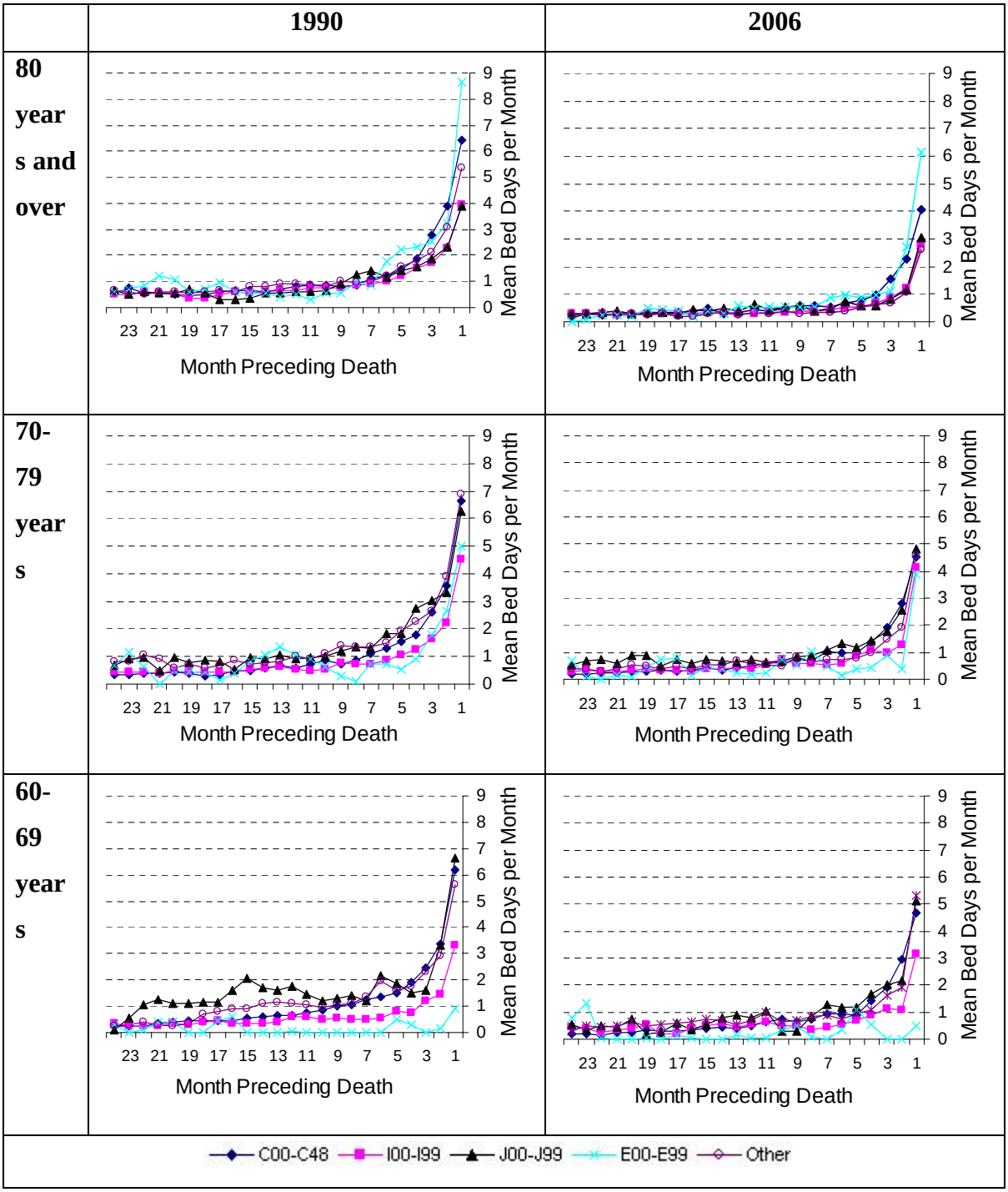
Different causes of death have different patterns of hospital use prior to death. Sudden causes of death, such as heart attacks and accidents, could be expected to be associated with relatively low hospital utilisation in the months leading up to death when compared to neoplasms, for example. Figure 3.2.2 shows the hospital use in the months prior to death for four leading causes of death over the time period 1990 to 2006.

Wider age groupings for ages 60 years and over are used due to small numbers and the results are presented for females only. Similar results were found for males, but deaths due to external causes had lower hospital use in the first few months prior to death for deaths at ages 80 years and over.

For young decedents (60 to 69) there is a clear cause-specific gradient in hospital use prior to death. External causes of death at ages 60 to 69 years are associated with the lowest hospital use in the six months immediately preceding death, followed by circulatory causes, with deaths to neoplasms and circulatory causes associated with similar, high levels of hospital use prior to death. This pattern is consistent over time. At ages 70-79 there is less difference in hospital use prior to death by cause of death in the last 6 months of life, with external causes of death associated with similar hospital use as circulatory causes.

For the oldest decedents, age group 80 years and over, the cause-specific pattern of hospital use in the last 6 months of life is markedly different. Hospital bed days due to external causes of death are higher than for all other causes of death. Deaths due to neoplasms are associated with the second highest hospital use in the last 6 months preceding death, while respiratory and circulatory causes are associated with very similar, low levels of hospital use. This pattern is also prevalent for males, but bed days due to deaths from external causes are not quite as high compared to deaths from other causes. For deaths occurring at the oldest ages (80+), other causes of death were associated with relatively low hospital use in the 6 months prior to death compared to the four leading causes of death, particularly in later years. For deaths at younger ages due to other causes however, hospital use in the last 6 months of life was as high, if not higher than for circulatory and respiratory causes of death.

Figure 3.2.2: Mean Number of Bed Days in Months Prior to Death, Four Leading Causes of Death*, 1990 and 2006, Females



*ICD10 Codes:

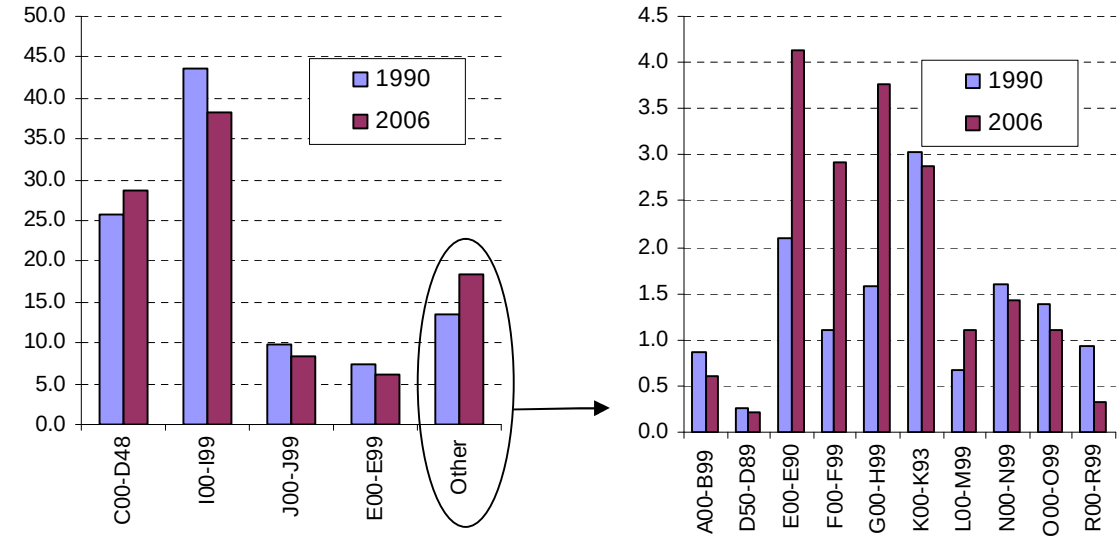
C00-C48 Neoplasms, **I00-I99** Circulatory, **J00-J99** Respiratory**, E00-E99** External Causes.

Prior to about 6 months before death, hospital use by cause of death tends to converge. This is seen for all age groups and time periods, however for younger decedents (60 to 69), hospital use in these months tends to be higher and more fluctuating than for

older decedents, particularly for respiratory causes of death. These fluctuations settle down over time, consistent with the trend of decreasing hospital use over time.

Neoplasms, Circulatory, Respiratory and External causes were the three most common causes of death over the time period 1990 to 2006 (see Figure 3.2.3).

Figure 3.2.3: Percentage of Deaths by Cause*, All Ages, Both Sexes, 1990 and 2006



* ICD10 codes

A00-B99 Certain infectious and parasitic
C00-D48 Neoplasms
D50-D89 Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90 Endocrine, nutritional and metabolic diseases
F00-F99 Mental and behavioural disorders
G00-H99 Diseases of the nervous system and sense organs
I00-I99 Circulatory
J00-J99 Diseases of the respiratory system
K00-K93 Diseases of the digestive system
L00-M99 Diseases of the skin, subcutaneous tissue, muscuoskeletal system and connective tissue
N00-N99 Diseases of the genitourinary system
O00-O99 Pregnancy, childbirth and the puerperium, certain conditions originating in the perinatal period, and congenital abnormalities
R00-R99 Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
E00-E99 External causes of morbidity and mortality

The proportion of deaths due to neoplasms increased by 11.2 percent over the time period, from 25.4 to 28.7 percent of all deaths. The proportion of deaths due to the other three main causes decreased by between 12.3 percent (for circulatory) and 16.0 percent (for external causes). Circulatory causes accounted for 43.6 percent of all deaths in 1990; by 2006 this had dropped to 38.3 percent. In 2006 respiratory and external causes accounted for 8.4 and 6.1 percent of deaths respectively. Other causes of death increased from 13.3 to 18.5 percent of all deaths, an increase of over a third. This increase in other causes of death was predominantly driven by the increasing proportion of deaths due to endocrine, metabolic and nutritional disorders (including diabetes), mental and behavioural

disorders, and diseases of the nervous system and sense organs. These three causes of death together accounted for 10.8 percent of all deaths in 2006, an increase from 4.8 percent in 1990. Unfortunately because of small numbers of deaths due to these causes, and the diverse nature of these causes rendering grouping nonsensical, it is not possible to examine hospital utilisation prior to death for these causes alone.

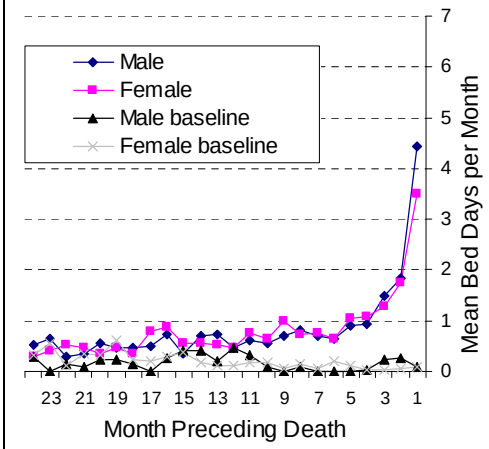
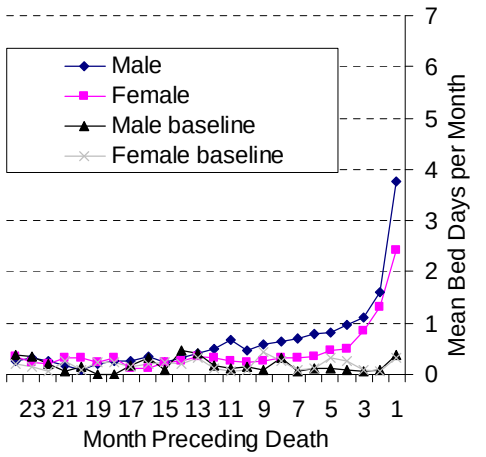
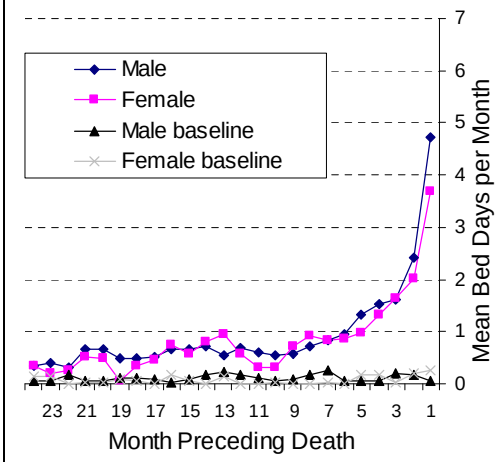
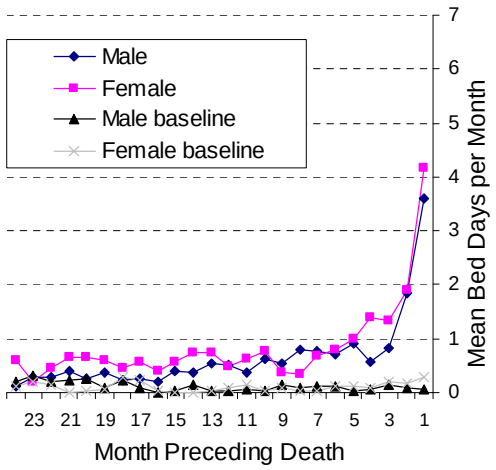
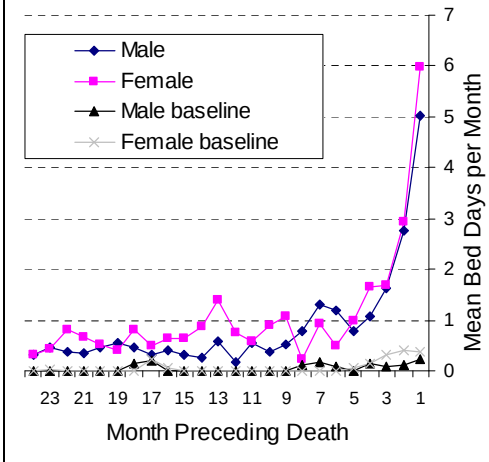
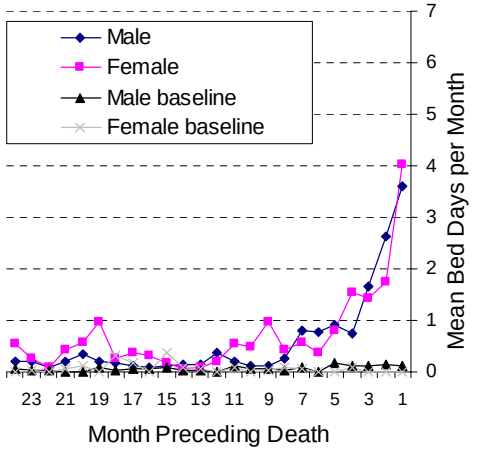
National trends in hospital utilisation prior to death by age, sex and cause have been presented. We now move on to the results from the area deprivation analysis.

3.2.3 Area Deprivation Results, Hospital Bed Days in Months Prior to Death, 1993-2006

Hospital use in the months prior to death is now examined by area deprivation decile. Due to the relatively small numbers of deaths in each decile it is not possible to conduct this analysis by cause. Results are presented by broad age groupings and sex, for the start and end of the time period observed (1993 to 2006). The analysis is restricted to a shorter time period than for the national analysis due to the lack of area deprivation data until 1991. Figures 3.2.4 and 3.2.5 present the results for deprivation deciles 1 and 10 respectively, decile 1 representing the 10 percent of Census Area Units (CAUs) with the lowest deprivation scores, and decile 10 representing the 10 percent of CAUs with the highest deprivation scores. Baseline hospital use in the absence of mortality is included in Figures 3.2.4 and 3.2.5. This is specific to the deprivation decile, year, age group, and sex for which it is presented and allows approximation of the point at onset of morbidity (this is the point at which morbidity associated with proximity to death rises above the background, baseline level of morbidity not associated with proximity to death).

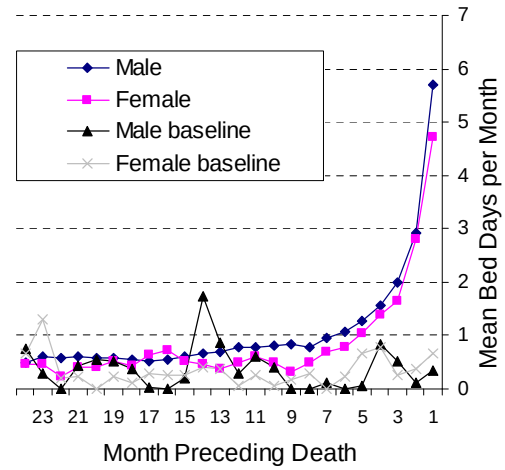
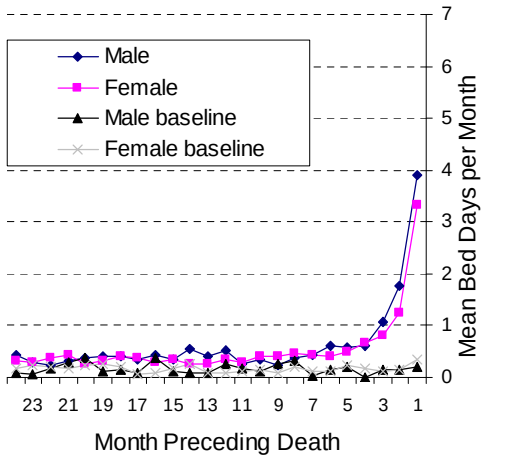
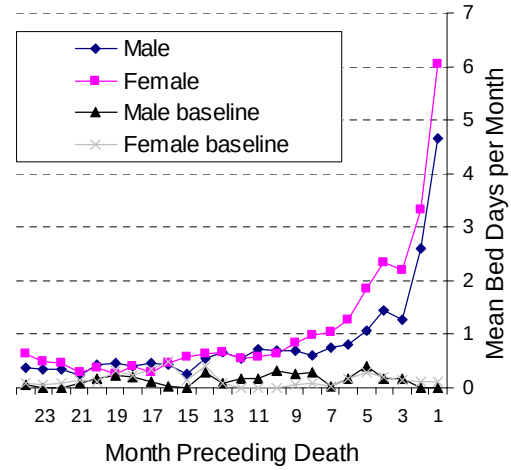
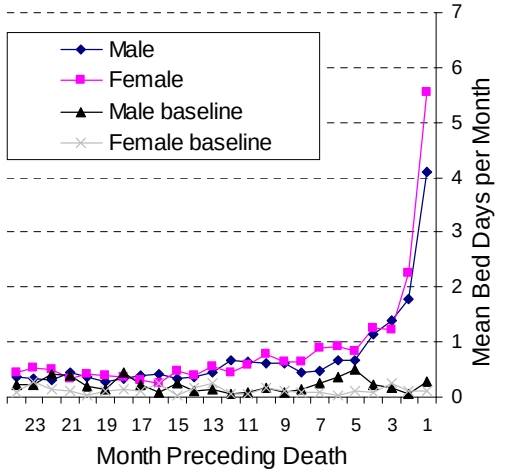
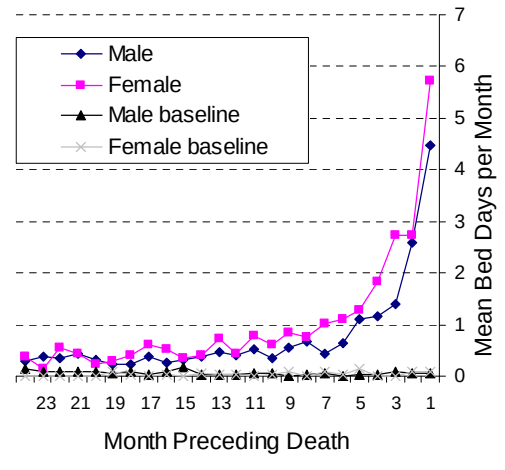
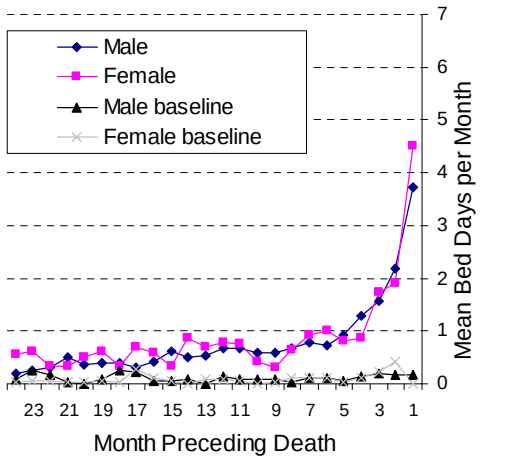
As was observed in the national results, at older ages of death (80 years and over) hospital use in the last month of life is higher for males than females, for both time periods and all deciles. A mixed picture is seen at ages 70 to 79, while at ages 60 to 69 the pattern is reversed: females have consistently higher hospital use in the last month of life than males. These sex-specific trends are seen regardless of deprivation decile or year.

Figure 3.2.4: Mean Number of Bed Days in Months Prior to Death, Actual and Baseline*, **Deprivation Decile 1**, 1993 and 2006, Selected Ages, by Sex

	Decile 1	
	1993	**2006**
80 years and over		
70-79 years		
60-69 years		

*Baseline bed days are the expected number of bed days in the *absence* of proximity to death

Figure 3.2.5: Mean Number of Bed Days in Months Prior to Death, Actual and Baseline*, **Deprivation Decile 10**, 1993 and 2006, Selected Ages, by Sex

	Decile 10	
	1993	**2006**
80 years and over		
70-79 years		
60-69 years		

*Baseline bed days are the expected number of bed days in the *absence* of proximity to death

For decedents aged 60 to 69 years there is little difference in hospital use in the months leading up to death by deprivation decile. However for older ages (70 to 79 years and 80 years and over) there is a gradient in hospital use by deprivation: increasing deprivation is associated with more days spent in hospital in the last two or three months of life. Although it is not possible to examine hospital utilisation in the months prior to death in more and less deprived areas by cause of death, it is useful to refer back to the cause-specific mortality rates reported in Paper One. These revealed that more deprived areas have less diversity in cause of death than less deprived areas, with a higher proportion of deaths due to the four leading causes: neoplasms, circulatory, respiratory and external causes. Circulatory external causes of death for older decedents are associated with higher hospital utilisation in the months prior to death than other causes of death (see Figure 3.2.2). This may partially account for the excess hospital bed days prior to death observed for older decedents in more deprived areas.

Towards the end of the time period this deprivation gradient is less apparent, and almost disappears for males dying in the oldest age group (80 years and over). Hospital use in the last month of life in 1993 for males dying at older ages (80 years and over) in decile 10 was 1.5 days higher than for decile 1, and for females this difference was 1.2 days. By 2006 hospital use for males in the last month of life in deciles 1 and 10 had converged at 3.8 and 2.9 days respectively, whereas female decedents in decile 10 aged 80 years and over still spent 0.9 days longer in hospital the last month of life than females in decile 10.

There is thus evidence of converging hospital use in the months immediately prior to death by deprivation decile over time, particularly at the oldest ages (which is where the gradient was largest). There is little evidence of such a gradient at younger ages (60 to 69 years), or at older ages prior to about 6 months before death.

The point of onset of morbidity is again difficult to define. The convergence between baseline and observed hospital bed days appears to occur slightly closer to death in 2006 than 1993, for all deciles and all ages. However this could be an artefact of the general pattern of declining observed hospital bed days in the months prior to death regardless of proximity to death, i.e. decreasing *severity* of morbidity. There is little discernable difference in this trend by deprivation decile. This is a positive finding, especially when coupled with the convergence between more and less deprived deciles of hospital use in the months immediately preceding death. It suggests that people dying in more deprived deciles are *not* experiencing an earlier point at onset of morbidity than people dying in less deprived deciles, *and* that the severity of morbidity (represented by

hospital use) after this point of onset in more deprived deciles is reducing to the same severity seen in less deprived deciles.

3.2.4 Conclusions from Reference Point Modelling

In conducting the reference point modelling, efforts were made to allow for the identification of the point at onset of morbidity. This included the construction of baseline hospital bed days for survivors in the months prior to death and the operationalisation of the point at onset of morbidity as the month prior to death in which baseline hospital bed days for survivors and actual hospital bed days for decedents converge. Despite this, identification of the point at onset of morbidity has been difficult to determine.

The all-cause national analysis echoed findings from Paper One: hospital use is decreasing over time, and from Paper Two: hospital use is decreasing fastest at the oldest ages over time. There was little evidence for a change in the point at onset of morbidity over time, but there *was* evidence for decreasing severity of morbidity (represented by hospital use) in the months prior to death. Further evidence for compression of morbidity was found when examining the magnitude of declines in hospital use in the months prior to death. Greater declines were found for months further from death, suggesting 'compression' of hospital use and thus morbidity into the months immediately prior to death. This evidence was however not particularly reliable due to large fluctuations in the magnitude of decreases from month to month.

The cause-specific analysis of the national results provides a background understanding to the all-cause decreases in hospital use over time. Different causes of death are associated with different patterns of hospital use prior to death and this *varied by age group*. External causes of death were associated with high hospital use in the months immediately preceding death for older decedents, but with low hospital use for younger decedents. An increasing proportion of deaths are due to neoplasms, yet hospital use prior to death for deaths from neoplasms has decreased substantially over the time period, for all ages at death examined. This results in a net decrease in all-cause hospital use prior to death. Hospital use in the months prior to death for deaths from circulatory causes did not decrease very much over the time period for deaths at most ages, yet this is offset by circulatory causes accounting for a decreasing proportion of deaths over time, resulting in a net decrease in all-cause hospital use prior to death.

It appears that there are no differences in the point at onset of morbidity by deprivation decile. Differences in hospital use in the last few months of life by deprivation

decile were observed for older decedents, but these reduced over time. The severity of morbidity (level of hospital utilisation) prior to death in more deprived deciles is improving to the levels seen in less deprived deciles.

The main limitation of the reference point modelling is the difficulty encountered in attempting to identify the point at onset of morbidity. Results from the person-month logistic regression, which aims to quantify the point at onset of morbidity, are now presented.

3.3 Person-Month Logistic Regression

Unlike reference point modelling, person-month based logistic regression quantifies the point at onset of morbidity. A slightly different operant is used, the point at onset of morbidity is conceptualised as the month in the past in which hospital use is no longer significantly associated with the chance of dying in the current month. Beyond this point (month) hospital use is not associated with subsequent mortality. Thus, if mortality in the current month is *not* associated with hospital use in the 12^{th} month previously, it is assumed to not be associated with hospital use in the 13^{th} and 14^{th} month previously either. As with reference point modelling, the point at which hospital use (morbidity) is no longer affected by proximity to death represents the point at onset of morbidity.

This section presents only one set of national results for the person-month logistic regression analysis. Further disaggregation by deprivation or cause was not possible due to small numbers of observations (deaths). The analysis is conducted only for the period 1990 to 1998 and 2004 (see Section 2 for more details).

3.3.1 National Results, Association between Hospital Use in a Previous Month and Death in the Current Month, 1990-1998 and 2004

Figures 3.3.1 to 3.3.3 present the results from the person-month logistic regression for three different ages, 75-77, 80-82 and 85-87. Findings are only presented for females, but are very similar to the findings for males. As individuals are followed for a period of 24 months (672 days) after attaining a specified age (e.g. 75 years), in reality the output represents individuals aged between 75 and 0 days and 75 and 672 days, and so on. For simplicity, these ages will be hereafter referred to as 75, 80 and 85, rather than 75-77, 80-82 and 85-87.

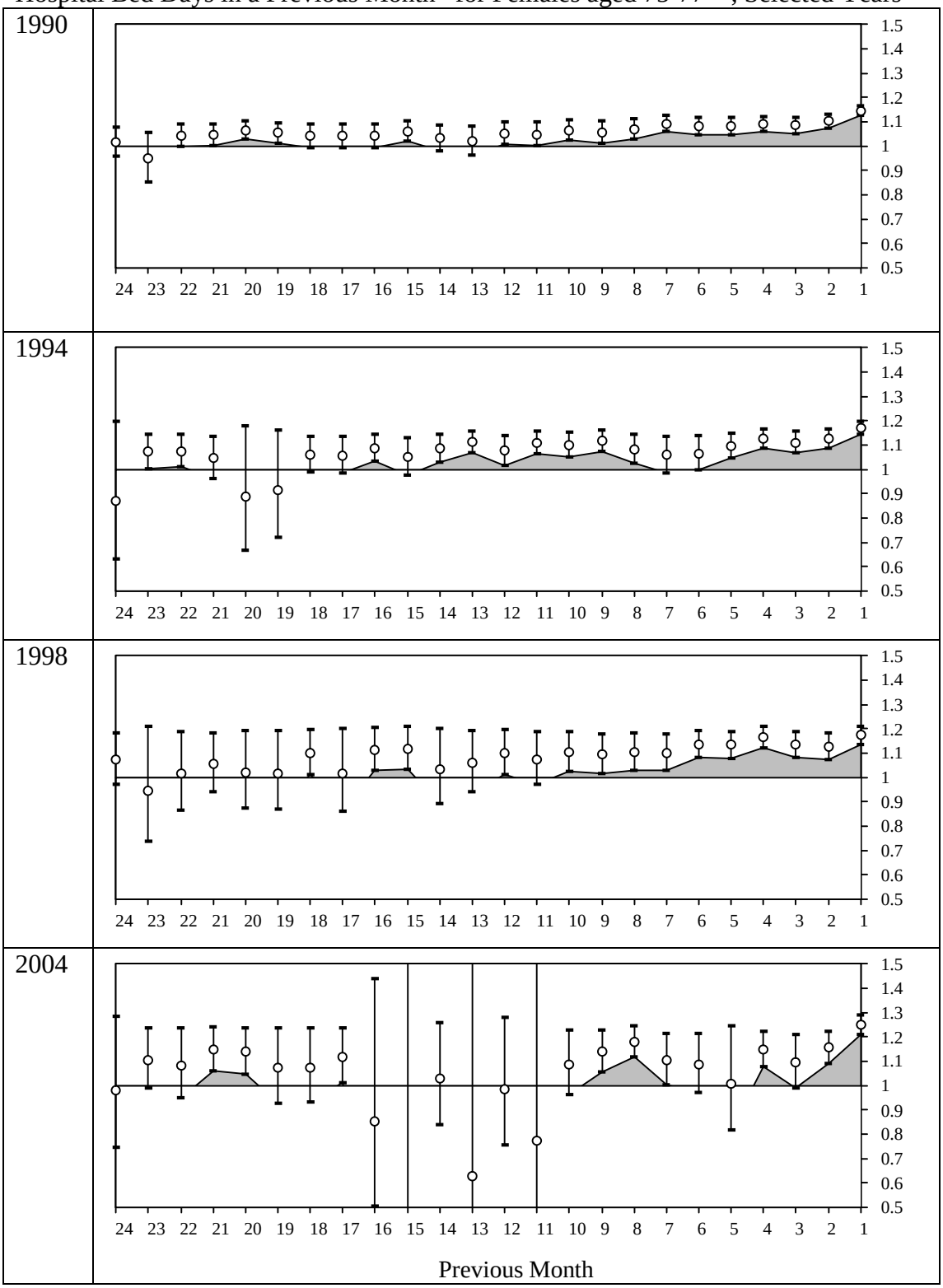
Before interpreting the graphs (Figures 3.3.1 to 3.3.3), it is necessary to discuss their construction. Each set of error bars represents a separate logistic regression model, one for each of the 24 months preceding the current month. The current month is fluid, and can be any month between age x (75, 80, or 85) and age x and 672 days. The number of the current month therefore relates directly to the age of the individual. For this reason, the month number of the current month is included in the model to adjust for age. The other predictor variable is the difference between the observed and expected number of days spent in hospital in a given previous month (going back a maximum of 24 months). This is hereafter referred to as the 'excess' hospital bed days in a previous month. The outcome variable is a dichotomous variable indicating the occurrence of death in the current month. Each set of error bars presents the odds ratio of death in the current month for each excess day in hospital in a previous month, with associated 95 percent confidence intervals.

The horizontal line across the graphs denotes an odds ratio of 1. If the lower confidence interval crosses this line, excess hospital use in that month is *not* significantly associated with the risk of dying in the current month.

So if a woman aged 75 in 1990 spent one excess day in hospital last month, her odds of dying in the current month, compared to someone who had spent the expected number of days in hospital last month, is 1.14 (see Figure 3.3.1). That is, she is 14 percent more likely to die in the current month for every excess day spent in hospital last month. The shading underneath the lower confidence interval is intended to directly replicate the Fries diagram. It highlights in which previous months excess hospital use is significantly associated with mortality in the current month (that is, the lower confidence interval for the odds ratio is above 1). If compression of morbidity was occurring, we would expect to see the grey areas become increasingly concentrated on the right side of the graphs, close to the current month.

Examining the figures on this basis it is rather unclear if compression is occurring. The first confidence interval to cross 1 does seem to move closer to the current month over time, but months much further back also appear to be significantly related to mortality in the current month. Only selected years are presented in figures 3.3.1 to 3.3.3, when all years are considered the trend appears even more haphazard. The pattern of error bars is very consistent in 1990, for all ages, but over time there are increasing fluctuations in both the size and the positioning of the confidence intervals.

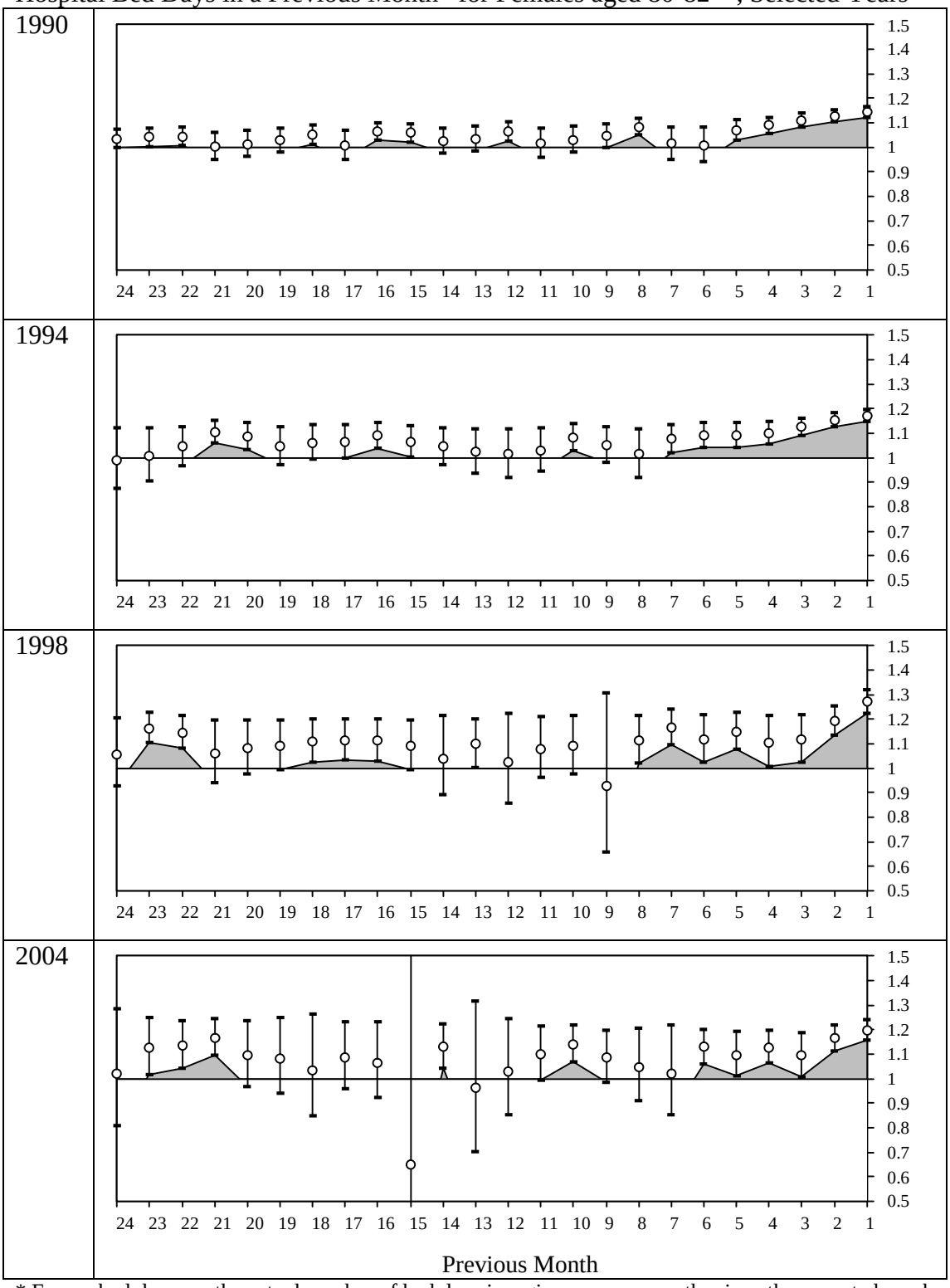
Figure 3.3.1: Odds Ratio and 95% C.I. of Death in the Current Month due to Excess Hospital Bed Days in a Previous Month* for Females aged 75-77**, Selected Years



* Excess bed days are the actual number of bed days in a given person-month minus the expected number of hospital bed days in that person-month for an individual of the same age and sex and turning 75 in the same year

** Individuals are aged 75 years and 0 days, to 75 years and 672 days (24 'months')

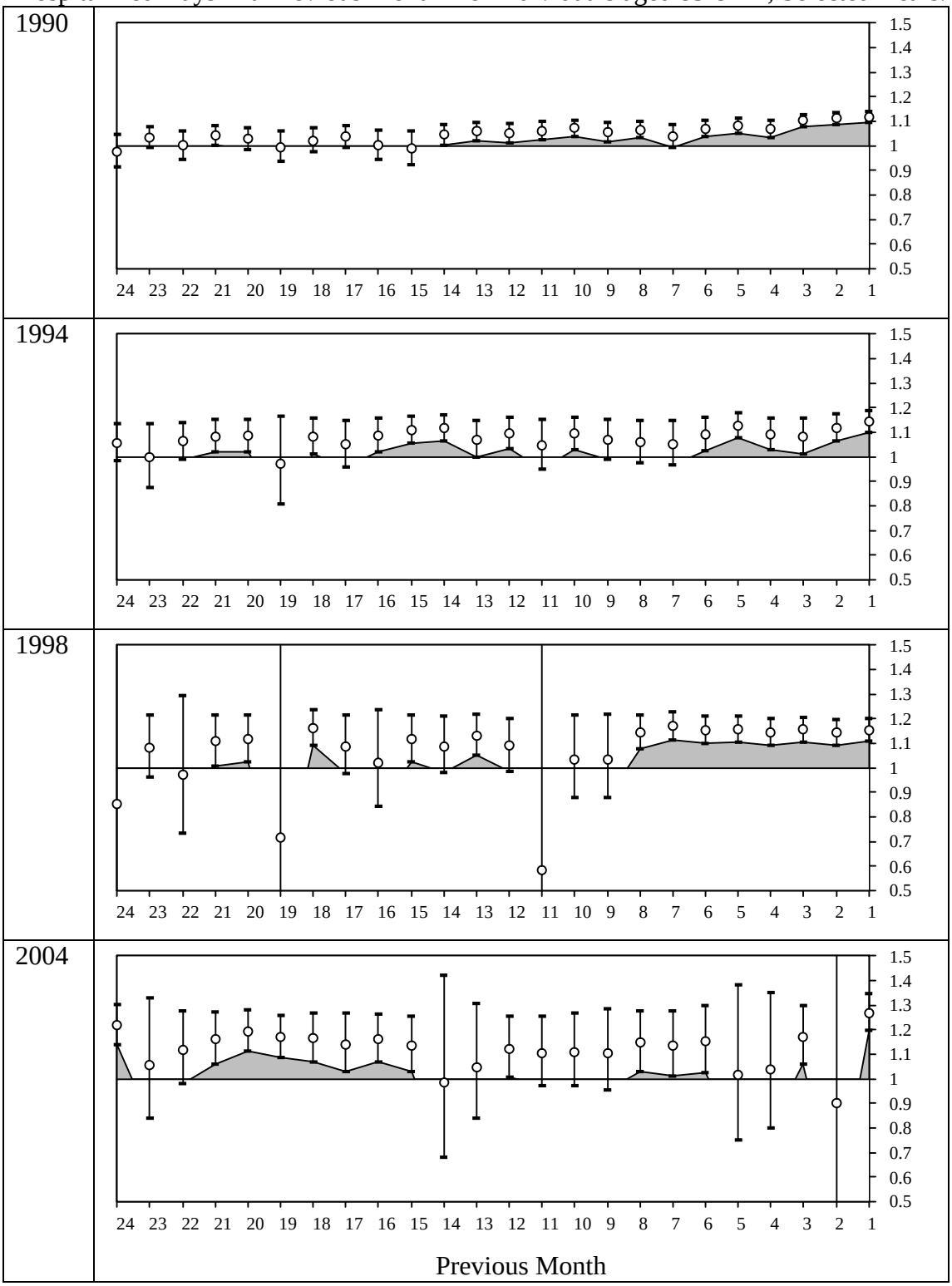
Figure 3.3.2: Odds Ratio and 95% C.I. of Death in the Current Month due to Excess Hospital Bed Days in a Previous Month* for Females aged 80-82**, Selected Years



* Excess bed days are the actual number of bed days in a given person-month minus the expected number of hospital bed days in that person-month for an individual of the same age and gender and turning 80 in the same year

** Individuals are aged 80 years and 0 days, to 80 years and 672 days (24 'months')

Figure 3.3.3: Odds Ratio and 95% C.I. of Death in the Current Month due to Excess Hospital Bed Days in a Previous Month* for individuals aged 85-87**, Selected Years.



* Excess bed days are the actual number of bed days in a given person-month minus the expected number of hospital bed days in that person-month for an individual of the same age and gender and turning 85 in the same year

** Individuals are aged 85 years and 0 days, to 85 years and 672 days (24 'months')

It is clear that neither support for, or evidence against the compression of morbidity hypothesis can be drawn from the person-month logistic regression modelling. Despite the fluctuations there are however two observations that can be drawn from these figures. Firstly, the further back the previous month, the wider the confidence interval tends to be. This is logical, in the first few months prior to the current month hospital use is strongly associated with death in the current month. This is reflected both in the high odds ratio, and the narrow confidence intervals seen in the first previous month for all ages and years. The confidence interval in the first month is never seen to cross 1. Excess hospital use in months further away from the current month are not as strongly associated with death in the current month, and this is reflected by both the lower odds ratios and the wider confidence intervals observed.

The second observation is that the confidence intervals get wider over time, even in the first month prior to death. While the confidence interval for the first month prior to the current month remains significant, it gets wider. This cannot be explained by a smaller number of deaths in later years: there is variation in the number of deaths from year to year but fluctuations in confidence interval width do not correspond to years with low numbers of deaths. It can however be explained by small numbers of certain *types* of observations. Each observation is a month in an individual's life, and each individual is followed for 24 months. There are two options, the individual can die in the current month or not, and they can spend time in hospital in a specified previous month or not. These observations can therefore be divided into four categories, as shown in Table 3.3.1.

Both death in the current month *and* hospital use in a specified previous month are very rare events. The reason for the increasing confidence intervals over time is a decreasing number of observations in the last of these categories: where there is a death in the current month and hospital utilisation in the previous month (1, 1). As hospital utilisation decreases both over time (as found in Papers One and Two) and in the months prior to death (as found in the reference point modelling), the number of observations in this category decreases. When the number of observations is compared to the size of confidence intervals in previous months, they are found to be strongly associated. After a suitable transformation (inverse, cubic or s-curve) the number of observations in the 1, 1 category is found to account for around 75 percent of the variation in confidence interval width.

Table 3.3.1: Categories of Observation* from the Person-Month Logistic Regression Analysis

	No death in current month	Death in current month
No hospital use in previous month	0, 0	1, 0
Hospital use in previous month	0, 1	1, 1

*Each observation is a month of an individual's life. There are four categories of observation:

0, 0 – No death in the current month and *no* hospital use in the previous month. Very common.

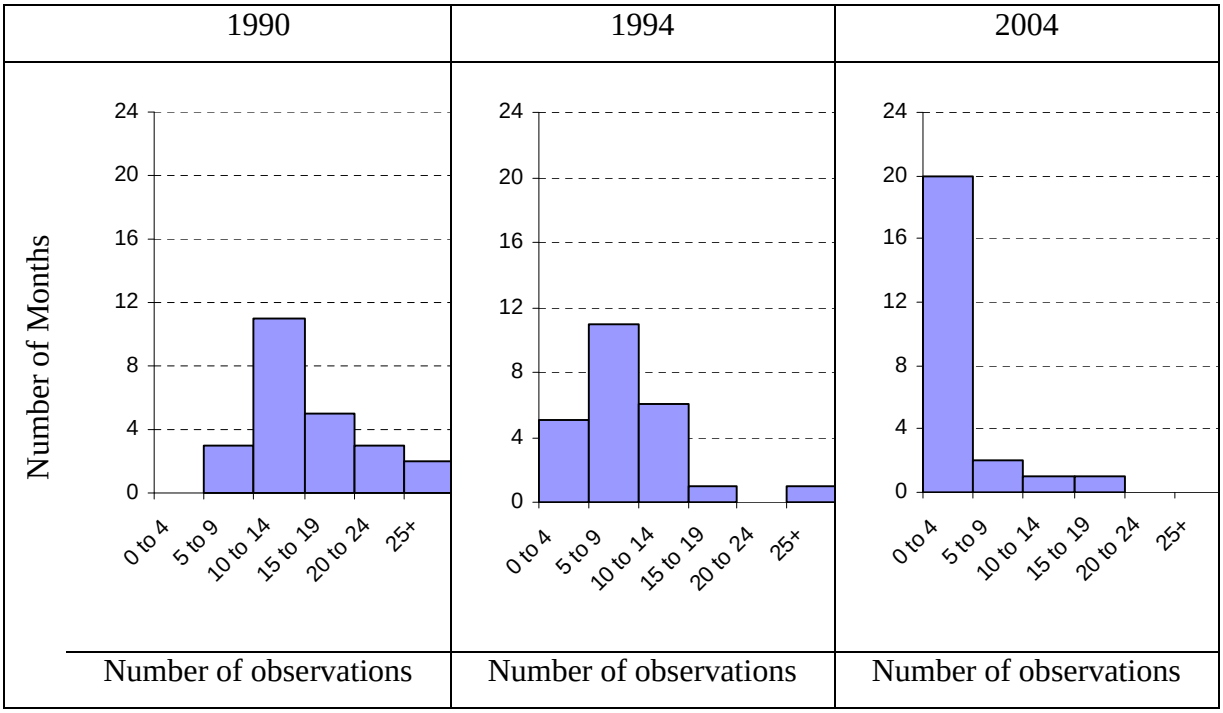
0, 1 – No death in the current month and hospital use in the previous month. Common.

1, 0 – Death in the current month and *no* hospital use in the previous month. Uncommon.

1, 1 – Death in the current month and hospital use in the previous month. Very uncommon.

The decreasing number of 1, 1 observations over time (where death in the current month is 1 and there are more observed than expected hospital bed days in the previous month (i.e. excess hospital bed days are positive)) is shown in Figure 3.3.4, for females aged 75 to 77. In 1990 there were no previous months where the number of 1, 1 observations was less than five, and two previous months where the number of observations was 25 or higher. However in 2004 most previous months had five 1, 1 observations or less (20 out of 24 previous months) including one month with no 1, 1 observations, and no previous months where the number of observations was 20 or higher.

Figure 3.3.4: The Changing Distribution of the Number of 1, 1 Observations* in each of the 24 Previous Months. Females aged 75-77*, Frequency Diagrams, Selected Years.



* A 1, 1 observation is where there is a death in the current month and hospital use in the previous month (higher observed than expected hospital bed days).

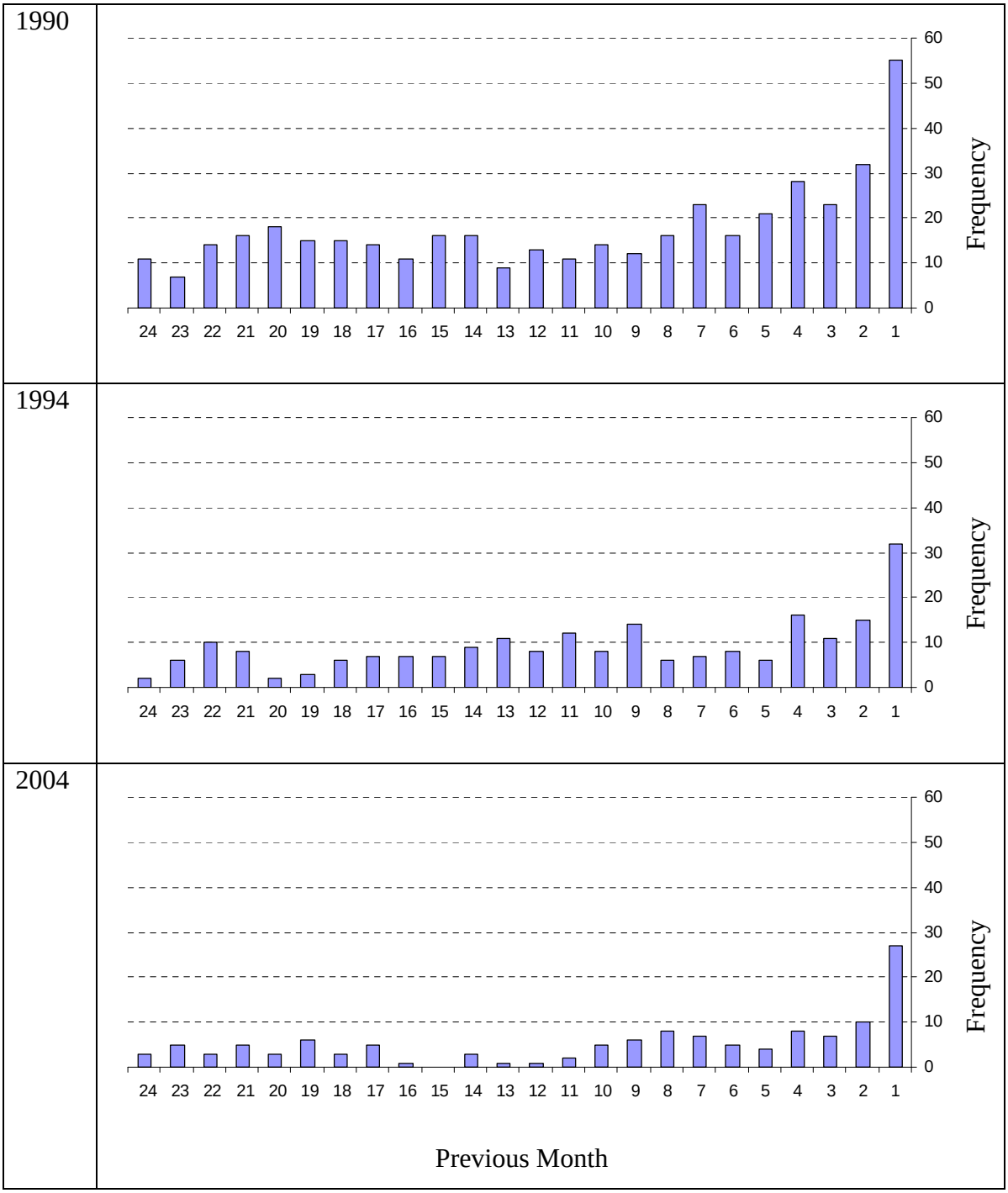
** Individuals are aged 75 years and 0 days, to 75 years and 672 days (24 'months')

The number of 1, 1 observations (where there is a death in the current month and hospital use in a previous month) is higher in the months immediately prior to the current month. This mirrors the findings from the reference point modelling. Thus the number of 1, 1 observations decreases the further back the previous month is (see Figure 3.3.5). It can be seen that previous months in which there were few 1, 1 observations have wide confidence intervals (compare Figure 3.3.5 with Figure 3.3.1). Examining Figure 3.3.5, previous month 23 in 1990, months 19, 20 and 24 in 1994, and months 11-16 in 2004 are seen to have fewer 1, 1 observations than neighbouring months. This corresponds with particularly large confidence intervals for these months in Figure 3.3.1.

It is worth noting however that for the *first* month previously the odds ratio of an additional day in hospital on the chances of dying in the current month *increased* significantly between 1990 and 2004. This occurred despite a decrease in the number of 1, 1 observations where there was a death in the current month and hospital use in the previous month. This was found to be true for ages 75-77 and 85-87 but the increase was not significant for ages 80-82. This is strong support for the strengthening of the

association between hospital utilisation last month and death in the current month, and thus support for compression of morbidity.

Figure 3.3.5: Number of 1, 1 Observations** in the 24 Previous Months. Females aged 75-77*, Bar Charts, Selected Years



* A 1, 1 observation is where there is a death in the current month and hospital use in the previous month (higher observed than expected hospital bed days).

** Individuals are aged 75 years and 0 days, to 75 years and 672 days (24 'months')

Whereas for other previous months decreasing numbers of 1, 1 observations have been associated with widening confidence intervals and lower odds ratios, in the first month prior to death there has been an increase in the odds ratio despite the number of 1, 1 observations decreasing by about 50 percent for females aged 75-77 (see Figure 3.3.5).

One solution to boost the numbers of 1,1 observations is to aggregate the units of analysis from months to quarters (periods of three months). This analysis was conducted, but the results are not shown here. Aggregating to person-quarter units resulted in a widening of confidence intervals over time, but fewer large fluctuations in confidence interval size. There was no evidence of a shift in the point at onset of morbidity.

3.3.2 Conclusions from Person-Month Logistic Regression

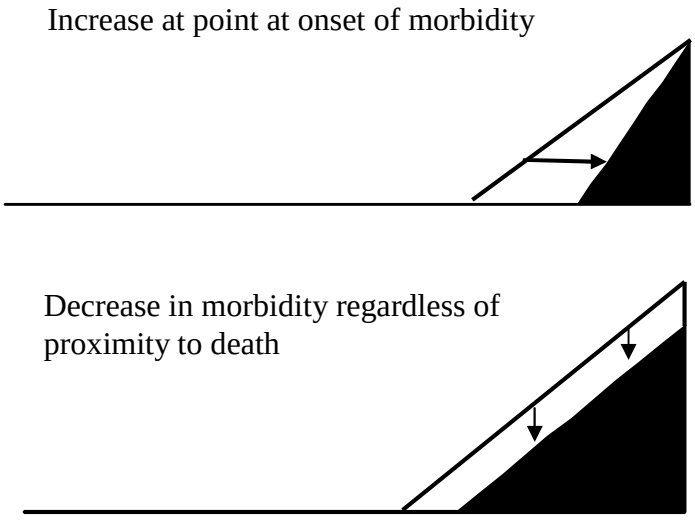
An examination of the logistic regression results did not yield any evidence of compression or expansion of morbidity (see Figures 3.3.1 to 3.3.3). However when the number of observations on which the regression analysis were based on was taken into consideration, some interesting findings emerged (see Figures 3.3.4, 3.3.5 and Table 3.3.1). The number of observations where there was a death in the current month *and* hospital use in the previous month (category 1, 1) was often very low, and was found to explain most (75 percent) of the variation in confidence interval size. However the odds ratio of death for additional excess days spent in hospital in the first month prior to death was found to increase, despite the number of observations in the 1, 1 category halving. Hospital utilisation in the months prior to death decreased, to the point where it was no longer significantly associated with death. This is in itself an important finding, suggestive of compression of morbidity. This could however represent either a function of decreasing relevance of hospital utilisation as a predictor of mortality, possibly through increased substitution of care, or a function of decreasing *severity* of morbidity (number of hospital bed days) rendering the prediction or mortality statistically insignificant. If the latter is true there is strong evidence for compression of morbidity, but further examination is necessary before such a conclusion can be arrived at. There could be other explanations for the decrease in hospital utilisation. An increase in substitution of health care from a public hospital setting to other settings, such as residential care homes would result in decreases in hospital utilisation that do *not* represent improvements in morbidity. The possibility for substitution of care is explored in Section 4.

3.5 Conclusion

The research conducted in this Paper is a reminder of the necessity to approach a research question from several different angles, using a range of techniques.

The reference point modelling found tentative support for compression of morbidity, not through an increase in age at onset of morbidity, but through a decrease in the *severity* of morbidity (hospital bed days) in the 24 months preceding death. Figure 3.4.1 demonstrates this type of decrease. It was however very difficult to visually assess the point at onset of morbidity. It was demonstrated that hospital use in the months prior to death varies substantially by cause of death, and by age. Convergence of hospital use in the months prior to death was seen between more and less deprived deciles, particularly for older men. Similar patterns in hospital use prior to death were found for men and women, with the exception that males dying in the oldest age group tended to have higher hospital use than females in the last month of life. It was also demonstrated that hospital use increases in the months leading up to death.

Figure 3.4.1: Compression of Morbidity Pathways



The person-month logistic regression analysis found little evidence either for or against compression of morbidity. No evidence was found of a shift in the point at onset of morbidity. Confidence intervals for the odds ratio of excess hospital use in a previous

month on chances of dying in the current month were found to get bigger further from the current month, and were also found to increase and become erratic over time. This trend was largely explained by differences in the number of observations where there is a death in the current month and hospital use in a previous month: confidence intervals were large when there were few of these observations. While the logistic regression itself did not yield any conclusions, an examination of the number of these observations suggested compression of morbidity was occurring. There are caveats associated with this conclusion however. As with the reference point modelling, the decrease in hospital utilisation could either represent improvements in health or the loss of individuals from the formal (public) health care system through substitution of public hospital care for other sources of health care. Mechanisms for substitution of care include residential care homes (hospital use is declining fastest at the oldest ages), private health insurance (although in Paper Two this explanation was discounted), and facilitation of palliative care and death in the home.

The next section examines the possible explanations for this phenomenon and assesses the relevance of the findings so far in light of this new evidence.

SECTION 4 – DISCUSSION

4.1 Introduction

The analyses in this paper have provided a diverse range of results. Both the reference point modelling and the logistic regression of person-month data aimed to answer the same question: is compression of morbidity occurring in New Zealand? While the reference point modelling graphically replicates the Fries diagram, denoting *severity* of morbidity (time spent in hospital) in each month prior to death, the person-month logistic regression analysis aimed to quantify the point at onset of morbidity.

The reference point modelling found evidence for the compression of morbidity hypothesis through decreasing hospital bed days in the months prior to death. While the results from the person-month based logistic regression were inconclusive, an examination of the reasons for these results suggested compression of morbidity through a decrease in severity of morbidity. However for both the reference point modelling and the person-month logistic regression, the possibility of substitution of care from public hospital to other settings needs to be examined. If substitution of care is occurring this would result in a decreasing number of hospital bed days prior to death, but would *not* represent a decrease in morbidity. In this section the results from the reference point modelling and the person-month logistic regression are reiterated and evidence for substitution of care (nursing home utilisation and the proportion of deaths taking place in public hospitals) are explored. It has been found previously (Paper Two) that private health insurance does not explain the declining public hospital bed day rates in New Zealand.

4.2 Reference Point Modelling, Discussion of Results

The results from the reference point modelling that are presented in the previous section examined hospital utilisation (a proxy for morbidity) in the months prior to death, and explored how this had changed over time, and how this varied by deprivation decile and age at death. This analysis was conducted in the context of the compression/expansion of morbidity hypotheses: decreasing hospital bed days prior to death would indicate compression whereas increasing hospital bed days in the months prior to death would

indicate expansion. The point at onset of morbidity was also examined, to identify in which month prior to death decedents first experienced raised hospital utilisation associated with proximity to death. If the point at onset of morbidity moved closer to death, this would indicate compression of morbidity, if it moved further away, expansion.

The results supported the compression of morbidity hypothesis, but through a reduction in severity of morbidity (hospital bed days) prior to death rather than through changes in the point at onset of morbidity. Cause-specific analysis identified how the interplay between the proportion of deaths attributable to different causes, and the average hospital utilisation prior to death for different causes of death have combined to result in lower all-cause hospital utilisation in the months prior to death. Favourable results for the compression of morbidity hypothesis were also found from the area deprivation analysis, which provided evidence of convergence of morbidity prior to death for elderly decedents in different deprivation deciles.

The results from the reference point modelling reinforced findings from Papers One and Two and provided support for the compression of morbidity hypothesis for both males and females, all levels of deprivation, and most causes of death. The analysis did not end here however. One further method, person-month based logistic regression was implemented in an attempt to quantify the point at onset of morbidity (a major limitation of reference point modelling).

4.3 Person-Month Logistic Regression, Discussion of Results

Logistic regression of data in person-month format was conducted to strengthen the validity of findings from the reference point modelling. The same research questions were addressed as for the reference point modelling (is compression of morbidity occurring over the time period, and how does this vary by age at death) but analysis by deprivation decile was not possible. Unlike the reference point modelling, which examined both the severity of morbidity *and* the point at onset of morbidity, the person-month logistic regression analysis focussed solely on the point at onset of the morbidity aspect of compression/expansion of morbidity. The point at onset of morbidity was difficult to determine from the reference point modelling

The findings from the person-month based logistic regression of current mortality based on previous hospital utilisation were inconclusive. No evidence was found either for

or against compression of morbidity (based on the point at onset of morbidity), but evidence was found to suggest that hospital utilisation became less associated with mortality over the time period, indicated by widening and increasingly erratic confidence intervals. This second finding was however dismissed when the number of different types of observation was examined. The width of confidence intervals was found to be largely explained by the number of observations where there was a death in the current month *and* hospital use in the previous month (observation type 1, 1. See Table 3.3.1 on page 189), which was often very small. Using this information, two points of interest emerged. Firstly, the odds ratio of death in the current month for each excess day spent in hospital in the first month previously increased significantly over the time period for two of the three age groups examined, despite large decreases in the number of 1, 1 observations. This indicates compression of morbidity into the first month prior to death. Secondly, decreases in hospital utilisation over the time period (due to decreased severity of morbidity) resulted in too few observations for hospital utilisation in many previous months to be a significant predictor of death in the current month. This again denotes compression of morbidity.

There is however an important caveat for these conclusions. Decreasing hospital utilisation over the time period (especially at older ages), may result from improvements in morbidity *or* increased substitution of care. If substitution of care from public hospital to other settings is occurring the validity of hospital bed days as a proxy for morbidity would decrease. The validity of hospital bed days that have been filtered to represent ill health as a proxy for morbidity is the cornerstone of all analyses in this paper (and in Papers One and Two). Thus potential confounders which may have resulted in the declining association between hospital utilisation and subsequent mortality need to be examined. Substitution of care from public hospitals to private hospitals was examined and dismissed in Paper Two. Two further forms of substitution of care remain plausible: an increase in the proportion of the population resident in nursing homes, and an increase in the facilitation of palliative care and death in the home. Both of these theories would account for the disproportionate declines in bed days observed at older age groups. These theories are discussed below, with reference to the declining association between hospital utilisation and mortality.

4.4 Institutionalised Healthcare in New Zealand, Implications for Public Hospital Utilisation Prior to Death

One possible explanation for low hospital use particularly at older ages is substitution of care from public hospitals to other settings, such as long term residential care (nursing homes, also called retirement homes). If substitution of care is found to be increasing over the time period (particularly among the very old) this may explain the decreasing hospital utilisation of both survivors and decedents (results from the reference point modelling). It would also help explain the decreasing number of months prior to death with hospital utilisation (results from person-month logistic regression). However, long term residential care information is not included in the public hospital dataset. In the early 2000's publically funded events occurring in private facilities, predominantly nursing homes, were temporarily included in the public hospital dataset (refer back to Figure 2.4.2 in Paper One on page 30). This had a huge effect on overall bed days, leading to the exclusion of publicly funded events occurring in private facilities with no procedure from the analyses. *Unfiltered* hospital bed days over this period increased from just under 3.5 million bed days per year in 1997 to over 5 million in 2003, and then dropped back down to just over 3 million bed days in 2006. This acts as a temporary window on publically funded long term residential care in New Zealand, and gives some idea of the magnitude of long term residential care utilisation at a population level.

Due to necessity (the difficulty of distinguishing recording errors from long-term care and the huge leverage long term care events have on overall hospital bed days) stays in excess of 365 days were also excluded from the hospital dataset.

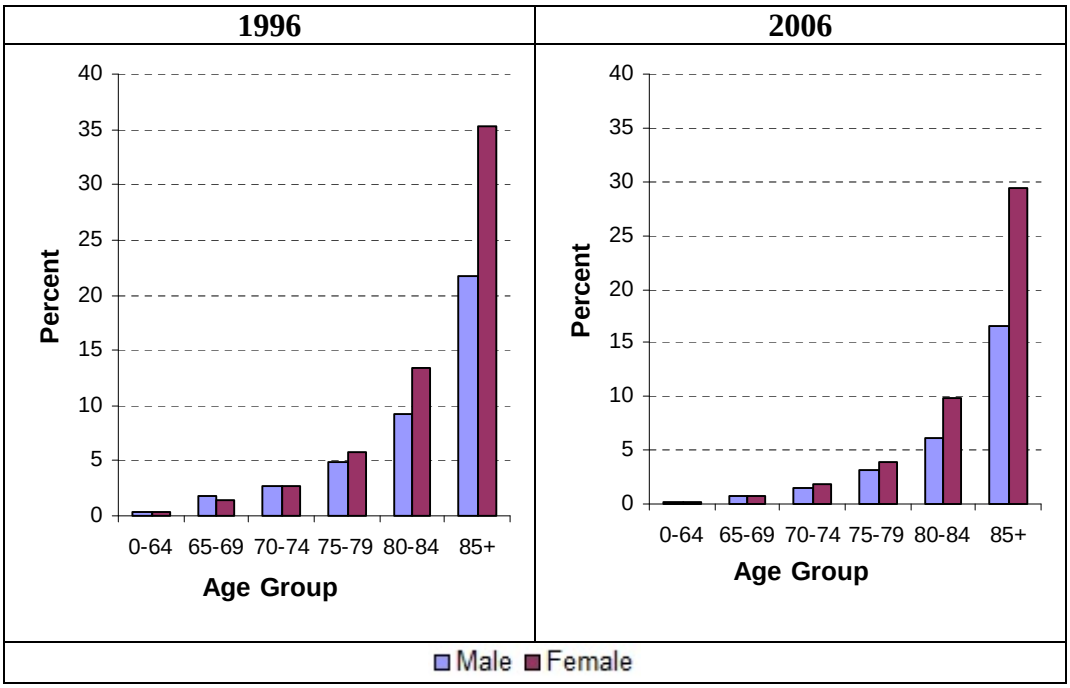
It is not therefore possible to ascertain the prevalence of substitution of care using the data available. The administrative public hospital data only include public hospital events, and even miss some of these: any public hospital stays longer than 365 days are excluded. Other sources need to be consulted for information on the proportion of the elderly population resident in nursing homes over time.

The number of individuals resident in non-private dwellings is published by Statistics New Zealand following each census. The three predominant long-term residential healthcare institutions listed in the census are retirement homes (also referred to as nursing

homes), private hospitals and public hospitals[28]. Of these three types of institution, only public hospitals are included in the datasets used in these papers, and most long-term residents are filtered out (stays exceeding 365 days). The percentage of the population in residential care (retirement homes, private and public hospitals) in 1996 and 2006 by age is displayed in Figure 4.4.1.

Three observations are immediately apparent: the percentage of the population resident in healthcare institutions increases with age; a larger percentage of females are resident in healthcare institutions than males, particularly at older age groups; and the percentage of the population resident in healthcare institutions declined over the time period 1996 to 2006 for every age group and both sexes. The data from the 2001 census supports these findings (Statistics New Zealand 2002). The higher prevalence of residential care for females than for males is well documented in the literature and has been noted in studies in New Zealand (Ministry of Health 2006) and elsewhere (Roos, Montgomery *et al.* 1987). Reasons for this include higher female than male life expectancy, leading to more female than male widows.

Figure 4.4.1: Percent of Usually Resident Population in Residential Healthcare Institutions* by Age and Sex, 1996 and 2006



*Healthcare institutions are retirement homes, public hospitals and private hospitals
Sources of data: (Statistics New Zealand 1998; Statistics New Zealand 2007a)

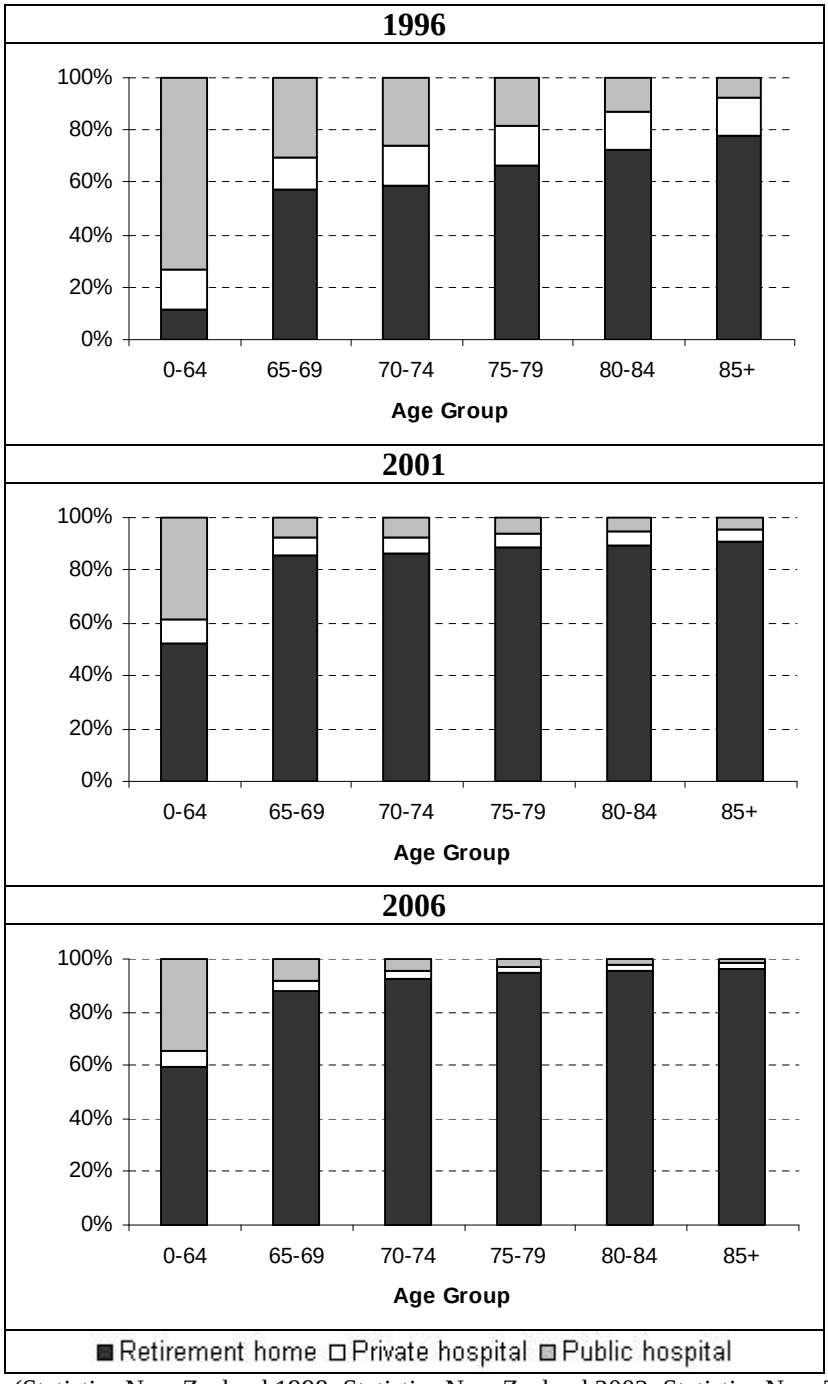
[28] Only long term residents of public hospitals are considered when discussing residential healthcare institutions, not all hospital events.

Figure 4.4.2 examines the percent of people resident in healthcare institutions by type of institution (retirement home, private hospital and public hospital). Most residents of healthcare institutions over the time period were residents of retirement homes (also known as residential care homes and nursing homes). This was the case for all age groups and years, with the exception of ages 0 to 64 years in 1996. In 1996, 73.6 percent of residents of healthcare institutions who were aged 0 to 64 years were residents of public hospitals, compared to just 11.7 percent for retirement homes.

The proportion of healthcare institution residents who are resident in retirement homes increased over the time period. This was seen for all age groups. Older individuals are much more likely to be resident in retirement homes than public or private hospitals: for residents of healthcare institutions aged 85 years and over in 2006, 96.4 percent were residents of retirement homes, compared to 1.8 and 1.9 percent resident in private and public hospitals respectively. Private hospitals have the smallest share for most age groups and at all three points in time. Public hospitals had a higher share, particularly at young age groups. Despite a declining share over time, 35 percent of residents of healthcare institutions aged 0 to 64 were residents of public hospitals in 2006. This share however drops sharply with age.

Several points can be extracted from these seemingly bewildering observations. Older people are more likely to be residents of healthcare institutions. And out of the three types of healthcare institution that have been defined, the datasets used in the analyses in this paper only include hospital events for public hospitals (where the length of stay is less than 365 days). Unfortunately public hospital residents comprise a smaller and smaller proportion of all healthcare institution residents over time, particularly at older ages. However although the hospital events for an increasing proportion of residents in healthcare institutions are excluded from the dataset (public hospital stays greater than 365 days), the proportion of the *population* that are resident in healthcare institutions is declining (see Figure 4.4.1).

Figure 4.4.2: Percent of Healthcare Institution Residents within each Age Group by Institution Type, both Sexes, 1996, 2001 and 2006



Sources of data: (Statistics New Zealand 1998; Statistics New Zealand 2002; Statistics New Zealand 2007a; Statistics New Zealand 2007b)

Kerse and Boyd provide clues as to why the proportion of the older population in healthcare institutions is declining (Figure 4.4.1). They point out that while the New Zealand population aged 65 years and over increased by 43 percent in the last 20 years, the number of residential care beds increased by just 3 percent (Kerse and Boyd 2010). They also found that while the proportion of the population in long term residential care has

decreased, the level of dependency of people in residential care has increased significantly. This suggests excess demand leading to stricter needs based allocation of available residential care beds. A recent Ministry of Health report found evidence to support this suggestion (Ministry of Health 2006). It compared the health of the population aged 65 years and over in residential healthcare and in private dwellings, however one major caveat of their results was the older age structure of the population in residential care, even within age groups. They found a higher prevalence of disability and of sedentary behaviour in the residential care population than in the private dwellings population, and attributed this to strict needs-based assessment for residential care. Thus excess demand and limited supply result in increased levels of dependency for individuals in residential care.

These findings are highly relevant to the results presented in Section 3. In 1996 over a third of females and a fifth of males in the oldest age group (85 years and over) were missed from the hospital data, but by 2006 this had dropped substantially to 29.6 and 16.4 percent for females and males respectively (see Figure 4.4.1). Declines were also seen at younger ages. This is evidence *against* substitution of care from public hospitals to residential care homes and private hospitals. Further evidence against substitution of care is found in the supply and demand components of residential care. Increasingly limited provision of residential care and needs-based assessment for residential care leads to substitution of care from public hospitals to retirement homes for *only* the most dependent individuals. This in turn results in high levels of unmet need for healthcare by the individuals who just missed out on the needs-based assessment for residential care. This should raise public hospital utilisation, particularly at older ages. The results however do not correspond with this, indeed the exact opposite has been observed.

A further test of substitution of care from public hospitals to other sources of healthcare is provided below.

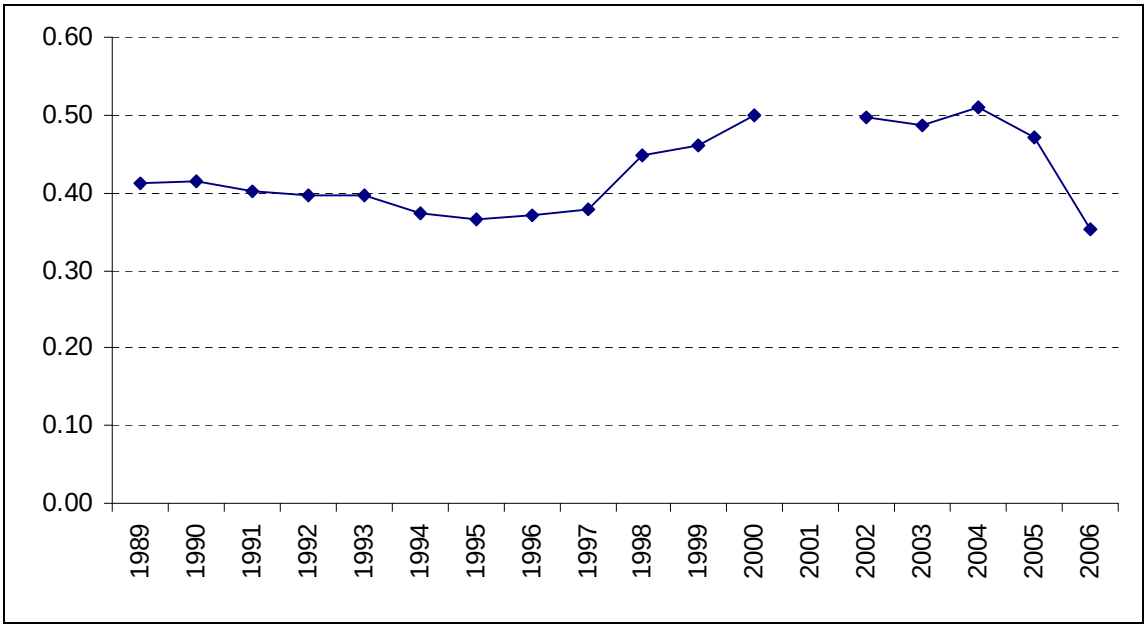
4.5 Changes in Place of Death over time in New Zealand, Implications for Public Hospital Utilisation Prior to Death

Substitution of health care from public hospital to other settings may explain the findings from Section 3: a decreasing number of hospital bed days in the months prior to death. However the analysis of the proportion of the older population in New Zealand in residential care has not found support for substitution of care from public hospitals to

residential care. There may however be substitution of palliative care from public hospitals to the home, through efforts to facilitate death in the home. The hospital and mortality datasets used for the analyses in these papers enable the calculation of the proportion of deaths that occur in a public hospital setting. A decreasing proportion of deaths occurring in public hospitals would suggest end of life care is increasingly occurring outside of a public hospital setting.

While there is no information on place of death provided in the mortality dataset, it is possible to ascertain the number of deaths that occurred in public hospitals by counting the number of discharges with the end type 'dead on discharge'[29]. In order to capture all public hospital deaths, the hospital events are *not* filtered. The number of hospital events ending in death divided by the number of deaths registered in a given year provides the proportion of deaths that occurred in public hospitals. Over the time period 1989 to 2006 the proportion of deaths that occurred in a public hospital decreased slightly from 0.41 to 0.37 between 1989 and 1997, after which it increased substantially, peaking at 0.51 in 2004, and then dropped to 0.35 in 2006 (see Figure 4.4.1).

Figure 4.4.1: Proportion of Deaths occurring in Public Hospitals, All Ages 1989-2006*



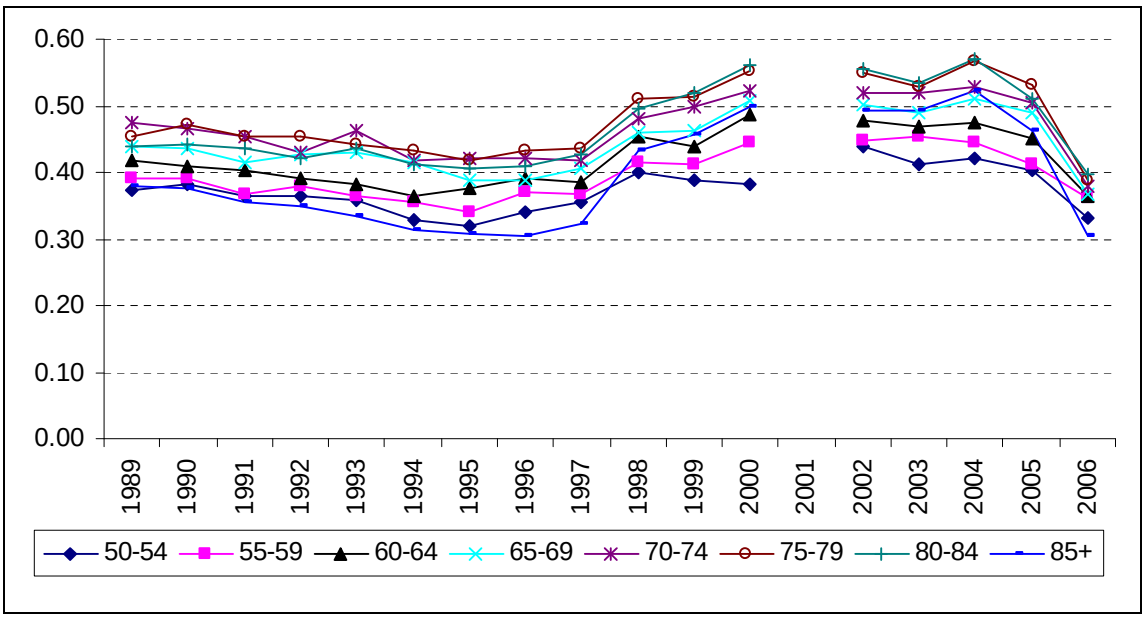
*Mortality data for 2001 missing

When this trend is examined by age (see Figure 4.4.2) a positive association between the proportion of deaths in hospital and increasing age at death is observed, for

[29] After adjusting for duplicate hospital events

ages 50 to 79. A different pattern is seen for the oldest age groups (80 to 84 years and 85 years and over). Decedents aged 74 to 79 were just as likely to die in hospital as decedents aged 80 to 84, while decedents aged 85 years and over were much less likely to die in hospital than other age groups: between 1989 and 1997 and in 2006 decedents aged 85 years and over had the lowest proportion of deaths in hospital out of all age groups 50 and older: a smaller proportion of deaths occur in hospital for the oldest decedents (85 years and over) than for younger decedents (50 to 84 years old).

Figure 4.4.2: Proportion of Deaths occurring in Public Hospitals, Selected Ages, 1989-2006*



*Mortality data for 2001 missing

The findings presented here are consistent with the findings in the literature that substitution of care (particularly in the form of nursing home utilisation) increases with age. This concurs with a recent Australian study which found that the oldest old were less likely to die in hospital than younger decedents, suggesting substitution of care from public hospitals to other settings such as nursing and retirement homes (Brameld, Holman *et al.* 1988).

The reason for the increase in the proportion of deaths accounted for in the public hospital dataset between 1997 and 2005 (especially at older ages) is the temporary inclusion of publicly funded nursing home data over this time period. This can be observed in Figure 2.4.2 on page 30 in Paper One. Recall that the data presented here is *unfiltered* whereas the data used for the analyses in Section 3 is filtered to exclude these cases. Once

this bump is removed, there appears to be a slight downward slope in the proportion of deaths occurring in public hospitals, declining from over 40 percent to about 35 over the time period 1990 to 2006 (see Figure 4.4.1).

Brameld *et al.*, used Australian data and found no evidence of an increasing proportion of deaths at older (or younger) ages occurring outside hospital. Once the spurious bump between 1997 and 2005 (due to temporary inclusion of nursing home data) is removed, Figure 4.4.1 and 4.4.2 show a moderate increase of 6 percent in the proportion of deaths that occur outside of a public hospital setting over the time period.

There is some evidence of substitution of end of life care from public hospital to other settings, which goes a little way towards explaining the decline in public hospital bed days in the last months of life. This is limited evidence of increasing substitution of end of life care from public hospitals to other settings over the time period.

4.6 Relevance of Findings to the Research Questions

The research question for this paper is very similar to the research question for Paper Two. It examines the evidence for compression of morbidity in New Zealand, at a national level, by deprivation, over time and by age at death (something not possible with the life table methods used in Paper Two). The two methods used approach this question from different angles, however analyses by deprivation was not possible for the person-month logistic regression. The research question is given below:

4. Taking hospital utilisation as a proxy for severe morbidity, do the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through analysing time spent in hospital in the last months of life?
    a. Does hospital use in the last months of life differ by deprivation decile?
    b. Does hospital use in the last months of life differ by age at death?
    c. How have these trends changed over time?

The evidence from the results presented in Section 3 support the compression of morbidity hypothesis through a reduction in the severity of morbidity rather than a change in the point at onset of morbidity prior to death. A reduction in hospital utilisation in the months prior to death is observed, but not a change in the point at onset of morbidity. The results from the person-month logistic regression however raise doubts over the continued validity of hospital bed days as a proxy for morbidity.

Hospital use in the last months of life was found to be similar across deprivation deciles, apart from for the oldest decedents (ages 80 years and over). In 1990 decedents aged 80 years and over in more deprived deciles had higher hospital utilisation in the months prior to death than decedents in less deprived deciles, but over the time period this difference diminished.

Hospital utilisation in the months immediately prior to death for decedents in age groups 65 to 69 years and above does not differ much by age, although the oldest decedents (85 years and over) tend to have slightly lower hospital utilisation in the last two or three months of life. This is not the case for months further from death. In months further from death the oldest decedents have higher hospital utilisation than other ages, and this is largely due to higher hospital utilisation at older ages regardless of proximity to death (as seen by high baseline hospital use).

Hospital use in the last few months of life declined for all age groups over the time period and convergence was seen between deprivation deciles. The largest declines were seen for deaths occurring at the oldest ages.

4.7 Limitations of this Research

The predominant limitation of analyses using public hospital data to infer health status is the absence of other forms of health service data (private hospital, primary, long term residential care, palliative care in the home). This limitation has been brought to the fore by the results presented in this paper, however little evidence has been found of increasing substitution of public hospital care to other sources. Other limitations include small numbers of observations that restrict disaggregation of results. Even though population-level data is used, once numbers of deaths are broken down by age group, sex, area deprivation decile and cause of death, cell counts become very small. This resulted in the person-month logistic regression being restricted to age and sex. It could not be conducted by cause, a major limitation given diverse cause-specific trends in hospital utilisation prior to death.

Administrative public hospital data was not collected for the purposes of health research, as has been demonstrated by the extensive filtering required in making the data comparable over time for health research. Administrative health data is also subject to changing trends in diagnosis. However the use of population-level administrative public

hospital data also has considerable strengths. They are free from limitations such as sample bias, and the addition of individual record linkage provides the benefit of facilitating longitudinal analyses.

4.8 Conclusion

The longitudinal analyses presented in this paper find support for the compression of morbidity hypothesis, but through a decline in the severity of morbidity prior to death rather than a shift in the point at onset of morbidity. The decline in hospital utilisation found previously (Papers One and Two) is reiterated in this paper: hospital utilisation is declining over time (particularly at older ages), both for decedents in the months prior to death and for survivors. This finding suggests substitution of health care from public hospitals to other settings, implying that the observed declines in hospital utilisation over the time period are artefactual and do *not* reflect improvements in health.

However it is found that substitution of care from public hospitals to private health care, nursing/retirement homes or palliative care in the home cannot fully explain the declining hospital utilisation prior to death over time. Further explanations could include changing health seeking behaviours, better provision of services leading to more effective preventive care at a primary level, or technological advances and less invasive techniques decreasing recovery times. This would reflect improved health: either health does not deteriorate to the stage where hospital admission is required, or when hospital utilisation is unavoidable, less invasive techniques with faster recovery times are employed. Having examined and largely dismissed substitution of care as an explanation for the declining association between bed days and mortality over time (with the exclusion of limited support for increasing substitution of end of life care), it must be concluded that declines in hospital bed days represent real improvements in morbidity.

CONCLUSION

This thesis brings together three papers, addressing four research questions in the broad area of health and mortality in New Zealand between 1974 and 2006. The research questions of this study are presented below:

1. How do the **patterns of mortality** in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of mortality differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

2. How do the **patterns of hospital use (**bed day rates) in New Zealand vary by area deprivation?
    a. Does the cause-specific distribution of hospital use differ by deprivation decile?
    b. Does the diversity of cause of death differ by deprivation decile?
    c. How have these trends changed over time?

3. Taking hospital utilisation as a proxy for severe morbidity, does the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through the use of Hospital Utilisation Expectancies (HUEs)?
    a. Does the lifetime expectation of hospital use at birth and at different ages differ by deprivation decile?
    b. How have these trends changed over time?

4. Taking hospital utilisation as a proxy for severe morbidity, does the mortality and hospital data support or contradict the **compression of morbidity hypothesis** through analysing time spent in hospital in the last months of life?
    a. Does hospital use in the last months of life differ by age at death?
    b. Does hospital use in the last months of life differ by deprivation decile?
    c. How have these trends changed over time?

In interpreting findings from the thesis as a whole, it is useful to group the research questions into two areas. While all of the research questions are concerned with health and mortality trends in New Zealand, they can be divided into general health patterns, and evidence for compression of morbidity. Research questions 1 and 2 are concerned with trends in mortality and hospital utilisation in New Zealand (by age, sex, deprivation decile and cause) whereas research questions 3 and 4 examine evidence for compression or expansion of morbidity in New Zealand (by age at death, sex, and deprivation decile). The

first area is examined in Paper One, and the second area in Papers Two and Three (using life table and longitudinal methodologies respectively).

This study contributes substantial new knowledge to both areas. In this conclusion the methodology behind the results is examined. This is followed by an exploration of the key findings from the two areas outlined above, highlighting contributions to the existing knowledge base. Limitations and caveats to interpretation of the results are outlined, followed by suggested directions for further research.

Reflection on the Data and Methods Used

Most research in the field of population health, and particularly in the field of compression of morbidity, has used subjective self reported health measures from surveys. This study takes a novel approach, both in terms of data (public hospital and mortality records) and methods (Hospital Utilisation Expectancies (HUEs) and longitudinal techniques). A detailed description of the data and methods is presented elsewhere (see the data and methods section in each paper and Appendixes A and B). The main advantages of the data and methods are outlined below, with emphasis on the aspects which make this research unique.

This research uses routinely collected administrative health (hospital and mortality) records. While these were not collected for the purposes of public health research, they are a valuable and underutilised data source. Administrative health records facilitate research at a population level, free of sample bias. While the mechanisms behind hospital utilisation may be influenced by many non-health related factors (health service policy, healthcare seeking behaviour, quality of primary healthcare services, use of private healthcare services), they are nonetheless an objective measure, remaining relatively unaffected by perceptions of health status that can vary from individual to individual, by age, and by ethnic group.

The first paper in this thesis contrasts age-specific and age-standardised hospital bed day and mortality rates, and examines differentials by area socioeconomic deprivation. This is an unusual analysis on two counts: the use of hospital bed days rather than discharges, and disaggregation by area socioeconomic deprivation. Most research of hospital utilisation in New Zealand (Jackson, Kelsall *et al.* 1998; Pool, Baxendine *et al.* 2006; Raymont 2008) and elsewhere (Billings, Zeitel *et al.* 1993; Hofer, Wolfe *et al.*

1998) has used discharge rates. Some New Zealand studies did use hospital bed day rates, but did not disaggregate results by socioeconomic deprivation (Pool, Baxendine *et al.* 2006; Raymont 2008). In New Zealand in recent years there has been a trend towards more, shorter hospital stays, particularly day patients. This results in more discharges, but does not represent an increase in morbidity. In addition, hospital discharge rates are not weighted by severity of morbidity. Hospital bed day rates avoid both of these disadvantages, providing a measure of morbidity that is weighted by severity (number of days spent in hospital) and is less prone to fluctuations brought about by changes in hospital policy.

The second and third papers in this thesis focus on evidence for compression of morbidity in New Zealand, using different methods: in the second paper period prevalence life table methods are employed, and in the third paper record linkage and longitudinal techniques. Period prevalence life tables are more traditionally used to calculate health expectancies, using self-reported health survey data. Paper Two uses HUEs, a variation of health expectancies that use hospital bed day rates instead of survey data and which represent the average number of days while still alive that an individual can expect to spend in hospital. Unlike health expectancies, HUEs are free from limitations of subjectivity and sample bias. HUEs were developed by Jit Cheung and colleagues at the University of Waikato, and have been used in research examining the compression of morbidity hypothesis in New Zealand (Cheung, Katzenellenbogen *et al.* 2001; Pool, Baxendine *et al.* 2009). This study builds on the existing research using HUEs in New Zealand and adds a new element by disaggregating results by area socioeconomic deprivation.

The third paper makes full advantage of the rich, individual and population level data sources by conducting individual record linkage and implementing longitudinal techniques. Due to the use of survey data in most research examining compression of morbidity, longitudinal methods are usually not possible. Compression of morbidity theories are concerned with health in the period prior to death, at a population level. Cross-sectional techniques that do not link health events with mortality at an individual level cannot fully explore health in the period prior to death: they can only examine age-specific averages of health and mortality events. Through individual record linkage this study is able to use longitudinal methods. In comparison to cross-sectional methods, longitudinal methods enable the researcher to identify the *actual* health experience of individuals in the period prior to death. These are then aggregated and presented at a population level. Only a

few studies have examined hospital utilisation in the months leading up to death and many of these studies are from a health economics perspective, concerned with the *cost* of dying (Brameld, Holman *et al.* 1988; Felder, Meier *et al.* 2000; Salas and Raftery 2001; Stearns and Norton 2004; Lynch, Holman *et al.* 2007). This study is unique on several counts: population level data provides enough cases to examine hospital utilisation in months prior to death, rather than quarters, and enough cases to disaggregate by area level socioeconomic deprivation. It is also the first study of its kind to use New Zealand data.

The fields of health inequalities and compression of morbidity have been well trodden in research. This study succeeds in bringing a unique approach, both in terms of the data and the methods used. The key findings from this study with reference to existing knowledge are discussed below.

Key Finding 1 - Declining Hospital Bed Day and Mortality Rates over time, and Convergence by Deprivation

Paper One examined patterns in hospital utilisation and mortality, disaggregating by age, sex, and area deprivation. There has been considerable research conducted in this field. Previous studies found deprivation gradients in both hospital discharge and mortality rates in New Zealand, with people in more deprived areas having higher hospital discharge and mortality rates (Jackson, Kelsall *et al.* 1998; Salmond and Crampton 1999). However these studies used hospital discharge rates rather than bed day rates, an important distinction. Pool *et al.* highlighted the differences between hospital bed day rates and hospital discharge rates, they found that while males tend to have higher hospital discharge rates than females, they have lower hospital bed day rates (Pool, Baxendine *et al.* 2009). In a rare study using hospital bed day rates Pool *et al.* found that hospital bed day and mortality rates declined substantially over time, for both sexes and most ages, although regional variation was found in these trends (Pool, Baxendine *et al.* 2009).

This research builds on the comprehensive study of hospital utilisation and mortality trends in New Zealand by Pool *et al.*. However unlike the study by Pool *et al.*, trends by area deprivation are examined. A clear area deprivation gradient in hospital bed day rates is determined, which had not been previously shown. Individuals living in more deprived areas are not only more likely to be admitted to hospital than individuals in less deprived areas, they are also more likely to spend more days in hospital.

Another result is that the extent of the deprivation gradient in mortality rates varies by age. A clear gradient is observed at middle to older ages, but little difference in mortality rates by area deprivation is observed for ages 80 years and over. Diversity of cause of hospital utilisation *and* cause of death is found to increase over time, as mortality rates due to the main killers (cardiovascular, neoplasms, respiratory and external causes) decrease.

Key Finding 2 - Compression of Morbidity is Occurring, Especially at Older Ages

Papers Two and Three used period prevalence life table and longitudinal techniques respectively to examine evidence for compression of morbidity. Studies examining compression and expansion of morbidity theories in New Zealand have found mixed results. Some offer support for compression of morbidity (Cheung, Katzenellenbogen *et al.* 2001; Pool, Baxendine *et al.* 2009), whereas others provide evidence that tentatively (Davis, Graham *et al.* 1999; Davis, Mathers *et al.* 2003) or strongly (Ministry of Health and Statistics New Zealand 2009) supports expansion of morbidity. However as with studies elsewhere, different measures of health and different methods yield different results. It is often unclear if conflicting results are real, or an artefact of the data and methods used.

This study finds strong evidence in favour of compression of morbidity, despite limited evidence for rectangularisation of the survival curve, or of a shift in the onset of morbidity. It is argued that rectangularisation of the survival curve is *not* a necessary prerequisite of compression of morbidity, as long as increases in healthy life expectancy continue to exceed increases in overall life expectancy. HUEs are found to decline and life expectancy increase over the period, indicating that levels of morbidity (hospital bed days) are decreasing at a faster rate than life expectancy is increasing. More people living to older ages (80 years and over) does *not* imply that morbidity at a population level will increase: HUEs are declining most rapidly at the oldest ages. While this study has examined evidence for compression of morbidity in a New Zealand setting, it agrees with international evidence that supports compression of morbidity despite no apparent limit to life expectancy (Vaupel 2010).

Evidence of considerable inequality in HUEs and life expectancy by area deprivation was observed at the start of the time period (1991), however there was also evidence of convergence over time, with more deprived areas improving to the level of less deprived areas by the end of the time period (2006). This finding was reiterated in the longitudinal analyses in Paper Three: at the start of the time period (1993) elderly decedents (80+) in more deprived areas spent more days in hospital in the last few months of life than elderly decedents in less deprived areas. But by the end of the time period (2006), hospital use in the months prior to death by elderly decedents in more deprived areas had reduced to levels comparable with less deprived areas. The longitudinal analyses also found evidence in support of compression of morbidity, with *severity* of morbidity (as measured by the number of days spent in hospital) in the months prior to death declining over time. There was no evidence of a shift in the point at onset of morbidity, however. But it is argued that compression of morbidity can also occur through decreasing severity of morbidity prior to death.

The late 1990s and early 2000s were a time of upheaval and reform in the New Zealand health system. Decreasing hospital utilisation may represent efficiency gains and improvements in morbidity, or increased substitution of healthcare from public hospitals to other settings, in recent years. Potential explanations for this (private health insurance, proportion of population in long term residential care facilities, facilitation of death and palliative care in the home) were examined and discounted, with the exception of death and palliative care in the home. A slight decrease in the proportion of deaths occurring in hospital was seen over the time period, suggesting end of life care in other settings, particularly at older ages. However this cannot account for the majority of the decrease in hospital utilisation over the time period, and it is therefore assumed that decreases in hospital utilisation, for the main part, reflect real improvements in morbidity. Improvements in morbidity may be at the preventative stage (better health-related behaviours, improved primary care, earlier diagnosis) or through improved efficiencies in public hospitals (technological advances and less invasive procedures).

The data and methods used in this study have many limitations, which must be considered when interpreting the results. These limitations and areas which would benefit from further research are now discussed.

Limitations and Further Research

The use of administrative public hospital data in health research has many limitations. The data was not collected for health research purposes, and thus extensive manipulation is required to make the data consistent over time and to adapt the data to represent a relevant proxy for morbidity. Other forms of health care (private hospitals, primary healthcare services, long term residential care) are absent from the data; thus if substitution of care from public hospitals to other settings occurs then the data may diminish in relevance as a proxy for morbidity. Population level public hospital records however also have considerable strengths compared to other measures of population health, namely objectivity and the absence of sample bias.

Routinely collected population level administrative health records are valuable and underutilised data in health research. While some adjustment is required to adapt them for the purpose of health research, they provide cheap, large scale, detailed information on health and death. Administrative health data are however at the whim of policy makers, they are prone to inconsistencies over time and to changes in their validity as a proxy for morbidity. This is an issue even if using hospital bed days, not discharges, and even after extensive filtering to ensure consistency over time. This research provides strong support for compression of morbidity in New Zealand over the time period studied, but also stresses the dangers of over-reliance on administrative health records alone. However, the use of different types of data and methods used in different studies can make it unclear if conflicting results represent real differences or an artefact of the type of data and methods.

There are two directions for further research. Research is required in New Zealand using different measures of health status (such as self-reported health from surveys) to test the continuing validity of hospital bed days as a proxy for morbidity over the time period. This research will examine to what extent the strong support for compression of morbidity in New Zealand that has been found by this study is in reality an artefact of changes in policy resulting in substitution of health care from public hospitals to other settings. The second direction for further research is to extend the use of HUEs to examine the compression of morbidity hypothesis in other countries. HUEs have only been used in a New Zealand setting however they make use of administrative health records that are routinely collected in many countries. Using the same type of data and method to examine evidence for compression of morbidity in other countries would enable cross-country comparisons, reducing the risk of an artefactual difference due to the type of data and methods used.

CONCLUSION

REFERENCES

Acheson, E. D. (1967). Medical Record Linkage. London, Oxford University Press.

Ajwani, S., Blakely, T., Robson, B., Atkinson, J. and Kiro, C. (2003). "Unlocking the numerator-denominator bias 3: adjustment ratios by ethnicity for 1981-1999 mortality data. The New Zealand Census-Mortality Study." New Zealand Medical Journal **116** (1175).

Banks, J., Marmot, M., Oldfield, Z. and Smith, J. P. (2006). "Disease and Disadvantage in the United States and in England." Journal of the American Medical Association **295** (17): 2037-2045.

Barnett, R., Moon, G. and Kearns, R. (2004). "Social Inequality and Ethnic Differences in Smoking in New Zealand." Social Science and Medicine **59**: 129-143.

BBC (2008). Social factors key to ill health. 28th August 2008. BBC News.

Billings, J., Zeitel, L., Lukomnik, J., Carey, T. S., Blank, A. E. and Newman, L. (1993). "Impact of Socioeconomic Status on Hospital Use in New York City." Health Affairs **12** (1): 162-73.

Blakely, T., Tobias, M., Robson, B., Ajwani, S., Bonne, M. and Woodward, A. (2005). "Widening ethnic mortality disparities in New Zealand 1981-99." Social Science and Medicine **61**: 2233-2251.

Blakely, T., Woodward, A., Pearce, N., Salmond, C., Kiro, C. and Davis, P. (2002). "Socio-economic factors and mortality among 25-64 year olds followed from 1991 to 1994: the New Zealand Census-Mortality Study." New Zealand Medical Journal **115** (1149): 93-97.

Blakely, T. A., Ajwani, S., Robson, B., Tobias, T. and Bonne, M. (2004). "Decades of Disparity: Widening Ethnic Mortality Gaps from 1980 to 1999." New Zealand Medical Journal **117** (1199).

Brameld, K. J., Holman, C. D. J., Bass, A. J., Codde, J. P. and Rouse, I. L. (1988). "Hospitalisation of the elderly during the last year of life: an application of record linkage in Western Australia." Journal of Epidemiology and Community Health **52**: 740-744.

Brazier, J. E., Harper, R., Jones, N. M. B., O'Cathain, A., Thomas, K. J., Usherwood, T. and Westlake, L. (1992). "Validating the SF-36 health survey questionnaire: new outcome measure for primary care." British Medical Journal **305**: 160-164.

REFERENCES

Case, A. and Paxson, C. (2005). "Sex differences in morbidity and mortality." Demography **42** (2): 189-214.

Cheung, J. (1999). Mortality, morbidity and population health dynamics, Doctor of Philosophy thesis, University of Waikato, Hamilton, New Zealand.

Cheung, J., Katzenellenbogen, J., Baxendine, S., Pool, I. and Jackson, G. (2001). "Hospital utilisation expectancies in New Zealand, 1980-98." Australian Health Review **24** (4): 46-56.

Collins English Dictionary 5th Edition (2000). Harper Collins Publishers.

Crampton, P., Salmond, C., Woodward, A. and Reid, P. (2000). "Socioeconomic Deprivation and Ethnicity Are Both Important for Anti-tobacco Health Promotion." Health Education and Behavior **27** (3): 317-327.

Davis, P., Graham, P. and Pearce, N. (1999). "Health expectancy in New Zealand, 1981-1991: social variations and trends in a period of rapid social and economic change." Journal of Epidemiology and Community Health **53**: 519-527.

Davis, P., Mathers, C. and Graham, P. (2003). Health Expectancy in Australia and New Zealand. Determining Health Expectancies. J. M. Robine, C. Jagger, C. D. Mathers, E. M. Crimmins and R. M. Suzman. Chichester, Wiley.

Dharmalingam, A., Pool, I., Baxendine, S. and Sceats, J. (2004). "Trends and patterns of avoidable hospitalisations in New Zealand: 1980-1997." New Zealand Medical Journal **117** (1198): 1-20.

Doblhammer, G. and Kytir, J. (2001). "Compression or Expansion of Morbidity? Trends in Healthy-Life Expectancy in the Elderly Australian Population between 1978 and 1998." Social Science and Medicine **52**: 385-391.

Dunn, H. L. (1946). "Record Linkage." American Journal of Public Health **36**: 1412-1416.

Eachus, J., Chan, P., Pearson, N., Propper, C. and Davey Smith, G. (1999). "An additional dimension to health inequalities: disease severity and socioeconomic position." Journal of Epidemiology and Community Health **53**: 603-611.

Felder, S., Meier, M. and Schmitt, H. (2000). "Health care expenditure in the last months of life." Journal of Health Economics **19**: 679-695.

Finau, S. A. and Tukitonga, C. (1999). Pacific Peoples in New Zealand. Health and Society in Aotearoa New Zealand. P. Davis and K. Dew. Auckland, Oxford University Press**:** 129-143.

Fries, J. F. (1980). "Aging, Natural Death, and the Compression of Morbidity." The New England Journal of Medicine **303** (3): 130-135.

Fries, J. F. (1983). "The Compression of Morbidity." The Milbank Quarterly **61** (3): 397-419.

Fries, J. F. (1984). "The Compression of Morbidity: Miscellaneous Comments about a Theme." The Gerontologist **24** (4): 354-359.

Fries, J. F. (1989). "The Compression of Morbidity: Near or Far?" The Milbank Quarterly **67** (2): 208-232.

Fries, J. F. (1990). "The Sunny Side of Aging." Journal of the American Medical Association **263** (17): 2354-2355.

Fries, J. F. (1992). "Strategies for Reduction of Morbidity." American Journal of Clinical Nutrition **55**: 1257S-1262S.

Fries, J. F. (2000). "Compression of Morbidity in the Elderly." Vaccine **18**: 1584-1589.

Fries, J. F. and Crapo, L. (1981). Vitality and Aging: Implications of the Rectangular Curve. New York, W.H. Freeman.

Gordis, L., Cold, E. and Seltser, R. (1977). "Privacy Protection in Epidemiologic and Medical Research: A Challenge and a Responsibility." American Journal of Epidemiology **105** (3): 163-168.

Graham, P., Blakely, T., Davis, P., Sporle, A. and Pearce, N. (2004). "Compression, Expansion, or Dynamic Equilibrium? The Evolution of Health Expectancy in New Zealand." Journal of Epidemiology and Community Health **58**: 659-666.

Gruenberg, E. M. (1977). "The Failures of Success." Milbank Memorial Fund Quarterly/Health and Society **55** (1): 3-24.

Hayflick, L. (1994). Why do we Live as Long as we Do? Aging, Concepts and Controversies. H. R. Moody, Pine Forge Press. **Fifth Edition:** 41-43.

Hayflick, L. (2000). "The Future of Ageing." Nature **408**: 267-269.

Health Funds Association of New Zealand (2001). An Insight into the New Zealand Health Insurance Industry. Wellington, Health Funds Association of New Zealand Inc.

Health Funds Association of New Zealand (2006). Health Insurance Statistics March 2006.

Health Utilisation Research Alliance (2006). "Ethnicity, socioeconomic deprivation and consultation rates in New Zealand general practice." Journal of Health Services Research and Policy **11** (3): 141-149.

Hefford, M., Crampton, P. and Foley, J. (2005). "Reducing health disparities through primary care reform: the New Zealand experiment." Health Policy **72**: 9-23.

Hessler, R. M., Eriksson, B. G., Dey, D., Steen, G., Sundh, V. and Steen, B. (2003). "The Compression of Morbidity Debate in Aging: an Empirical Test using the

Gerontological and Geriatric Population Studies in Göteborg, Sweden." Archives of Gerontology and Geriatrics **37**: 213-222.

Hills, J., Brewer, M., Jenkins, S., Lister, R., Lupton, R., Machin, S., Modood, T., Rees, T. and Riddell, S. (2010). An Anatomy of Economic Inequality in the UK: Report of the National Equality Panel.

Hinde, A. (1998). Demographic Methods, Oxford University Press.

Hofer, T. P., Wolfe, R. A., Tedeschi, P. J., McMahon, L. F. and Griffith, J. R. (1998). "Use of Community Versus Individual Socioeconomic Data in Predicting Variation in Hospital Use." Health Services Research **33** (2): 243-259.

Holman, C. D. J., Bass, A. J., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., Brook, E. L., Trutwein, B., Rouse, I. L., Watson, C. R., De Klerk, N. H. and Stanley, F. J. (2008). "A decade of data linkage in Western Australia: strategic design, applications and benefits of the WA data linkage system." Australian Health Review **32** (4): 766-777.

Hopkins, S. and Cumming, J. (2001). "The impact of changes in private health expenditure on New Zealand households." Health Policy **58**: 215-229.

House, J. S., Kessler, R. C., Herzog, A. R., Mero, R. P., Kinney, A. M. and Breslow, M. J. (1990). "Age, Socioeconomic Status, and Health." The Milbank Quarterly **68** (3): 381-411.

Jackson, G., Kelsall, L., Parr, A. and Papa, D. (1998). Socio-Economic Inequalities in Health Care, A Preliminary Analysis of the Link Between Health Status and Socio-Economic Status in the North Health Region, Health Funding Authority.

Jagger, C. (2000). "Compression or Expansion of Morbidity - What does the Future Hold?" Age and Ageing **29**: 93-94.

Johansson, L. A. and Westerling, R. (2000). "Comparing Swedish hospital discharge records with death certificates: implications for mortality statistics." International Journal of Epidemiology **29**: 495-502.

Johansson, S. R. (1991). "The Health Transition: The Cultural Inflation of Morbidity during the Decline of Mortality." Health Transition Review **1** (1): 39-68.

Johnstone, K., Cheung, J., Pool, I., Dharmalingam, A. and Hillcoat-Nalletamby, S. (1998). Health Measures: Principles and Applications to New Zealand Data. A Report Prepared for the Ministry of Health. Hamilton, Population Studies Centre, University of Waikato.

Judge, K. and Paterson, I. (2001). Poverty, Income Inequality and Health. Treasury Working Paper 01/29.

Katzenellenbogen, J., Baxendine, S., Cheung, J., Pool, I. and Jackson, G. (2001). "Getting New Zealand hospital discharge data in shape for a time series documenting National Minimum Data Set milestones." Health Manager **8** (3): 14-18.

Kelman, C. W., Bass, A. J. and Holman, C. D. J. (2002). "Research use of linked health data - a best practice protocol." Australian and New Zealand Journal of Public Health **26** (3): 251-255.

Kerse, N. and Boyd, M. (2010). "Improving care for older people in residential care." New Zealand Medical Journal **123** (1308): 13-15.

Khaw, K. (1997). "Healthy Aging." British Medical Journal **315** (1090-1096).

Leibson, C. L., Ballard, D. J., Whisnant, J. P. and Melton, L. J. (1992). "The Compression of Morbidity Hypothesis: Promise and Pitfalls of Using Record-linked Data Bases to Assess Secular Trends in Morbidity and Mortality." The Milbank Quarterly **70** (1): 127-154.

Lynch, C., Holman, C. D. J. and Moorin, R. E. (2007). "Use of Western Australian linked hospital morbidity and mortality data to explore theories of compression, expansion and dynamic equilibrium." Australian Health Review **31** (4): 571-581.

Magnusson, R. S. (2002). "Data Linkage, Health Research and Privacy: Regulating Data Flows in Australia's Health Information System." Sydney Law Review **24** (5): 5-55.

Manton, K. G. (1982). "Changing Concepts of Morbidity and Mortality in the Elderly Population." Milbank Memorial Fund Quarterly/Health and Society **60** (2): 183-244.

Manton, K. G., Stallard, E. and Tolley, H. D. (1991). "Limits to Human Life Expectancy: Evidence, Prospects, and Implications." Population and Development Review **17** (4): 603-637.

Maori Health Website. (2010). Retrieved 28th January 2010, from www.maorihealth.govt.nz.

Ministry of Health (2000). Social Inequalities in Health, New Zealand 1999: A summary. Wellington, Ministry of Health.

Ministry of Health (2002). Health Expenditure Trends in New Zealand 1980-2000. Wellington, Ministry of Health.

Ministry of Health (2005). Health Expenditure Trends in New Zealand 1992-2003. Wellington, Ministry of Health.

REFERENCES

Ministry of Health (2006). Older People's Health Chart Book. Wellington.

Ministry of Health (2008). Health Expenditure Trends in New Zealand 1996-2006. Wellington, Ministry of Health.

Ministry of Health and Statistics New Zealand (2009). Longer Life, Better Health? Trends in health expectancy in New Zealand, 1996-2006. Wellington, Statistics New Zealand.

Moorin, R. E. and Holman, C. D. J. (2008). "The cost of in-patient care in Western Australia in the last years of life: A population-based data linkage study." Health Policy **85**: 380-390.

Mor, V. (2005). "The Compression of Morbidity Hypothesis: A Review of Research and Prospects for the Future." Journal of the American Geriatrics Society **53** (9): S308-S309.

New Zealand Health Information Service (NZHIS) Website. (2008). Retrieved 2nd February 2010, from http://www.nzhis.govt.nz/moh.nsf/pagesns/435/$File/2008+domicile_code_tab1.3.xls.

Newcombe, H. B. and Kennedy, J. M. (1962). "Record Linkage: Making Maximum Use of the Discriminating Power of Identifying Information." Communications of the Association for Computing Machinery **5** (11): 563-566.

Olshansky, S. J. and Ault, A. B. (1986). "The Fourth Stage of the Epidemiologic Transition: The Age of Delayed Degenerative Diseases." The Milbank Quarterly **64** (3): 355-391.

Olshansky, S. J., Rudberg, M. A., Carnes, B. A., Cassel, C. K. and Brody, J. A. (1991). "Trading off Longer Life for Worsening Health. The Expansion of Morbidity Hypothesis." Journal of Aging and Health **3** (2): 194-216

Omran, A. R. (1971). "The Epidemiologic Transition: A Theory of the Epidemiology of Population Change." Milbank Memorial Fund Quarterly **49** (4): 509-538.

Omran, A. R. (1983). "The Epidemiologic Transition Theory. A Preliminary Update." Journal of Tropical Pediatrics **29**: 305-316.

Omran, A. R. (1998). "The Epidemiologic Transition Theory Revisited Thirty Years Later." World Health Statistics Quarterly **51**: 99-119.

Pampel, F. C. (2002). "Cigarette Use and Sex Differentials in Mortality." Population and Development Review **28** (1): 77-104.

Paul, C. (2002). "Socio-economic factors and death rates." New Zealand Medical Journal **115** (1149): 91-92.

Pearce, J. and Dorling, D. (2006). "Increasing Geographical Inequalities in Health in New Zealand, 1980-2001." International Journal of Epidemiology **35**: 597-603.

Peto, R., Lopez, A. D., Boreham, J., Thun, M. and Heath, C. (1992). "Mortality from tobacco in developed countries: Indirect estimation from national vital statistics." Lancet **339**: 1268-1278.

Pool, I. (1994). Cross-Comparative Perspectives on New Zealand's Health. Social Dimensions of Health and Disease, New Zealand Perspectives. J. Spicer, A. Trlin and J. A. Walton. Palmerston North, The Dunmore Printing Company Ltd.

Pool, I., Baxendine, S., Cheung, J., Coombs, N., Dharmalingam, A., Jackson, G., Katzenellenbogen, J., Sceats, J. and Cooper, J. (2009). Restructuring and Hospital Care: Sub-national Trends, Differentials, and their Impacts in New Zealand from 1981, The Population Studies Centre, University of Waikato.

Pool, I., Baxendine, S., Katzenellenbogen, J. and Howard, S. (2006). "New Zealand Regions, 1986-2001: Hospitalisation and some Related Health Facts." Population Studies Centre Discussion Papers **63**.

Raymont, A. (2008). "Hospital discharges in New Zealand 1991-2005: changes over time and variation between districts." New Zealand Medical Journal **121** (1279): 66-74.

Riddell, T. (2005). "Heart failure hospitalisations and deaths in New Zealand: patterns by deprivation and ethnicity." New Zealand Medical Journal **118** (1208).

Robine, J., Romieu, I. and Cambois, E. (1999). "Health expectancy indicators." Bulletin of the World Health Organisation **77** (2): 181-185.

Robling, M. R., Hood, K., Houston, H., Pill, R., Fay, J. and Evans, H. M. (2004). "Public attitudes towards the use of primary care patient record data in medical research without consent: a qualitative study." Journal of Medical Ethics **30**: 104-109.

Rodgers, S. E., Lyons, R. A., Desilva, R., Jones, K. H., Brooks, C. J., Ford, D. V., John, G. and Verplancke, J. (2009). "Residential Anonymous Linking Fields (RALFs): a novel information infrastructure to study the interaction between the environment and individual's health." Journal of Public Health **31** (4): 582-588.

Rogers, A., Rogers, R. G. and Belanger, A. (1990). "Longer Life but Worse Health? Measurement and Dynamics." The Gerontologist **30** (5): 640-649.

Rogers, R. G., Everett, B. G., Saint Onge, J. M. and Krueger, P. M. (2008). Sex Differentials in Mortality: The Importance of Demographic, Socioeconomic, and

Behavioural Factors, Presented to the Division of Social Statistics, University of Southampton, June 10, 2008.

Rogers, R. G. and Hackenberg, R. (1987). "Extending Epidemiologic Transition Theory: A New Stage." Social Biology **34** (3-4): 234-243.

Roos, N. P., Montgomery, P. and Roos, L. L. (1987). "Health Care Utilization in the Years Prior to Death." The Milbank Quarterly **65** (2): 231-254.

Salas, C. and Raftery, J. P. (2001). "Economic Issues in Testing the Age Neutrality of Health Care Expenditure." Health Economics **10**: 669-671.

Salmond, C. and Crampton, P. (1999). Deprivation and Health. Social Inequalities in Health: New Zealand 1999. Wellington, Ministry of Health.

Salmond, C. and Crampton, P. (2001). "NZDep96 - What Does it Measure?" Social Policy Journal of New Zealand (17): 82-100.

Salmond, C. and Crampton, P. (2002). NZDep2001 index of Deprivation User's Manual, Department of Public Health, Wellington School of Medicine and Health Sciences.

Schneider, E. L. and Brody, J. A. (1983). "Aging, Natural Death and the Compression of Morbidity: Another View." The New England Journal of Medicine **309** (14): 854-855.

Scott, K. M., Sarfati, D., Tobias, M. I. and Haslett, S. J. (2000). "A challenge to the cross-cultural validity of the SF-36 health survey: factor structure in Māori, pacific and New Zealand European ethnic groups." Social Science and Medicine **51**: 1655-1664.

Seshamani, M. and Gray, A. (2004). "Ageing and health-care expenditure: the red herring argument revisited." Health Economics **13**: 303-314.

Shryock, H. S., Siegel, J. S. and Associates (1976). The Methods and Materials of Demography, Academic Press.

Sibthorpe, B., Kliewer, E. and Smith, L. (1995). "Record linkage in Australian epidemiological research: health benefits, privacy safeguards and future potential." Australian Journal of Public Health **19** (3): 250-256.

Siegel, J. S. and Swanson, D. A. (2004). The Methods and Materials of Demography (Second Edition), Academic Press.

Spillman, B. C. and Lubitz, J. (2000). "The Effect of Longevity on Spending for Acute and Long-Term Care." The New England Journal of Medicine **342** (19): 1409-1415.

Statistics New Zealand (1979). New Zealand Period Life Tables, 1975-1977. Wellington.

Statistics New Zealand (1998). Families and Households: 1996 Census of Population and Dwellings. Wellington.

Statistics New Zealand (2002). 2001 Census of Population and Dwellings, National Summary. Wellington.

Statistics New Zealand (2007a). 2006 Census of Population and Dwellings: Quickstats about New Zealand's Population and Dwellings. Wellington.

Statistics New Zealand (2007b). New Zealand's 65+ Population: A Statistical Volume. Wellington.

Statistics New Zealand (2008). New Zealand Period Life Tables, 2005-2007. Wellington.

Stearns, S. C. and Norton, E. C. (2004). "Time to include time to death? The future of health care expenditure predictions." Health Economics **13**: 315-327.

Sullivan, D. F. (1971). "A Single Index of Mortality and Morbidity." Health Services and Mental Health Administration (HSMHA) Health Reports **86** (4): 347-354.

Tobias, M. and Cheung, J. (2003). "Monitoring health inequalities: life expetancy and small area deprivation in New Zealand." Population Health Metrics **1** (2).

Tobias, T., Bhattacharya, A. and White, P. (2008). "Cross classification of the New Zealand population by ethnicity and deprivation: trends from 1996 to 2006." Australian and New Zealand Journal of Public Health **32** (5): 431-436.

Trutwein, B., Holman, C. D. J. and Rosman, D. L. (2006). "Health Data Linkage Conserves privacy in a Research-Rich Environment." Annals of Epidemiology **16** (4): 279-280.

United Nations Population Division Website. (2009). Retrieved 2nd February 2010, from http://www.un.org/esa/population/.

University of Otago Website. (1988-2001). Retrieved 2nd February 2010, from http://www.wnmeds.ac.nz/academic/dph/research/socialindicators.html.

Vaupel, J. W. (2010). "Biodemography of human ageing." Nature **464**: 536-542.

Wald, N., Law, M., Meade, T., Miller, G., Alberman, E. and Dickinson, J. (1994). "Use of Personal Medical Records for Research Purposes." British Medical Journal **309** (6966): 1422-1424.

Wamala, S., Blakely, T. and Atkinson, J. (2006). "Trends in absolute socioeconomic inequalities in mortality in Sweden and New Zealand. A 20-year gender perspective." BMC Public Health **6** (164).

REFERENCES

Washington State Department of Health Website. (2009). "Assessment Guidelines: Confidence Intervals." Retrieved 2nd February 2010, from http://www.doh.wa.gov/data/guidelines/ConfIntguide.htm.

White, P., Gunstan, J., Salmond, C. and Crampton, P. (2008). Atlas of Socioeconomic Deprivation in New Zealand NZ*dep*2006. Wellington, Ministry of Health.

Whiteman, D. (2005). "Privacy and medical research." International Medicine Journal **35**: 441-442.

Wilmoth, J. R. and Horiuchi, S. (1999). "Rectangularization Revisited: Variability of Age at Death within Human Populations." Demography **36** (4): 475-495.

Wolfbane Cybernetic Ltd. Website. (2007). "International Classification of Diseases." Retrieved 2nd February 2010, from http://www.wolfbane.com/icd/index.html.

World Health Organization. (2010). "WHO Statistical Information System." Retrieved 4th March 2010, from http://www.who.int/whosis/.

Zeweifel, P., Felder, S. and Meiers, M. (1999). "Ageing of Population and Health Care Expenditure: a Red Herring?" Health Economics **8**: 485-496.

Zingmond, D. S., Ye, Z., Ettner, S. L. and Liu, H. (2004). "Linking hospital discharge and death records - accuracy and sources of bias." Journal of Clinical Epidemiology **57**: 21-29.

APPENDIX A
DATA PROCESSING

Introduction

This appendix is a rather technical section that describes the process of data manipulation undertaken in order to make the data suitable for analyses. The data were obtained in a combination of Microsoft Access and Text formats, and extensive processing was needed to enable the homogenisation of the data and insertion of ICD and NZ*dep* variables for different years.

Six steps were involved in extracting the data from Access and into SPSS, adding the NZ*dep* area deprivation scores, converting the ICD codes into a suitable numerical format for grouping purposes, and developing variables to filter the hospital data. The datasets involved are listed in Table 2.2.1 in Paper One and the steps involved in preparing the data are shown in Figure 2.2.1 in Paper One. The steps covered in this appendix are discussed in the following sections:

Step 1 - Export data from Access to Text, 1974-2000/01
Step 2 - Import data into SPSS from Text, 1974-2006/mid 2008
Step 3 - Add NZdep
Step 4 - Convert ICD codes to numerical format and add ICD description
Step 5 - Create grouped ICD and age variables
Step 6 - Create filter variables for hospital data

See Appendix Tables 2 and 3 for the names of variables included in the data, and in which years data for certain variables are missing. **Note:** hospital variables end with _h and mortality variables end with _m. This enables you to see at a glance whether you are looking at hospital or mortality data and also eases data linkage between hospital and mortality datasets at a later stage (see Appendix B). The variable names in Appendix Tables 2 and 3 give little information of what the variables represent: informative variable labels and other information are added when the data are imported into SPSS. The variables included in the hospital and mortality datasets, along with the variable type and label are listed in Tables A.1 and A.2 respectively. **Note:** All variables are kept in text format until imported into SPSS, when they are changed to string, numeric or date format as appropriate.

Table A.1: Variables in Hospital Dataset

	Name	Type	Label
Original data	year_h	Numeric	Year of hospital discharge
	event_id_h	Numeric	Unique hospital event identifier
	enc_nhi	String	Encrypted NHI, person identifier
	dob_date_h	Date	Date of birth (1988 on)
	dob_year_h	Numeric	Year of birth (before 1988)
	NZ_res_h	String	Residency status/non-NZ dom code identifier
	age_dis_h	Date	Age at hospital discharge
	ethnicgp_h	String	Prioritised ethnicity
	gender_h	String	Gender
	ethnicg1_h	String	Ethnicity 1
	ethnicg2_h	String	Ethnicity 2
	ethnicg3_h	String	Ethnicity 3
	dom_code_h	String	Domicile code
	nhi_dod_h	Date	Date of death (hospital record)
	adm_src_h	String	Admission source
	adm_type_h	String	Admission type
	end_type_h	String	Hospital event end type
	event_end_date_h	Date	Hospital discharge date
	event_start_date_h	Date	Hospital admission date
	fac_type_h	String	Facility Type (from 1988 on)
	event_lvd_h	Numeric	Number of days absent from hospital during stay
	los_h	Numeric	Length of stay in hospital (days) ignoring event_lvd_h and coding day patients as 0.5
	diag01_h	String	Diagnosis ICD code, first
	diag02_h	String	Diagnosis ICD code, second
	diag03_h	String	Diagnosis ICD code, third
	op01_h	String	Operation ICD code, first
	op02_h	String	Operation ICD code, second
	op03_h	String	Operation ICD code, third
	ecode01_h	String	Ecode ICD code, first
	ecode02_h	String	Ecode ICD code, second
	ecode03_h	String	Ecode ICD code, third
	accflag01_h	String	Accident flag, first
Flag variables (discussed in Step 6)	mh_flag_h	String	Mental health flag
	dss_flag_h	String	DSS flag
	respite_flag_h	String	Respite care flag
	rehab_flag_h	String	Rehabilitation flag
	ungroupable_flag_h	String	Ungroupable flag
	boarder_flag_h	String	Boarder flag
	well_babies_flag_h	String	Well babies flag
	short_stay_flag_h	String	Short stay flag
	non_che_flag_h	String	Non-CHE flag
NZdep	Score_h	Numeric	NZ*dep* area deprivation decile score
	Scale_h	Numeric	NZ*dep* area deprivation decile scale

(Step 3)	CAU_h	Numeric	Census Area Unit
	NZdep_version_h	Numeric	NZ*dep* version used
	diag01_num_h	Numeric	Primary diagnosis, recoded numerically
	diag02_num_h	Numeric	Second diagnosis, recoded numerically
	diag03_num_h	Numeric	Third diagnosis, recoded numerically
Numerical	op01_num_h	Numeric	First procedure code, recoded numerically
ICD	op02_num_h	Numeric	Second procedure code, recoded numerically
(Step 4)	op03_num_h	Numeric	Third procedure code, recoded numerically
	ecode01_num_h	Numeric	First ecode, recided numerically
	description_h	String	Primary diagnosis code description
	ICDversion_h	String	ICD version used
Grouped	diaggroup_h	Numeric	Primary diagnosis code, grouped
ICD and	diaggroup_b_h	Numeric	Primary diagnosis code, broad groupings
age	age_grouped_h	Numeric	Age at discharge, grouped
(Step 5)	age_grouped_b_h	Numeric	Age at discharge, broad groupings
	filter1_h	Numeric	Non-NZ residents
	filter2_h	Numeric	Boarders
	filter3_h	Numeric	Obstetrics/Pregnancy Related
Filter	filter4_h	Numeric	Well babies
variables	filter5_h	Numeric	Baby Boarder
(Step 6)	filter6_h	Numeric	Mental Health
	filter7_h	Numeric	Respite care
	filter8_h	Numeric	Rehabilitation
	filter9_h	Numeric	Disability Support Services
	filter10_h	Numeric	Non-CHE
	filter11_h	Numeric	Long stay (<=365 days)
	filter12_h	Numeric	Day patients
	filter13_h	Numeric	Chemotherapy, Radiotherapy, Renal Dialysis.
	filter14_h	Numeric	Same day gastroscopies, colonoscopies and cystoscopies
	filter15_h	Numeric	Blood transfusions
	filter_16_h	Numeric	Publicly funded events in private facilities without procedure
	filter0_h	Numeric	Not filtered, equals 1 when none of the filters apply, 0 otherwise

Table A.2: Variables in Mortality Dataset

	Name:	Type	Description
Original Data	regyear_m	Numeric	Registration year of death
	enc_nhi	String	Encrypted NHI, person identifier
	dob_m	Date	Date of birth
	age_m	Date	Age at death
	cob_m	Numeric	Country of birth
	yrsinNZ_m	Numeric	Years in New Zealand
	occupation_m	String	Occupation
	ethnicgp_m	String	Prioritised ethnicity
	gender_m	String	Gender
	ethnicg1_m	String	Ethnicity 1
	ethnicg2_m	String	Ethnicity 2
	ethnicg3_m	String	Ethnicity 3
	dom_cd_m	String	Domicile code
	dod_m	Date	Date of death
	icd01_m	String	Cause of death ICD code, primary
	icd02_m	String	Cause of death ICD code, second
	icd03_m	String	Cause of death ICD code, third
	death_certifier_m	Numeric	Death certifier code
	death_info_src_m	Numeric	Primary source of cause of death information
NZdep info (Step 3)	Score_m	Numeric	NZ*dep* area deprivation decile score
	Scale_m	Numeric	NZ*dep* area deprivation decile scale
	CAU_m	Numeric	Census Area Unit
	NZdep_version_m	Numeric	NZ*dep* version used
Numerical ICD (Step 4)	icd01_num_m	Numeric	Primary cause of death code, recoded numerically
	description_m	String	Primary cause of death code description
	ICDversion_m	String	ICD version used
Grouped ICD and age (Step 5)	icdgroup_m	Numeric	Grouped cause of death
	icdgroup_b_m	Numeric	Grouped cause of death, broad groupings
	age_grouped_m	Numeric	Age at death, grouped
	age_grouped_b_m	Numeric	Age at death, broad groupings

Step 1 - Export data from Access to Text, 1974-2000/01

The older data are stored as individual year tables in Microsoft (MS) Access, for years 1974-2000 (mortality) and 1974-2001 (hospital). Given that the data have been collated over many years and are stored over many different tables, the variable names differ. It is necessary to rename the variables and make the variable type and size consistent across all tables for later merging into one, 'flat' file.

This is achieved by making field names consistent across all tables, deleting unnecessary fields (e.g. hospital diagnoses to 4 to 20 in 2006) and adding empty fields for

tables where these data are missing (e.g. prioritised ethnicity, years before 1998 (hospital) and 1996 (mortality)). This process is sufficient for the mortality data, which is now ready for exporting to text. However, due to very large table sizes, problems arise for the hospital data:

Problem 1: For hospital date of birth, there are two different formats. It is a date from 1988 on (dd/mm/yyyy), and numerical before 1988 (yyyy). From 1988 it needs to be coded as a date, before this it needs to be coded numerically. **Solution 1:** split into two variables, dob_date_h, which is a date variable and is blank before 1988, and dob_year_h, which is numerical and is blank from 1988.

Problem 2: There is no hospital event ID for years 1974-1987. **Solution 2:** insert a large random number and check it is unique in each year (there are no duplicates).

Problem 3: Hospital event ID may not be unique across all years, especially for 1974-1987. This is not a problem while there is a separate table for each year, but it would be problematic when the data are converted into 'flat' file and imported into SPSS. **Solution 3:** Make event ID unique for each year by inserting year at start of event ID. This is achieved by Visual Basic (VB) code as shown below, where 'yr_88_02' is the table for 1988.

```
Sub random ()
s_time = Time
Set rst_1988 = CurrentDb.OpenRecordset ("yr_88_02")
With rst_1988
   Do While Not .EOF
   i_recordnbr = i_recordnbr + 1
      s_text = !event_id
       s_evntid = "1988" & s_text
      .Edit
         !event_id = s_evntid
      .Update
skip_update:
      .MoveNext
   Loop
End With
MsgBox "I updated " & i_recordnbr & " records in " & Time - s_time & "
seconds"
End Sub
```

The field type needs to be changed to text and the field size needs to be made consistent across all tables before exporting the data to text. It is best to keep all variables

stored as text until they are imported into SPSS. If domicile code is changed to a numerical variable for instance, where 0 is the first digit, this will be deleted.

Problem 4: For hospital tables, the database is too large to change field type or size in design view. The memory is insufficient, even after compacting and removing all other tables from the database. The field size for most variables is unnecessarily large (250 characters), and because of it is impossible to export to text due to memory restrictions. **Solution 4:** It is not possible to change the field size in design view, but it cannot be exported because the field size too large. VB is used to change the field size and type:

```
Sub insert_table ()
sql_1 = "Insert into template select * from yr_79_02"
DoCmd.RunSQL sql_1
End Sub
```

This forces the data from the year 79 table into a template table with correct field names, size and type, where ‘template’ is a blank table with correct field names, size and type, and ‘yr_79_02’ is the table for year 79, with correct field names but incorrect field size and/or type.

The hospital data can now be exported to text.

Step 2 – Import data into SPSS from text, 1974-2006/mid08

The new data obtained from the New Zealand Ministry Of Health (MOH) is stored as individual year text files for hospital data (2002-mid2008), and as a single text file for the mortality data (2002-2006). The text files created in Step 1 and the new data can now be imported into SPSS using the ‘read text data’ dialogue boxes. The correct variable names and type are specified, and because the process needs to be repeated several times, import templates for the new and old data are saved. However for the new data the correct variable names and type need to be defined during the import, with unnecessary variables not imported. After importing, extra empty variables are created to make it comparable with the variables present in the old data.

Once imported, the SPSS datasets can be merged to create one big dataset for mortality data (approx 800,000 events) and one big dataset for hospital data (approx 20 million events).

Step 3 – Add NZdep

NZ*dep* scores apply to meshblocks (very small areas which we do not have access to) or Census Area Units (CAUs) (areas of approx 2000 individuals). In the hospital and mortality data, area of residence is given by domicile code, which is supposed to link one to one with CAUs. NZ*dep* consists of a scale, a large numerical value, and a score, a value between 1 and 10 which is calculated from the scale to represent deciles of deprivation. There are two processes needed when inserting a relevant NZ*dep* score for each event, retrieval of the NZ*dep* score for the given CAU and year, and mapping of domicile code to CAU.

We have domicile code information from 1988. Domicile codes, and area boundaries for domicile codes change over time, they are redefined at each census (1986, 1991, 1996, 2001, 2006). NZdep scores are calculated at each census, but only from 1991.

According to the National Minimum Data Set 2009 data dictionary for hospital events[30] and the 2004 data dictionary for mortality events (no longer available online) the domicile codes recorded are as shown in Table A.3.

Table A.3: Domicile Code versions used in Hospital and Mortality Data

	Hospital		Mortality	
Year of census	From	Until	From	Until
1991*	1988	30th June 1998	1988	1997
1996	1st July 1998	30th June 2003	1998	2002
2001	1st July 2003	30th June 2008	2003	
2006	1st July 2008			

*Data from 1988-1992 was recorded using the 1986 code but has since been mapped to 1991

Access is used to derive a table for each census year, with *all* CAUs from 1991-2006 (regardless of if they were redundant in that census year) domicile code (which is needed for merging with the mortality and hospital data) and relevant NZ*dep* scale and score. However the NZ*dep* Scale and Score will be blank if the dom code/CAU was not in use that census year. The reason it is done it this way is to allow identification between real non-matches where there is no CAU relating to a domicile code (where the domicile code is incorrect) and where the domicile code is relating to a CAU that does not exist for the current census year, but **did** exist in the past/will exist at the next census year.

[30] Available at http://www.nzhis.govt.nz/moh.nsf/pagesns/238

Tables listing NZ*dep* by CAU are available online from the University of Otago website (University of Otago Website 1988-2001). Access is used to create a table from these sources containing all CAU codes for each census year, with relevant NZ*dep* scale and score, but no domicile code. An extract is shown in Table A.4. There are four rows for every CAU code - one for every census year, with the relevant NZ*dep* score and scale. If a CAU code is only valid in one census year, it still has 4 rows (one for every census year), but there would only be NZ*dep* information available for one of these years.

Table A.4

CAU	NZdep Score	NZdep Scale	Year
609023	926	2	1991
609023	923	2	1996
609023	920	2	2001
609023	935	3	2006
609024	895	1	1991
609024	902	1	1996
609024			2001
609024			2006
609025	922	2	1991
609025	893	1	1996
609025			2001
609025			2006

So in this case, CAU 609023 was a valid CAU for all four census years, increasing from decile 2 to 3 over this period. CAUs 609024 and 609025 were only valid for census years 1991 and 1996.

Now the relevant domicile codes need to be added. The CAU/domicile code mapping table from NZHIS website (New Zealand Health Information Service (NZHIS) Website 2008) is copied into Access. Three columns are of interest, Dom Code, CAU and Year. A ‘make new table’ query is run to merge these two tables into a new table with CAU, NZ*dep* Score, NZ*dep* Scale, Year **and** domicile code. The join is dependent on CAU and Year in both tables being the same, and includes all records from the first table (from the University of Otago website) and only the records from the second table (from NZHIS website) that match. An extract of this table is shown in table A.5.

Table A.5

CAU	NZdep Score	NZdep Scale	Year	Domicile code
609023	926	2	1991	3066
609023	923	2	1996	3066
609023	920	2	2001	3066
609023	935	3	2006	3066
609024	895	1	1991	3067
609024	902	1	1996	3067
609024			2001	3067
609024			2006	3067
609025	922	2	1991	3106
609025	893	1	1996	3106
609025			2001	3106
609025			2006	3106

This table is then split into four separate tables, one for each census year, exported to text and imported into SPSS, where the CAU, NZdep Score and NZ*dep* Scale are merged with the mortality and hospital data, using domicile code as the 'match cases on' variable.

It was initially intended to add the NZ*dep* variables to the data in Access before exporting the data to text. However because the hospital data switches between domicile codes (e.g. 1991 to 1996 boundaries) in mid-year (30th June 1998), it is necessary to for instance merge with 1991 CAUs for the first half of 1998, and with 1996 CAUs for the second half of 1998. This can be achieved using the advanced date and time functions available in SPSS. The date variables are stored as numerical values in Access and there is no simple 'design view' way of telling Access that they are dates. NZ*dep* info needs to be added once the data are already in SPSS. In addition to the CAU, NZ*dep* score and scale, variables are created in SPSS (NZdep_version _m and NZdep_version _h) that identify which NZdep version to use for each event, based on the registration year of death/date of hospital discharge. These are created using the SPSS syntax below:

```
IF regyear_m>=1988 AND regyear_m<1998 NZdep_version _m=1991.
IF regyear_m>=1998 AND regyear_m<2003 NZdep_version _m=1996.
IF regyear_m>=2003 NZdep_version _m=2001.
EXECUTE.

IF event_end_date_h>=DATE.DMY (01,01,1988) AND
        event_end_date_h<DATE.DMY (01,07,1998) NZdep_version _h=1991.
IF event_end_date_h>=DATE.DMY (01,07,1998) AND
        event_end_date_h<DATE.DMY (01,07,2003) NZdep_version _h=1996.
IF event_end_date_h>=DATE.DMY (01,07,2003) NZdep_version_h=2001.
EXECUTE.
```

Four variables are created for the hospital and mortality datasets:

NZdep_version_h

CAU_h

Scale_h

Score_h

NZdep_version_m

CAU_m

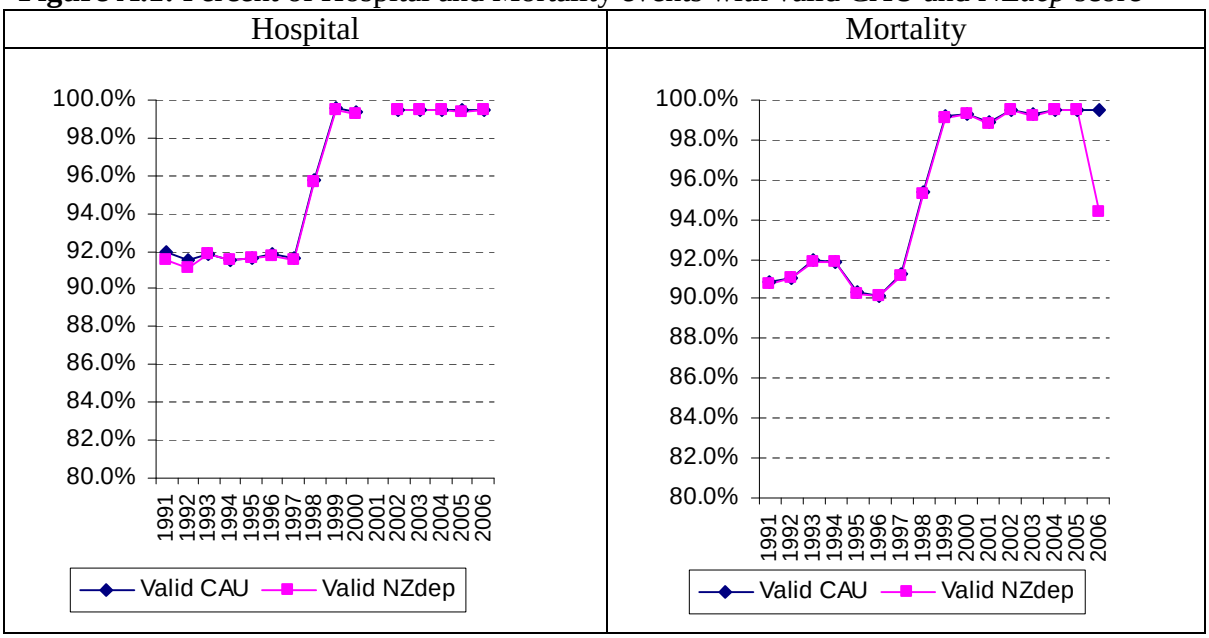
Scale_m

Score_m

Once the NZ*dep* variables have been added to the hospital and mortality tables, it is apparent that there are a relatively large proportion of non-matches (up to about 10%). A bit of investigation has suggested that this is largely due to the reporting of 'old' domicile codes: there were lots of cases where the CAU value was present, but not the scale or score. This would represent, for instance, reporting a 1996 domicile code in 2004, when it may have ceased to exist. The percentage missing decreases from about 1999, moderately for hospital (down to about 4 or 5%) and sharply for mortality (down to about 0.5%), which suggests better reporting of current domicile codes for mortality data.

The changes in the domicile code versions used can be clearly seen when the percentage of events with a relevant CAU is plotted. See figure A.1. In 1991 and 1992, the proportion of mortality events with a valid CAU was quite low, about 92 percent, and of these events only 99.5 percent could be matched to a valid NZ*dep* score. This links in with the updating of data from 1988-1992, from the 1986 domicile code version to the 1991 domicile code version. The large, linear increase in the percent of mortality events with a valid CAU between 1997 and 1999 corresponds with the changing from the 1996 domicile code version to the 2001 domicile code version that occurred halfway through 1998. In short, mapping domicile codes to CAUs when the 1991 domicile code version was used, was much less successful than when the 1996 or 2001 versions were used. And identifying a valid NZ*dep* score for a CAU was usually highly successful, but least so in 1991 and 1992 when domicile codes were originally recorded in the 1986 version, but had been mapped to the 1991 version.

Figure A.1: Percent of Hospital and Mortality events with valid CAU and NZ*dep* score



Step 4 – Convert ICD codes to numerical format and add ICD description

There are 4 different ICD versions used in the data, is shown in Table A.6, the ICD versions reported do not change at the same time for hospital and mortality data. In the mortality data, icd01_m, icd02_m and icd03_m (the first three causes of death) use ICD codes, and in the hospital data, diag01_h, diag02_h, diag03_h (the first three diagnosis codes), op01_h, op02_h, op03_h (the first three procedure codes), ecode01_h, ecode02_h and ecode03_h (the first three external cause codes) use ICD codes.

Problem 5: SPSS will not recognise ranges unless the variable is numerical. The fourth digit of ICD 8 and 9 is sometimes an X (to denote any number between 0 and 9). Codes in ICD8, 9 and 9CM sometimes start with a 0, which when stored as a numerical value disappears. ICD10 codes and some codes in ICD9CM use a letter at the start of the code.

Three digit codes for ICD 8 and 9 are perfectly adequate for our purposes. **Solution 5:** delete fourth digit (which is sometimes an X). Export diagnosis/cause of death variable to text and import again with a break between each digit. This results in four separate variables, each one digit wide. Delete fourth variable, export to text and re-import with no breaks between variables.

Table A.6: ICD versions used in Hospital and Mortality Data

	Hospital	Mortality
1974	ICD8	ICD8
1975		
1976		
1977		
1978		
1979		ICD9
1980	ICD9	
1981		
1982		
1983		
1984		
1985		
1986		
1987		
1988	ICD9 CM	ICD9 CM
1989		
1990		
1991		
1992		
1993		
1994		
1995		
1996		
1997		
1998		
1999		
2000		ICD10 AMv2
2001		
2002	ICD10 AMv2	
2003		
2004		
2005		
2006		
2007		
2008		

For ICD9CM and 10, Access can deal with ranges that start with letters using queries, however this is unpractical and time consuming. There is also still the problem of codes starting with a 0. To make the ICD coding flexible there must be one-to-one conversion from the ICD code to a numerical variable, which can then easily be grouped in SPSS to give diagnoses/cause of death groups. These groups also need to be easily changeable and flexible.

Excel codebooks for ICD9CM and 10AM were obtained from Chris Lewis at the Ministry of Health and the Waikato server, and codebooks for ICD8 and 9 were obtained from the Wolfbane website (Wolfbane Cybernetic Ltd. Website 2007). Searchable versions of ICD9CM and 10 are available online,[31] although the latter is just for ICD10, not the Australian Modification. **Problem 6:** given that the later ICD versions consist of many thousands of codes, with additional procedure codes (the 'op' variables in the hospital data) and ecodes, it is unfeasible to write the syntax manually to convert these codes to numerical values. **Solution 6:** A way around this is to use Excel to write the code automatically. Using the contatenate function in Excel, SPSS syntax is written to recode each ICD code into a numerical value, with each ICD code relating to a different numerical value. An example for ICD9CM is given in table A.7.

[31] Available at http://icd9.chrisendres.com/ and http://apps.who.int/classifications/apps/icd/icd10online2006/ respectively.

Table A.7: Generating SPSS code in Excel, example from ICD9CM

	A	B	C	D
1	ICD Code	Description	ICD_recode	SPSS Syntax
2	0010	Cholera due to Vibrio cholerae	1	RECODE diag01_h ('0010'=1) INTO diag01_num_h.
3	0011	Cholera due to Vibrio cholerae	2	RECODE diag01_h ('0011'=2) INTO diag01_num_h.
4	0019	Cholera, unspecified	3	RECODE diag01_h ('0019'=3) INTO diag01_num_h.
5	0020	Typhoid fever	4	RECODE diag01_h ('0020'=4) INTO diag01_num_h.
6	0021	Paratyphoid fever A	5	RECODE diag01_h ('0021'=5) INTO diag01_num_h.
7	0022	Paratyphoid fever B	6	RECODE diag01_h ('0022'=6) INTO diag01_num_h.
8	0023	Paratyphoid fever C	7	RECODE diag01_h ('0023'=7) INTO diag01_num_h.
9	0029	Paratyphoid fever, unspecified	8	RECODE diag01_h ('0029'=8) INTO diag01_num_h.
10	0030	Salmonella gastroenteritis	9	RECODE diag01_h ('0030'=9) INTO diag01_num_h.
11	0031	Salmonella septicaemia	10	RECODE diag01_h ('0031'=10) INTO diag01_num_h.

Where diag01_h is the ICD code (column A) and diag01_num_h is a new, numerical variable (with values taken from column C).

The code in cell D2 is:

```
=CONCATENATE ("RECODE diag01_h ('",A2,"'=",C2,") INTO
diag01_num_h.")
```

This cell can be copied down to the last ICD code. These cells can be pasted directly into the SPSS Syntax window.

The same process can also be used to insert a description string variable into SPSS. While this is not really necessary, it allows you to see at a glance what condition a particular code relates to. Unlike the numerical ICD variable, in the syntax the descriptions need to be in inverted commas ‘’ because they are string values. This can be achieved by putting a column either side of the description column in Excel, containing a space followed by ‘ (in Excel, putting a ‘ in a cell on its own makes it disappear, a space is required before it), as shown in Table A.8.

Table A.8

'	Cholera due to Vibrio cholerae	'
'	Cholera due to Vibrio cholerae el tor	'
'	Cholera, unspecified	'
'	Typhoid fever	'
'	Paratyphoid fever A	'
'	Paratyphoid fever B	'
'	Paratyphoid fever C	'
'	Paratyphoid fever, unspecified	'
'	Salmonella gastroenteritis	'
'	Salmonella septicaemia	'

This can then be exported to text and re-imported back into excel with no breaks, and ready for using with the concatenate function to produce SPSS syntax for a description variable, as shown in Table A.9.

Table A.9

' Cholera due to Vibrio cholerae '
' Cholera due to Vibrio cholerae el tor '
' Cholera, unspecified '
' Typhoid fever '
' Paratyphoid fever A '
' Paratyphoid fever B '
' Paratyphoid fever C '
' Paratyphoid fever, unspecified '
' Salmonella gastroenteritis '
' Salmonella septicaemia '

However, for the description syntax, when pasted directly into the syntax window, SPSS inserts “” around each line. This may be due to the additional ‘’ in the code that is necessary for string variables. The description syntax needs to be copied to text, imported into SPSS as data, and then copied and pasted into the syntax window. This avoids SPSS inserting “” around each line, which prevents it running correctly.

In addition to the numerical and description ICD variables, variables are created for the hospital and mortality datasets that identify the ICD version in use for each event. This is similar to the NZdep_h and NZdep_m variables created in step 3.

```
STRING ICDversion_m (a5).
IF regyear_m<1979 ICDversion_m='8'.
IF regyear_m>=1979 AND regyear_m<1989 ICDversion_m='9'.
IF regyear_m>=1989 AND regyear_m<2000 ICDversion_m='9CM'.
IF regyear_m>=2000 ICDversion_m='10AM'.
EXECUTE.
```

```
STRING ICDversion_h (a5).
IF year_h<1980 ICDversion_h='8'.
IF year_h>=1980 AND year_h<1988 ICDversion_h='9'.
IF year_h>=1988 AND year_h<2002 ICDversion_h='9CM'.
IF regyear_m>=2002 ICDversion_h='10AM'.
EXECUTE.
```

The Excel ICD codebooks with the SPSS syntax are kept as reference for when creating grouped ICD variables. Using the techniques above, nine new variables are created for the hospital data:

diag01_num_h

diag02_num_h

diag03_num_h

ecode01_num_h

op01_num_h

op02_num_h

op03_num_h

description_h

ICDversion_h

And three new variables are created for the mortality data:

icd01_num_m

description_m

ICDversion_m

Note: only the primary cause of death will be examined in the analysis, so only the first ICD code needs to be converted to numerical format. This could also be done for the second and third cause of death codes. All three diagnosis and procedure codes and the first ecode will be required to calculate the filters for the hospital data, so these variables are all converted to numerical format.

Note: This process has to be done separately for each ICD version using a DO IF/END IF command to isolate particular time periods/ICD versions. Given that the numerical values will not relate to the same ICD codes across versions, the ICD codes also need to be grouped separately for each ICD version.

Step 5 – Create grouped ICD and age variables

Two grouped variables are created for both primary diagnosis/cause of death, and age. Both narrow and broad groupings are calculated. For age, the narrow grouping is 5-year age groups:

```
RECODE age_dis_h (0 thru 4=1) (5 thru 9=2) (10 thru 14=3) (15 thru 19=4) (20
thru 24=5) (25 thru 29=6) (30 thru 34=7) (35 thru 39=8) (40 thru 44=9) (45 thru
49=10) (50 thru 54=11) (55 thru 59=12) (60 thru 64=13) (65 thru 69=14) (70 thru
74=15) (75 thru 79=16) (80 thru 84=17) (85 thru Highest=18)
   INTO age_grouped_h.
EXECUTE.

RECODE age_m (0 thru 4=1) (5 thru 9=2) (10 thru 14=3) (15 thru 19=4) (20 thru
24=5) (25 thru 29=6) (30 thru 34=7) (35 thru 39=8) (40 thru 44=9) (45 thru 49=10)
(50 thru 54=11) (55 thru 59=12) (60 thru 64=13) (65 thru 69=14) (70 thru 74=15)
(75 thru 79=16) (80 thru 84=17) (85 thru Highest=18)
   INTO age_grouped_m.
EXECUTE.
```

And the broader age groups are:

```
RECODE age_dis_h (0 thru 19=1) (20 thru 39=2) (40 thru 49=3) (50 thru 59=4)
(60 thru 69=5) (70 thru 79=6) (80 thru 89=7) (90 thru Highest=8)
INTO age_grouped_b_h.
EXECUTE.

RECODE age_m (0 thru 19=1) (20 thru 39=2) (40 thru 49=3) (50 thru 59=4) (60
thru 69=5) (70 thru 79=6) (80 thru 89=7) (90 thru Highest=8)
INTO age_grouped_b_m.
EXECUTE.
```

For primary cause of death/diagnosis, the narrow groups included all main diagnostic categories given in the ICD10 codebook on the WHO website, and several sub-categories for neoplasms and circulatory, for which there are large numbers of hospital events and deaths. Identifying the same diagnostic groups for ICD9CM, 9, and 8 is relatively easy, unlike tracking individual codes across ICD versions. The broader ICD groups were calculated by recoding the narrow groups into fewer categories and focussed on the main diagnostic categories, grouping these together for uncommon causes. The categories for the broad ICD groups are shown in Table A.10. Results in the analysis are generally presented by broad ICD group. The narrow ICD groups were used for examining cause of death/diagnosis trends for the main causes of death in more detail.

Table A.10: Broad ICD groupings, by ICD version

ICD codes				Description
ICD10AM	ICD9CM	ICD9	ICD8	
A00-B99	001-139	001-139	000-136	Certain Infectious and Parasitic
C00-D48	140-239	140-239	140-239	Neoplasms
D50-D89	280-289	280-289	240-279	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90	240-279	240-279	280-289	Endocrine, nutritional and metabolic diseases
F00-F99	290-319	290-319	290-315	Mental and behavioural disorders
G00-H99	320-389	320-389	320-389	Nervous system and sense organs, eye, adnexa, ear and mastoid process
I00-I99	390-459	390-459	390-429	Circulatory system
J00-J99	460-519	460-519	460-519	Respiratory system
K00-K93	520-579	520-579	520-577	Digestive system
L00-M99	680-709	680-709	680-709	Diseases of the skin, subcutaneous tissue, musculoskeletal and connective tissue
N00-N99	580-629	580-629	580-629	Genitourinary system
O00-Q99	630-679	630-677	630-678	Pregnancy, childbirth, puerperium, conditions originating in perinatal, congenital malformations, deformations and chromosomal abnormalities
R00-R99	780-799	780-799	780-796	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
S00-T98	800-999	800-999	800-999	Injury, poisoning and certain other consequences of external causes (ICD9 E codes to determine cause/intent) HOSPITAL ONLY
V01-Y98	E800-E99	800-999	800-999	External causes of morbidity and mortality (ICD9 E codes to determine cause/intent) MORTALITY ONLY
Z00-Z99	V01-V89	V01-V82	Y00-Y89	Factors influencing health status and contact with health services HOSPITAL ONLY
U00-U99				Codes for special purposes (ICD10 only) HOSPITAL ONLY

The eight variables added to the data at this stage are:

age_grouped_h

age_grouped_b_h

diaggroup_h

diaggroup_b_h

age_grouped_m

age_grouped_b_m

icdgroup_m

icdgroup_b_m

Step 6 – Create filter variables for hospital data

The filters used in the monograph were replicated as closely as possible. This was a bigger challenge for ICD 8 (before the scope of the monograph) and ICD10AM (the monograph only goes up to 2000). In addition to reading the documentation on the Waikato server regarding previous filtering of the data, details of how the Ministry of Health filters hospital data was obtained, from in Appendix 1 and 2 of the hospital throughput reports for 2004/05 and 1998/99 respectively[32]. There are 16 filter variables, although few of them can be implemented over the full time period 1974-mid2008. The filter variables are shown in Table A.11.

Table A.11: Filter Variable Descriptions

filter1_h	Non-NZ residents
filter2_h	Boarders
filter3_h	Obstetrics/Pregnancy Related
filter4_h	Well babies
filter5_h	Baby Boarder
filter6_h	Mental Health
filter7_h	Respite care
filter8_h	Rehabilitation
filter9_h	Disability Support Services
filter10_h	Non-CHE
filter11_h	Long stay (>365 days)
filter12_h	Day patients
filter13_h	Same day Chemotherapy, Radiotherapy, Renal Dialysis.
filter14_h	Same day gastroscopies, colonoscopies and cystoscopies
filter15_h	Same day Blood transfusions
filter16_h	Publicly funded private hospital events with no procedure
filter0_h	Not filtered, equals 1 when none of the filters apply, 0 otherwise

An extra variable used in the monograph was a Starship[33] filter. However when examined, the Starship variable was found to contain no data, so Starship patients could not be filtered out. An additional filter was implemented to exclude publicly funded hospital events at private facilities for which there was no procedure (filter 16). There was a large increase in bed days between 2003 and 2006, which are thought to be due to the Ministry of Health temporarily including rest homes and other long-stay, non-acute facilities in the dataset.

[32] Available at http://www.moh.govt.nz/moh.nsf/indexmh/hospital-throughput-0405?Open and http://www.moh.govt.nz/moh.nsf/pagesmh/786?Open

[33] Starship is a childrens hospital in Auckland, events taking place at Starship hospital were temporarily included in the hospital events dataset.

Tracking individual codes across ICD versions is painstaking and time consuming. Several of the filters relied on procedure codes, but procedure codes were not used before ICD9CM. In the old data, a range of 'flag' variables that were used in the filtering process are retained. However there were two versions of many of the 'flag' variables (such as respite, rehab, DSS, mental) present, and it is not always clear which one was an MOH flag, and which was a flag created for the monograph, so these have been avoided where possible. In addition, none of the 'flag' variables go back to 1974, and it is often unclear how these flag variables were calculated. Where a filter can be calculated through diagnosis or procedure codes instead of a flag (as for mental health, respite care and rehabilitation) the flag is not used. However this is not possible for all of the filters and for all ICD versions. Details of the criteria used to calculate the filters are given in Table A.12.

Table A.12: Calculation of the Filter Variables (by ICD code)

	ICD version			
	ICD10	ICD9CM	ICD9	ICD8
1	dom_code_h = 9999	dom_code_h = 9999	dom_code_h = 9999	dom_code_h = 9999
2	diag01_h Z763, Z764 First procedure Null	diag01_h V650 First procedure Null	diag01_h V65	diag01_h Y71
3	diag01_h O00-O99, Z34, Z35, Z36, Z37, Z50	diag01_h 630-679, V22, V23, V24, V27, V28	diag01_h 630-677	diag01_h 630-678
4	diag01_h Z38	diag01_h V30-39	diag01_h V30-39	diag01_h Y80-Y89
5	diag01_h Z762 op01_h null	diag01_h V201 op01_h null	diag01_h V20 op01_h null	diag01_h Y62 op01_h null
6	diag01_h Z046, F20-F639, F68-F99 op01_h null	diag01_h V701, 295-301, 306-319 op01_h null	diag01_h 295-301, 306-319	diag01_h 295-301, 305-315
7	MOH flag or diag01_h Z59	MOH flag	MOH flag	
8	MOH flag or diag01_h Z39	MOH flag or diag01_h V57	MOH flag	
9	MOH flag	MOH flag	MOH flag	
10		MOH flag	MOH flag	
11	los_h>365 days	los_h>365 days	los_h>365 days	los_h>365 days
12	los_h = 0.5 days except when event_end_type_h = "DD"	los_h = 0.5 days except when event_end_type_h = "DD"		
13	diag01_h or diag02_h Z510-Z512 diag01_h or diag02_h Z292 diag01_h or diag02_h Z49	op01_h 9925 or diag01_h V581, V0739 op01_h 9221-9229 or diag01_h V580 op01_h 3995, 5498 or diag01_h V560, V568		

	Second and Third procedure null los_h = 0.5 non-acute	Second and Third procedure null los_h = 0.5 non-acute		
14	op01_h 3681201…* op02_h null or 3681201…** op03_h null los_h = 0.5 non-acute	op01_h 4413-4414, 4513-4516, 4523-4525, 5732-5733 Second procedure null or as above. op03_h null los_h = 0.5 non-acute		
15	diag01_h Z7513 OR op01_h 1370601, 1370602, 1370603, 9206000 op02_h null, 1370601, 1370602, 1370603, 9206000 op03_h null los_h = 0.5 non-acute	diag01_h V582 OR op01_h 9902-9905 op02_h null, 9902-9905 op03_h null los_h = 0.5 non-acute		
16	fac_type_h=”02” and	fac_type_h=”02” and		
	op01_h null	op01_h null		

Not all of the filters were used for the analyses in this paper, and slightly different filters were used for national analyses and analyses by area deprivation. See Paper One, Chapter Two for further information. Appendix Figures 3 and 4 show graphically the number of discharges and bed days for each filter variable over time and by gender.

Conclusions

The data required a considerable amount of processing to make it suitable for analysis. The decisions made during the processing stage, particularly the creation of filter variables, have substantial impacts on the results from analyses, and their interpretation, therefore accountability of the data processing is essential. Filtering was essential to allow time-trend analyses, exclude those hospital events which do not represent ill health, allow comparability between male and female hospital utilisation, and to smooth the effect of coding changes and short term policy changes. Without filtering, hospital data are a very poor proxy for morbidity. While necessary, the filtering process was rather severe, with up to half of all bed days excluded from the analysis. Other aspects of the data processing which will affect the integrity of results are the integration of the NZ*dep* values and the

recoding and grouping of the ICD variables. If errors were made in incorporating these variables, it would make analyses over time problematic.

APPENDIXES

APPENDIX B
DATA LINKAGE AND PREPARATION FOR PAPER THREE

Introduction

This appendix describes the steps (1 to 11) required to prepare the data for the longitudinal analyses presented in Paper Three. Both the methods used in Paper Three (reference point modelling and person-month logistic regression) require individual record linkage between the hospital records and between the hospital and mortality records. This is discussed in Step 1.

Most variables discussed will be defined in the syntax presented. However some variables are not. Please refer back to variable tables A.1 and A.2 on pages 228-230 for the original variables present in the hospital and mortality datasets. Steps 1 to 6 are similar for both the reference point modelling and the person-month logistic regression, and are presented for the reference point modelling analysis. Steps 7 and 8 are specific to reference point modelling. Step 9 describes the data preparation for the person-month logistic regression, and how this differs from the data preparation for reference point modelling as shown in steps 1 to 6. Also presented in Step 9 are the conversion of the data from person-event to person-month format and other preparation required prior to conducting person-month logistic regression.

The production of output from the reference point modelling and person-month logistic regression is described in Steps 10 and 11 respectively. This is followed by a conclusion.

Step 1 – Individual Record Linkage

The first step in linking the hospital and mortality datasets is to check for duplicates. The individual identifier used to link records is the encrypted National Health Index (NHI), enc_nhi. In the mortality dataset there can only be one record for every unique enc_nhi (person). If this was not the case the data could not be merged. In the hospital dataset there can be several records for every enc_nhi, one for each hospital event. The mortality dataset therefore needs to be checked for duplicate enc_nhi values, and duplicates deleted. The

dataset used is called mort_nhi.sav, this is the mortality dataset including only records which have an enc_nhi value (deaths registered from 1990 onwards). There were approximately 120 records in this period with a missing enc_nhi value which were deleted from the mort_nhi.sav dataset.

However because this dataset will be transposed from variables to cases and includes hospital records with no linked mortality record (H'), mortality records with no linked hospital record (M') and linked hospital and mortality records (MH) it is necessary for ease of interpretation of the final dataset to add a variable indicating which dataset the record originally came from. For the mortality dataset this is mortds and for the hospital dataset this is hospds.

```
GET file 'filepath:\mort.sav'
/KEEP regyear_m enc_nhi dob_m dod_m gender_m scale_m icdgroup_b_m.
DATASET NAME mort.

SORT CASES by enc_nhi (A) dod_m (A).
EXECUTE.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

IF (fileseq=1 OR enc_nhi NE LAG (enc_nhi)) morbseq=1.
IF (fileseq NE 1 AND enc_nhi=LAG (enc_nhi)) morbseq=LAG (morbseq)+1.
VARIABLE LABEL morbseq "Morbidity sequence".
FORMATS morbseq (F1).
EXECUTE.

FREQUENCIES morbseq.
EXECUTE.

SELECT IF morbseq=1.
EXECUTE.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

COMPUTE mortds = 1.
FORMATS mortds (F1).
EXECUTE.

DELETE VARIABLES morbseq.
SAVE outfile 'M:\mort_nhi.sav'.
DATASET NAME mortnhi.
EXECUTE.
```

The morbidity sequence variable (morbseq) increases sequentially for each record relating to the same individual. It is used extensively when manipulating the hospital

dataset, and is used here to identify where an individual has more than one mortality record. A morbidity sequence value of 1 indicates the first record for that enc_nhi value. Where it is 2, there is a second record for the same enc_nhi, indicating a duplicate. There are 65 duplicate records where morbseq is 2. Given that the data is sorted within enc_nhi by ascending date of death (dod_m), records where morbseq is 2 are deleted. This leaves the earlier death date for that enc_nhi. The file sequence (fileseq), indicates the 'correct' order of the records when enc_nhi is sorted ascending, and is used when calculating the morbidity sequence. However after deleting duplicates the file sequence is now incomplete and so is computed again, and the dataset is saved under a new name (mort_nhi.sav). The hospital dataset for records with an enc_nhi number (hosp_nhi.sav) is now merged with the mortality dataset (mort_nhi.sav) and saved as merged_all.sav. This includes ALL individuals in the datasets if they have either a mortality or a hospital event. This is the entire area M', MH and H' in the conceptual diagram on page 158. However the merge means that for individuals with linked hospital and mortality records, the mortality record only appears for the first hospital record. It is necessary to copy the mortality variables (dob_m dod_m gender_m regyear_m icdgroup_b_m scale_m) and the dataset variable down the file to be the same for all records relating to the individual. Events with no valid encrypted nhi number are excluded (1790 records).

The variable dataset is created from hospds and mortds to identify which dataset the record came from. That is, whether the record came from the mortality dataset, the hospital dataset, or both (linked mortality *and* hospital records). To achieve this, an extra variable was created in both the hospital and mortality datasets before linkage: hospds and mortds respectively. These variables are computed to be equal to 1. Thus, in the linked dataset containing all hospital records, all mortality records and linked mortality and hospital records, if hospds is equal to 1 and mortds is missing, the record came from only the hospital dataset. If hospds is missing and mortds is equal to 1, the record came from only the mortality dataset. And if both hospds and mortds are equal to 1, linked hospital *and* mortality information are present in the record. The variable created from this information (dataset) is useful later in interpreting the records when the data is in person-month format. One further adjustment necessary is to convert the gender variables from the hospital and mortality datasets from string to numeric variables.

```
GET file 'filepath:\hosp.sav'
/KEEP year_h event_id_h enc_nhi dob_date_h gender_h end_type_h event_end_date_h
event_start_date_h scale_h.
```

```
DATASET NAME hospnhi.
EXECUTE.
SORT CASES by enc_nhi (A) event_start_date_h (A) event_end_date_h (A).
EXECUTE.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

SELECT IF fileseq>=1790.
EXECUTE.

COMPUTE hospds = 1.
FORMATS hospds (F1).
EXECUTE.

MATCH FILES /FILE=*
  /FILE='mortnhi'
  /BY enc_nhi.
EXECUTE.
DATASET CLOSE mortnhi.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

IF (fileseq=1 OR enc_nhi NE LAG (enc_nhi)) morbseq=1.
IF (fileseq NE 1 AND enc_nhi=LAG (enc_nhi)) morbseq=LAG (morbseq)+1.
VARIABLE LABEL morbseq "Morbidity sequence".
FORMATS morbseq (F1).
EXECUTE.

IF (hospds=1) dataset=1.
IF (mortds=1) dataset=2.
IF (hospds=1 AND mortds=1) dataset=3.
EXECUTE.
DELETE VARIABLES hospds mortds.
FORMATS dataset (f1).
VARIABLE LABELS  dataset 'Dataset case obtained from'.
VALUE LABELS dataset 1 hospital
2 mortality
3 both hospital and mortality.
EXECUTE.

RECODE gender_m ('M'=1) ('F'=2) INTO gender_m1.
RECODE gender_h ('M'=1) ('F'=2) INTO gender_h1.
FORMATS gender_m1 (F1).
FORMATS gender_h1 (F1).
DELETE VARIABLES gender_m gender_h.
RENAME VARIABLES gender_m1=gender_m gender_h1=gender_h.
VALUE LABELS gender_m gender_h 1 male
2 female.
EXECUTE.

IF (morbseq=1) #flag=0.
IF (#flag=1) regyear_m=LAG (regyear_m).
IF (#flag=1) dob_m=LAG (dob_m).
IF (#flag=1) gender_m=LAG (gender_m).
```

```
IF (#flag=1) dod_m=LAG (dod_m).
IF (#flag=1) scale_m=LAG (scale_m).
IF (#flag=1) icdgroup_b_m=LAG (icdgroup_b_m).
IF (#flag=1) dataset=LAG (dataset).
DO IF (morbseq>1 AND #flag=0).
COMPUTE regyear_m=LAG (regyear_m).
COMPUTE dob_m=LAG (dob_m).
COMPUTE gender_m=LAG (gender_m).
COMPUTE dod_m=LAG (dod_m).
COMPUTE scale_m=LAG (scale_m).
COMPUTE icdgroup_b_m=LAG (icdgroup_b_m).
COMPUTE dataset=LAG (dataset).
COMPUTE #flag=1.
END IF.
EXECUTE.

SAVE outfile 'filepath\merged_all.sav'.
DATASET NAME mergedall.
```

The dataset is now reduced to include only individuals who have a mortality record *and* a linked hospital record. This file is saved as merged_mort_cut.sav. This is the area MH in the conceptual diagram on page 158, and (after reincorporating mortality events with no linked hospital event) is later used for the reference point modelling.

```
SELECT IF dataset=3.
EXECUTE.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

SAVE outfile 'filepath\merged_mort_cut.sav'
/KEEP enc_nhi event_id_h dob_date_h end_type_h gender_h event_end_date_h event_start_date_h
dob_m dod_m gender_m morbseq fileseq year_h regyear_m dataset.
DATASET NAME mergedmortcut.
DATASET ACTIVATE mergedmortcut.
DATASET CLOSE mergedall.
```

Step 2 – Probabilistic Matching Process

Now the records have been deterministically linked by encrypted NHI, the dataset is examined for false matches: where two different individuals happen to have the same encrypted NHI. An individual identification number such as an encrypted NHI is the holy grail of individual record linkage for health research; however as will be apparent (and as was demonstrated by the presence of duplicate encrypted NHI numbers in the mortality dataset) even this is not perfect. The probabilistic matching process is demonstrated for the merged dataset including only linked hospital and mortality records (MH)

(merged_mort_cut.sav), the dataset later used for reference point modelling, after incorporating back in mortality events with no linked hospital events (M').

The match is tested on three criteria: date of birth (on multiple hospital records for an individual and/or between hospital and mortality records); a hospital discharge date after the date of death for that individual; and gender (again on multiple hospital records for an individual and/or between hospital and mortality records).

Date of birth is commonly recorded incorrectly on hospital records, especially for older people. A variable (dobdiff) is computed to show any difference between the date of birth between the hospital and mortality records. This is then converted into a 'block' variable (dobdiffblock) that indicates whether the individual has a mismatch between hospital and mortality dates of birth for any of their records (not that record specifically).

```
COMPUTE  dobdiff=DATEDIF (dob_date_h, dob_m, "days").
VARIABLE LABEL  dobdiff "Difference between date of birth on hospital and mortality record, in
days".
FORMATS  dobdiff (F1.0).
EXECUTE.

COMPUTE dobdiffblock=0.
IF (morbseq=1) #flag=0.
IF (#flag=1) dobdiffblock=1.
DO IF (dobdiff NE 0 AND #flag=0).
COMPUTE dobdiffblock=1.
COMPUTE #flag=1.
END IF.
EXECUTE.
SORT CASES by fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=1) dobdiffblock=1.
DO IF (dobdiff NE 0 AND #flag=0).
COMPUTE dobdiffblock=1.
COMPUTE #flag=1.
END IF.
VARIABLE LABEL dobdiffblock "Individual indicator of incorrect date of birth - 1 if the
individual has at least one hospital record with an incorrect date of birth, 0 otherwise".
FORMATS dobdiffblock (F1.0).
EXECUTE.
SORT CASES by fileseq (A).
```

A similar difference (dddiff) and 'block' variable (dddiffblock) is created for the second criteria: hospital discharges occurring after the date of death. It is assumed that an individual being discharged from hospital several months after their date of death was not, in fact, dead at the point when they were discharged from hospital. This suggests that either the same encrypted NHI has been applied to two separate individuals, or there is an error in the data.

```
COMPUTE  dddiff=DATEDIF (dod_m, event_end_date_h, "days").
VARIABLE LABEL  dddiff "Difference between date of death and the date of discharge from
hospital - if negative, hospital discharge date occured after date of death".
FORMATS  dddiff (F1.0).
EXECUTE.

COMPUTE dddiffblock=0.
IF (morbseq=1) #flag=0.
IF (#flag=1) dddiffblock=1.
DO IF (dddiff < 0 AND #flag=0).
COMPUTE dddiffblock=1.
COMPUTE #flag=1.
END IF.
EXECUTE.
SORT CASES by fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=1) dddiffblock=1.
DO IF (dddiff < 0 AND #flag=0).
COMPUTE dddiffblock=1.
COMPUTE #flag=1.
END IF.
VARIABLE LABEL dddiffblock "Individual indicator of hospital discharge after death - 1 if the
individual has at least one hospital discharge date after the date of death, 0 otherwise".
FORMATS dddiffblock (F1.0).
EXECUTE.
SORT CASES by fileseq (A).
```

And finally a difference (sexdiff) and ‘block’ (sexdiffblock) variable is created for sex mismatches.

```
COMPUTE sexdiff=0.
IF (gender_h NE gender_m) sexdiff=1.
VARIABLE LABEL sexdiff "Record indicator of gender mismatch in the hospital and mortality
records - 1 if genders do not match, 0 otherwise".
FORMATS sexdiff (F1.0).
EXECUTE.
FREQUENCIES sexdiff.
EXECUTE.
SORT CASES by sexdiff (D).
EXECUTE.

COMPUTE sexdiffblock=0.
IF (morbseq=1) #flag=0.
IF (#flag=1) sexdiffblock=1.
DO IF (sexdiff NE 0 AND #flag=0).
COMPUTE sexdiffblock=1.
COMPUTE #flag=1.
END IF.
EXECUTE.
SORT CASES by fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=1) sexdiffblock=1.
DO IF (sexdiff NE 0 AND #flag=0).
COMPUTE sexdiffblock=1.
COMPUTE #flag=1.
END IF.
```

```
VARIABLE LABEL sexdiffblock "Individual indicator of incorrect gender - 1 if the individual has
at least one hospital record with an incorrect gender, 0 otherwise".
FORMATS sexdiffblock (F1.0).
EXECUTE.
SORT CASES by fileseq (A).
```

Table B.1 presents the number of individuals who would be excluded from the dataset based on the criteria above, a total of 14.4 percent. The majority of the exclusions are due to an incorrect date of birth (13.1 percent).

Table B.1: Percent of Individuals Excluded from Merged* Dataset Before Probabilistic Matching

		Number of individuals		
	Total number of individuals	**417,160**	%	
Number of individuals who have at least one hospital record with:	Incorrect date of birth	54,547	13.1	
	discharge date after date of death	4,992	1.2	
	Incorrect gender	1,749	0.4	
	Number not excluded	**357,038**	85.6	
			14.4	% excluded

*M' and MH (all mortality records and hospital records only where mortality record is present)

In all cases where a mismatch is found, the mortality record is assumed to be correct and the hospital record incorrect. Unlike hospital records, the death certificate (from which date of birth, date of death and gender values in the mortality data are derived) is a legal document and (although not infallible) more likely to be reliable. The process of probabilistically matching the excluded individuals is detailed below and in Table B.2. After examining patterns in incorrect dates of birth it was found that the vast majority of mismatches were less than or equal to 31 days (one month) different. Differences of about one/two/or three years were also common. It was decided that if an individual was discharged from hospital dead (with an incorrect date of birth on the hospital record) but that the date of discharge agreed with the date of death on the mortality record, they were likely to be the same person. Also, if this incorrect date of birth had been identified as an error (rather than a different person) through comparison with the date discharged dead and the date of death, then other hospital records for that individual using the *same* incorrect date of birth were allowed.

Hospital discharges occurring one or two days after the date of death are quite common, and not unfeasible. People may die overnight in hospital and be discharged the following day, or die over the weekend and be held until the following week. After examination of the pattern of the difference between the date discharged dead and the date

of death, a reasonable period between the date discharged dead and the date of death was identified as 10 days.

Table B.2: Probabilistic Matching Criteria

Incorrect date of birth	Allow a mismatch of up to 31 days
	Allow mismatches that are about a year or about two years out etc., up to 20 years, allowing two days either side
	Allow an incorrect hospital date of birth if that stay ends in death on discharge, and the discharge date is up to 10 days after the date of death
	Allow if a subsequent discharge for that individual with **the same** incorrect hospital date of birth is identified as ok (dead on discharge, within 10 days after date of death
Discharge date after date of death	Allow discharges between 1 and 10 days after death, as long as the discharge type is "dead on discharge"
Incorrect sex	No adjustments made

With sex mismatches, there was little option but to assume that sex is fixed and exclude any individual with different sexes recorded on different records. Changes in sex are very rare and there is no way to identify if the mismatch is a recording error, a real change of sex, or a different person. Deterministic exclusion is therefore used for sex mismatches.

Applying these criteria, the percentage of individuals excluded from the dataset is reduced from 14.4 percent to 2.5 percent (see Table B.3).

Table B.3: Percent of Individuals Excluded from the Merged* Dataset After Probabilistic Matching

		Number of individuals		
	Total number of individuals	**417,160**	%	
Number of individuals who have at least one hospital record with:	Incorrect date of birth	8,002	1.9	
	discharge date after date of death	1,024	0.2	
	Incorrect gender	1,749	0.4	
	Number not excluded	**406,711**	97.5	
			2.5	% excluded

*M' and MH (all mortality records and hospital records only where mortality record is present)

The syntax used to affect the probabilistic matching criteria is given below. For incorrect date of birth, records with a mismatch of up to 31 days, and mismatches that are about one/two/three years out etc. are identified by the creation of the variable dobcategory.

```
IF (dobdiff=0) dobcategory=0.
IF (dobdiff>=-31 AND dobdiff<=31 AND dobdiff NE 0) dobcategory=0.5.
IF ( (dobdiff<=-363 AND dobdiff>=-367) OR (dobdiff>=363 AND dobdiff<=367)) dobcategory=1.
IF ( (dobdiff<=-728 AND dobdiff>=-732) OR (dobdiff>=728 AND dobdiff<=732)) dobcategory=2.
IF ( (dobdiff<=-1093 AND dobdiff>=-1097) OR (dobdiff>=1093 AND dobdiff<=1097))
dobcategory=3.
IF ( (dobdiff<=-1458 AND dobdiff>=-1462) OR (dobdiff>=1458 AND dobdiff<=1462))
dobcategory=4.
IF ( (dobdiff<=-1823 AND dobdiff>=-1827) OR (dobdiff>=1823 AND dobdiff<=1827))
dobcategory=5.
IF ( (dobdiff<=-2188 AND dobdiff>=-2192) OR (dobdiff>=2188 AND dobdiff<=2192))
dobcategory=6.
IF ( (dobdiff<=-2553 AND dobdiff>=-2557) OR (dobdiff>=2553 AND dobdiff<=2557))
dobcategory=7.
IF ( (dobdiff<=-2918 AND dobdiff>=-2922) OR (dobdiff>=2918 AND dobdiff<=2922))
dobcategory=8.
IF ( (dobdiff<=-3283 AND dobdiff>=-3287) OR (dobdiff>=3283 AND dobdiff<=3287))
dobcategory=9.
IF ( (dobdiff<=-3648 AND dobdiff>=-3652) OR (dobdiff>=3648 AND dobdiff<=3652))
dobcategory=10.
IF ( (dobdiff<=-4013 AND dobdiff>=-4017) OR (dobdiff>=4013 AND dobdiff<=4017))
dobcategory=11.
IF ( (dobdiff<=-4378 AND dobdiff>=-4382) OR (dobdiff>=4378 AND dobdiff<=4382))
dobcategory=12.
IF ( (dobdiff<=-4743 AND dobdiff>=-4747) OR (dobdiff>=4743 AND dobdiff<=4747))
dobcategory=13.
IF ( (dobdiff<=-5108 AND dobdiff>=-5112) OR (dobdiff>=5108 AND dobdiff<=5112))
dobcategory=14.
IF ( (dobdiff<=-5473 AND dobdiff>=-5477) OR (dobdiff>=5473 AND dobdiff<=5477))
dobcategory=15.
IF ( (dobdiff<=-5838 AND dobdiff>=-5842) OR (dobdiff>=5838 AND dobdiff<=5842))
dobcategory=16.
IF ( (dobdiff<=-6203 AND dobdiff>=-6207) OR (dobdiff>=6203 AND dobdiff<=6207))
dobcategory=17.
IF ( (dobdiff<=-6568 AND dobdiff>=-6572) OR (dobdiff>=6568 AND dobdiff<=6572))
dobcategory=18.
IF ( (dobdiff<=-6933 AND dobdiff>=-6937) OR (dobdiff>=6933 AND dobdiff<=6937))
dobcategory=19.
IF ( (dobdiff<=-7298 AND dobdiff>=-7302) OR (dobdiff>=7298 AND dobdiff<=7302))
dobcategory=20.
VARIABLE LABEL dobcategory "Specific errors in hospital date of birth that may occur by
accident. 0 if no mismatch between hospital and mortality date of birth".
FORMATS dobcategory (F10.1).
EXECUTE.
RECODE dobcategory (MISSING=-9).
EXECUTE.
```

Hospital records with an incorrect date of birth, but that end in death on discharge up to 10 days after the date of death from the mortality record are identified next.

```
IF (dobcategory=-9 AND end_type_h="DD" AND dddiff>=-10 AND dddiff<=0) dobcategory=99.
EXECUTE.
```

This is then extended to any other hospital record for that individual with the same, incorrect date of birth.

```
SORT CASES fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=dob_date_h) deadondis=1.
DO IF (dobcategory=99 AND #flag=0).
COMPUTE #flag=dob_date_h.
COMPUTE deadondis=1.
END IF.
SORT CASES by fileseq (A).
EXECUTE.
IF (deadondis=1 AND dobcategory=-9) dobcategory=90.
EXECUTE.
```

All incorrect date of birth probabilistic matching criteria have been met. A ‘block’ variable (dobdiffblock2) is created which is then used to exclude individuals who still have mismatches in date of birth after probabilistic matching. Redundant variables are deleted.

```
COMPUTE dobdiffblock2=0.
IF (morbseq=1) #flag=0.
IF (#flag=1) dobdiffblock2=1.
DO IF (dobcategory=-9 AND #flag=0).
COMPUTE dobdiffblock2=1.
COMPUTE #flag=1.
END IF.
EXECUTE.
SORT CASES by fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=1) dobdiffblock2=1.
DO IF (dobcategory=-9 AND #flag=0).
COMPUTE dobdiffblock2=1.
COMPUTE #flag=1.
END IF.
VARIABLE LABEL dobdiffblock2 "Adjusted individual indicator of incorrect date of birth - 1 if
individual has at least one record with unexplained difference in date of birth between mortality and
hospital records, 0 otherwise".
FORMATS dobdiffblock2 (F1.0).
EXECUTE.
SORT CASES by fileseq (A).

SELECT IF dobdiffblock2=0.
EXECUTE.
DELETE VARIABLES deadondis dobdiff dobdiffblock dobcategory.
```

The probabilistic matching process for hospital discharges after the date of death is more straightforward. A ‘block’ variable is created to identify individuals with a hospital discharge up to10 days after death, where the discharge type was ‘dead on discharge’ (dddiffblock2). This block variable is then used to exclude individuals with a discharge date after the date of death, except where the discharge is up to 10 days after the date of death and the discharge type is ‘dead on discharge’.

```
COMPUTE dddiffflag=0.
IF (dddiff<-10) dddiffflag=1.
```

```
IF (dddiff>=-10 AND dddiff<0 AND end_type_h NE "DD") dddiffflag=1.
EXECUTE.

COMPUTE dddiffblock2=0.
IF (morbseq=1) #flag=0.
IF (#flag=1) dddiffblock2=1.
DO IF (dddiffflag=1 AND #flag=0).
COMPUTE dddiffblock2=1.
COMPUTE #flag=1.
END IF.
EXECUTE.
SORT CASES by fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=1) dddiffblock2=1.
DO IF (dddiffflag=1 AND #flag=0).
COMPUTE dddiffblock2=1.
COMPUTE #flag=1.
END IF.
VARIABLE LABEL dddiffblock2 "Adjusted Individual indicator of hospital discharge after death -
1 if the individual has at least one hospital discharge date MORE THAN TEN DAYS after date of
death or not dead on discharge, 0 otherwise".
FORMATS dddiffblock2 (F1.0).
EXECUTE.
SORT CASES by fileseq (A).

SELECT IF dddiffblock2=0.
EXECUTE.
DELETE VARIABLES dddiffblock dddiffflag dddiffblock2.
EXECUTE.
```

No adjustments are made for sex mismatches, all that remains is to delete individuals with a sex mismatch, compute a new file sequence variable, and save the dataset (merged_mort_cut6.sav).

```
DELETE VARIABLES sexdiffblock.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

SAVE outfile 'filepath\merged_mort_cut6.sav'.
DATASET NAME mergedmortcut6.
```

Step 3 – National and Area Deprivation Filters

Up until this point, the unfiltered datasets have been used. At this stage it is necessary to apply the national and deprivation filters. Once the filters have been applied, all manipulation needs to be conducted separately for the national and the deprivation datasets.

The hospital dataset that includes the filter variables (hosp_all_filter.sav) is opened, and the filter variables merged with the merged mortality and hospital dataset (merged_mort_cut6.sav). This dataset (and associated filter variables) was created when preparing the datasets for the analyses in Papers One and Two (see Appendix A).

```
Get file 'filepath\hosp_all_filter.sav'
/KEEP event_id_h filter1 filter2 filter3 filter4 filter5 filter6 filter7 filter8 filter9 filter10 filter11
filter12 filter13 filter14 filter15 filter16 dataset.
DATASET NAME hospallfilter.
DATASET ACTIVATE hospallfilter.

SELECT IF dataset=1 OR dataset=3.

SORT CASES BY event_id_h.

DATASET ACTIVATE mergedmortcut6.
SORT CASES BY event_id_h.
MATCH FILES /FILE=*
  /TABLE='mergedall'
  /BY event_id_h.
EXECUTE.
DATASET CLOSE hospallfilter.

GET FILE 'filepath\merged_all.sav'
/KEEP enc_nhi dob_m dod_m gender_m regyear_m icdgroup_b_m scale_m dataset.
DATASET NAME mergedall.
DATASET ACTIVATE mergedall.
SELECT IF (dataset=2).
EXECUTE.

DATASET ACTIVATE mergedmortcut6.
ADD FILES /FILE=*
  /FILE='mergedall'.
EXECUTE.
DATASET CLOSE mergedall.

SORT CASES BY enc_nhi (A), event_start_date_h (A), event_end_date_h (A).

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.
```

After merging in the filter variables, and adding in the unlinked mortality records the dataset is saved under a new name, merged_mort_filter1.sav. First the national filter is applied, the unselected records deleted, new file and morbidity sequence variables created, and the file saved as merged_mort_nat1.sav. Merged_mort_filter1.sav is then reopened and the process repeated to apply the deprivation dataset, and the file is saved as merged_mort_dep1.sav.

```
SAVE outfile 'filepath\merged_mort_filter1.sav'.
DATASET NAME mergedmortfilter1.
```

```
SELECT IF ( (dataset=2) OR (filter1=0 and filter2=0 and filter3=0 and filter4=0 and filter5=0 and
filter11=0 and filter12=0 and filter16=0 and icdgroup_b_m>0)) natfilter=0.
EXECUTE.

DELETE VARIABLES filter1 filter2 filter3 filter4 filter5 filter6 filter7 filter8 filter9 filter10 filter11
filter12 filter13 filter14 filter15 filter16.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

IF (fileseq=1 OR enc_nhi NE LAG (enc_nhi)) morbseq=1.
IF (fileseq NE 1 AND enc_nhi=LAG (enc_nhi)) morbseq=LAG (morbseq)+1.
VARIABLE LABEL morbseq "Morbidity sequence".
FORMATS morbseq (F1).
EXECUTE.

SAVE outfile 'filepath\merged_mort_nat1.sav'.
DATASET NAME mergedmortnat1.

Get file 'filepath\merged_mort_filter1.sav'.
DATASET NAME mergedmortfilter1.
DATASET CLOSE mergedmortnat1.

SELECT IF year_h>1990.

SELECT IF (filter1=0 and filter2=0 and filter3=0 and filter4=0 and filter5=0 and filter11=0 and
filter13=0 and filter14=0 and filter15=0 and filter16=0 and icdgroup_b_m>0 and  scale_m>0).
EXECUTE.

DELETE VARIABLES filter1 filter2 filter3 filter4 filter5 filter6 filter7 filter8 filter9 filter10 filter11
filter12 filter13 filter14 filter15 filter16.

COMPUTE fileseq=$casenum.
VARIABLE LABEL fileseq "File sequence".
FORMATS fileseq (F1).
EXECUTE.

IF (fileseq=1 OR enc_nhi NE LAG (enc_nhi)) morbseq=1.
IF (fileseq NE 1 AND enc_nhi=LAG (enc_nhi)) morbseq=LAG (morbseq)+1.
VARIABLE LABEL morbseq "Morbidity sequence".
FORMATS morbseq (F1).
EXECUTE.

SAVE outfile 'filepath\merged_mort_dep1.sav'.
DATASET NAME mergedmortdep1.
```

Step 4 – Duplicate and Overlapping Hospital Stays

An individual can only spend a maximum of one day in hospital per calendar day. Duplicate and overlapping hospital stays can artificially inflate the number of bed days spent in hospital in a given time period. First, duplicates are deleted. These are hospital stays with the same start and end date. In the national dataset (merged_mort_nat1.sav)

there were 538 duplicates. When identifying duplicates it is important that the file is sorted by ascending enc_nhi, discharge date, and then admission date (the criteria for the file sequence variable).

```
COMPUTE duplicate=0.
IF (enc_nhi = LAG (enc_nhi) AND event_end_date_h =LAG (event_end_date_h) AND
event_start_date_h=LAG (event_start_date_h)) duplicate=1.
EXECUTE.

SELECT IF duplicate=0.
EXECUTE.
DELETE VARIABLES duplicate.

COMPUTE fileseq=$casenum.
EXECUTE.

COMPUTE morbseq=0.
IF (fileseq=1 OR enc_nhi NE LAG (enc_nhi)) morbseq=1.
IF (fileseq NE 1 AND enc_nhi=LAG (enc_nhi)) morbseq=LAG (morbseq)+1.
EXECUTE.
```

Overlapping hospital stays are dealt with slightly differently. A transfer sequence variable (overlapseq) is created to identify which hospital stays for an individual overlap. Where overlapseq is 1, the hospital event is first in a transfer sequence. Where overlapseq is 2, the hospital event is second in a transfer sequence and overlaps with the previous hospital event, and so on. In absence of overlapping hospital stays, overlapseq will always be 1. The discharge date for the first hospital event in an overlap sequence is adjusted to be the same as the discharge date for the *last* hospital event in that overlap sequence, and all but the first event in the overlap sequence deleted. Thus no days in hospital are counted twice. If analysis was being conducted by cause of hospital bed days, this would not be a suitable solution as hospital bed days by cause may be over (or under) represented. However the analyses conducted require the reduction of the hospital data to one line per individual containing simply length of stay information, so any event-specific hospital information (such as ICD codes) will be lost anyway.

```
COMPUTE overlapseq=0.
IF (morbseq>=2 AND LAG (event_end_date_h)>event_start_date_h) overlapseq=1.
IF (morbseq>=2 AND overlapseq=1 AND LAG (overlapseq)>=1) overlapseq=LAG (overlapseq)+1.
FORMATS overlapseq (F1.0).
VARIABLE LABEL overlapseq "Transfer sequence: greater than one if one or more overlapping
hospital stays".
EXECUTE.
```

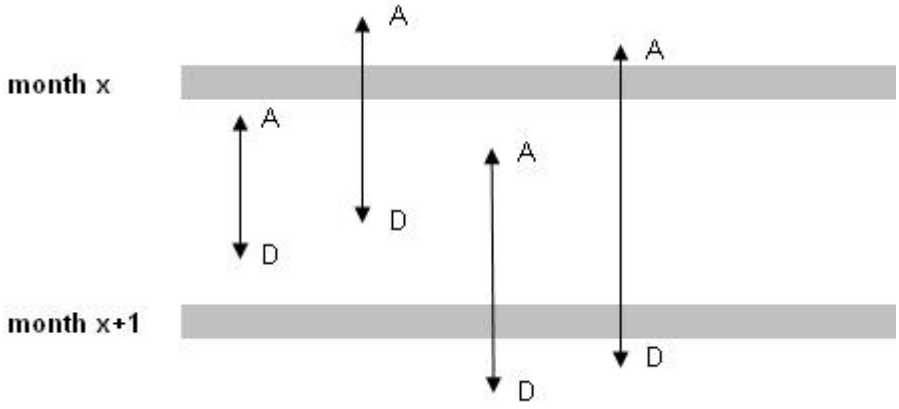
For the national dataset (merged_mort_nat1) there were 827 hospital events which overlapped the previous event (807 overlapseq=2 and 20 overlapseq=3). The discharge date for the first event in the overlap sequence is adjusted to be the same as for the last event in the overlap sequence (new variable enddate_h), and all but the first event in the sequence are deleted. The file sequence and morbidity sequence variables are re-created and the dataset saved as merged_mort_nat3.sav. This process is repeated for the deprivation dataset.

```
SORT CASES BY fileseq (D).
COMPUTE enddate_h=event_end_date_h.
IF (LAG (overlapseq)>0) enddate_h=LAG (enddate_h).
FORMATS enddate_h (EDATE10).
VARIABLE LABEL enddate_h "Event end date adjusted for overlapping hospital stays".
EXECUTE.

SELECT IF overlapseq=0.
EXECUTE.
COMPUTE fileseq=$casenum.
EXECUTE.

COMPUTE morbseq=0.
IF (fileseq=1 OR enc_nhi NE LAG (enc_nhi)) morbseq=1.
IF (fileseq NE 1 AND enc_nhi=LAG (enc_nhi)) morbseq=LAG (morbseq)+1.
EXECUTE.

DELETE VARIABLES overlapseq, event_end_date_h.
SAVE outfile 'filepath\merged_mort_nat3.sav'.
DATASET NAME mergedmortnat3.
```

Step 5 – Length of Stay variables for Months Prior to Death

Month-specific length of stay variables are created for each separate hospital event, before being aggregated and the datasets reduced to one line per individual. The date of each month prior to death is calculated first (dodm1 to dodm24. Do repeat commands are used to calculate the date of the 2nd to the 24th months prior to death (dodm1 to dodm24). These dates are then used to calculate the number of days spent in hospital in each month, for each hospital event (also using do repeat commands) (losm1 to losm24). Do repeat commands require the variable names to be the same, with sequential numerical endings. As the date of death (dod_m) marks the end of the first month prior to death, the length of stay for the first month prior to death needs to be calculated separately. There are four different possibilities, shown in Figure A.1.

Figure B.1: Length of Stay in Months Prior to Death, Four Possibilities



Notes:
A represents the date of admission
D represents the date of discharge
Month x represents the date of a given month prior to death
Month x+1 represents the date of month x plus 28 days

An individual can be admitted and discharged within the same month prior to death, admitted in the previous month and discharged the current month, admitted in the current month and discharged in the next month, and admitted in the previous month and discharged in the next month. The four 'IF' commands shown in the following syntax represent these four possibilities. If an individual is in hospital for an entire month (the fourth possibility), the length of stay is 28 days. A stay of 0 (admitted and discharged on the same day) is recoded as 0.5, a stay of half a day. Missing values (where there are no hospital bed days for a given month) are coded as 0. The redundant date variables are deleted when the length of stay variables have been created and the file (with length of stay variables) is saved as hosp_nat_merged2.sav. The same process is repeated for the deprivation dataset.

```
COMPUTE #counter=0.
FORMATS #counter (F10.0).
DO REPEAT dod=dodm1 to dodm24.
COMPUTE #counter= (#counter+28).
COMPUTE dod=DATESUM (dod_m,- (#counter),"days",'closest').
FORMATS dod (EDATE10).
END REPEAT PRINT.
EXECUTE.

IF (enddate_h<=dod_m AND enddate_h>dodm1 AND event_start_date_h<=dod_m AND
event_start_date_h>dodm1) losm1=DATEDIF (enddate_h, event_start_date_h, "days").
IF (enddate_h<=dod_m AND enddate_h>dodm1 AND event_start_date_h<=dodm1)
losm1=DATEDIF (enddate_h, dodm1, "days").
IF (enddate_h>dod_m AND event_start_date_h<=dod_m AND event_start_date_h>dodm1)
losm1=DATEDIF (dod_m, event_start_date_h, "days").
IF (enddate_h>dod_m AND event_start_date_h<=dodm1) losm1=28.
RECODE losm1 (0=0.5).
```

```
FORMATS losm1 (F8.1).

DO REPEAT los=losm2 to losm24
/dod=dodm1 to dodm23
/dod1=dodm2 to dodm24.
IF (enddate_h<=dod AND enddate_h>dod1 AND event_start_date_h<=dod AND
event_start_date_h>dod1) los=DATEDIF (enddate_h, event_start_date_h, "days").
IF (enddate_h<=dod AND enddate_h>dod1 AND event_start_date_h<=dod1) los=DATEDIF
(enddate_h, dod1, "days").
IF (enddate_h>dod AND event_start_date_h<=dod AND event_start_date_h>dod1) los=DATEDIF
(dod, event_start_date_h, "days").
IF (enddate_h>dod AND event_start_date_h<=dod1) los=28.
RECODE los (0=0.5).
FORMATS los (F8.1).
END REPEAT.
EXECUTE.

DO REPEAT los=losm1 TO losm24.
RECODE los (SYSMIS=0) (ELSE=COPY) INTO los.
END REPEAT.
EXECUTE.

DELETE VARIABLES dodw1 dodw2 dodw3 dodw4 dodm1 dodm2 dodm3 dodm4 dodm5 dodm6
dodm7 dodm8 dodm9 dodm10 dodm11 dodm12 dodm13 dodm14 dodm15 dodm16 dodm17
dodm18 dodm19 dodm20 dodm21 dodm22 dodm23 dodm24.

SAVE OUTFILE  "filepath\hosp_nat_merged2.sav".
DATASET NAME hospnatmerged2.
```

Step 6 – Aggregate Length of Stay Variables and Reduce to One Record per Individual

The length of stay variables in the months prior to death for each hospital event, are added up in preparation for reducing the data to one record per individual. These are referred to as *total* length of stay variables (totlos1 to totlos24). The cumulative hospital-event specific length of stay is calculated down the file, thus the length of stay variables in the last record for an individual will be the sum of all their hospital stays in each month prior to death. This is then copied *up* the file by turning the file ‘upside down’ (sorting by descending file sequence).

```
DO REPEAT totlos=totlosm1 TO totlosm24
/ los=losm1 TO losm24
/ #counter=#counter1 TO #counter24.
COMPUTE totlos=0.
RECODE los (sysmiss=0).
DO IF (morbseq=1).
COMPUTE #counter=0.
COMPUTE #flag=0.
```

```
END IF.
DO IF (#flag=1).
COMPUTE #counter= (#counter+los).
COMPUTE totlos=#counter.
END IF.
DO IF (los>0 AND #counter=0).
COMPUTE #counter=los.
COMPUTE totlos=#counter.
COMPUTE #flag=1.
END IF.
FORMATS totlos (F8.1).
END REPEAT.
EXECUTE.

IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
DO REPEAT totlos=totlosm1 TO totlosm24.
IF (#flag=1) totlos=LAG (totlos).
DO IF (#flag=0 AND enc_nhi=LAG (enc_nhi)).
COMPUTE totlos=LAG (totlos).
COMPUTE #flag=1.
END IF.
END REPEAT.
EXECUTE.
SORT CASES by fileseq (A).
```

If an individual is discharged and then readmitted on the same day, it can push the length of stay for that month over 28 to 28.5. This is not a problem in its own right, but to make sure that no days are counted more than once, the total length of stay variables for each month need to be recoded so they do not exceed 28.

```
DO REPEAT totlos=totlosm1 TO totlosm24.
IF (totlos=28.5) totlos=28.
END REPEAT.
EXECUTE.
```

Finally, the redundant event-specific length of stay variables can be deleted and the dataset reduced to one record per individual. The redundant morbidity sequence and any hospital event specific variables can be deleted at this point. The file is saved as merged_nat_indiv.sav. This process is repeated for the deprivation dataset.

```
DELETE VARIABLES losw1 losw2 losw3 losw4 losm1 losm2 losm3 losm4 losm5 losm6 losm7
losm8 losm9 losm10 losm11 losm12 losm13 losm14 losm15 losm16 losm17 losm18 losm19 losm20
losm21 losm22 losm23 losm24.

SELECT IF morbseq=1.
EXECUTE.

DELETE VARIABLES event_id_h event_start_date_h enddate_h morbseq.

COMPUTE fileseq=$casenum.
EXECUTE.

SAVE outfile 'filepath\merged_nat_indiv.sav'.
```

```
DATASET NAME mergednatindiv.
```

Step 7 – Incorporate Linked Dataset back into Mortality Dataset (Reference Point Modelling)

Now the probabilistic matching, the identification and resolving of duplicate and overlapping hospital stays, the creation of month-specific length of stay variables, and the reduction of the data to one line per individual have been carried out, the linked dataset needs to be merged back into the mortality dataset (mort_nhi_valid), and saved as a new dataset (merged_nat_indiv2.sav). This provides us with the denominator for the reference point modelling, which needs to include *all* deaths, regardless of if the individual spent time in hospital in the 24 months prior to death. As before, this needs to be repeated for the deprivation dataset.

```
Get file 'filepath\mort_nhi_valid.sav'
/KEEP enc_nhi, dod_m, gender_m, scale_m, regyear_m, icdgroup_b_m,
age_grouped_m.
DATASET NAME mort_nhi_valid.

DATASET ACTIVATE mortnhivalidcut.
SORT CASES by enc_nhi (A).
MATCH FILES /FILE=*
  /TABLE='mergednatindiv'
  /BY enc_nhi.
EXECUTE.

SAVE outfile  'filepath\merged_nat_indiv2.sav'.
DATASET NAME hospnatfake1.
```

Step 8 –Baseline Hospital Use (Reference Point Modelling)

The reference point modelling requires the calculation of baseline hospital bed days in the months prior to a ‘fake’ death. This provides a comparison between hospital use in the months prior to death, and hospital use that would be expected *in the absence of proximity to death*. This is calculated by taking the distribution of age at death (in days) for individuals who died over the time period (by sex and deprivation decile) and applying this to all available data (both the hospital and the mortality datasets) to impose a ‘fake’ date of death for each individual. The hospital and mortality datasets together are the closest thing

available to an approximation of the general population of New Zealand over this time period. Once checking that the individual is not already known to have died by their 'fake' date of death, bed days in the months prior to this 'fake' death date are calculated as in Steps 5 and 6. This process is demonstrated for the national analysis, for which distribution of age at death is obtained from, and applied to males and females separately. For the deprivation analysis the distribution of age at death is obtained from, and applied to 20 different groups separately (males, females, and deprivation deciles 1 to 10).

First, the mortality dataset containing all mortality events with a valid NHI number (M) is opened (merged_nat_allmort.sav). We are only interested in the distribution of age at death from 1990 onwards (with a 2 year observation window prior to death and records from 1988, only deaths from age 1990 are included in the analysis). This is 1993 onwards for the deprivation analysis.

```
GET FILE  'filepath\mort_nhi_valid.sav'
/KEEP enc_nhi dod_m gender_m scale_m regyear_m icdgroup_b_m age_grouped_m dob_m.
DATASET NAME mortnhivalid.

SELECT IF regyear_m>=1990.
EXECUTE.
```

The age at death in days (agedays) is calculated for each individual and then the dataset is split into different files for males and females (mortvalidmale, mortvalidfemale). For the deprivation analysis the dataset is split into different files for males, females, and deprivation deciles.

```
COMPUTE  agedays=DATEDIF (dod_m, dob_m, "days").
VARIABLE LABEL  agedays "Age at death in days".
VARIABLE LEVEL  agedays (SCALE).
FORMATS  agedays (F5.0).
VARIABLE WIDTH  agedays (5).
EXECUTE.

DATASET ACTIVATE mortnhivalid.
DATASET COPY  mortvalidfemale.
DATASET ACTIVATE mortvalidfemale.
FILTER OFF.
USE ALL.
SELECT IF (gender_m=1).
EXECUTE.

DATASET ACTIVATE mortnhivalid.
DATASET COPY  mortvalidmale.
DATASET ACTIVATE mortvalidmale.
FILTER OFF.
USE ALL.
SELECT IF (gender_m=2).
EXECUTE.
```

```
DATASET CLOSE mortnhivalid.
```

A random number generator is used to calculate a random number large enough to be unique for each individual (randsort). The dataset is then sorted by this random number, and a file sequence variable for this randomly sorted file (randseq) is created.

```
DATASET ACTIVATE mortvalidfemale.
SORT CASES BY enc_nhi (A).
SET SEED 2000000.
COMPUTE randsort=UNIFORM (1).
FORMATS randsort (F8.7).
EXECUTE.
SORT CASES BY randsort (A).
EXECUTE.
COMPUTE randseq=$casenum.
VARIABLE LABEL randseq "Random sequence".
FORMATS randseq (F1).
EXECUTE.

DATASET ACTIVATE mortvalidmale.
SORT CASES BY enc_nhi (A).
SET SEED 2000000.
COMPUTE randsort=UNIFORM (1).
FORMATS randsort (F8.7).
EXECUTE.
SORT CASES BY randsort (A).
EXECUTE.
COMPUTE randseq=$casenum.
VARIABLE LABEL randseq "Random sequence".
FORMATS randseq (F1).
EXECUTE.
```

The hospital dataset is opened, reduced to include individuals with a valid enc_nhi and filters applied, and saved as hosp_nat_mort.sav. This dataset is then split into separate files for females and males (hospnatfemale and hospnatmale) and a random number is generated (randsort). This random number is set to repeat after the number of records in the associated mortality dataset (natallfemale or natallmale). In this case the associated female mortality dataset had 214,608 records and the male mortality dataset had 225,805 records. The dataset is sorted by the random number, and a file sequence variable created (randseq). The age at death in days variable (agedays) from the associated mortality dataset (natallfemale or natallmale) is then merged into the dataset, using the random number (randsort) as a key. Once this has been conducted for both sexes, they are merged together and saved as a new dataset, hosp_nat_fake1.sav.

```
DATASET ACTIVATE hospnat.
DATASET COPY  hospnatfemale.
DATASET ACTIVATE  hospnatfemale.
```

```
SELECT IF (gender_h=1).
EXECUTE.

SET SEED=2000000.
COMPUTE randsort=UNIFORM (1).
FORMATS randsort (F8.7).
EXECUTE.
SORT CASES BY randsort (A).
EXECUTE.
IF ($casenum EQ 1) #counter=0.
IF (#counter EQ 214608) #counter=0.
COMPUTE #counter=#counter + 1.
COMPUTE randseq=#counter.
FORMATS randseq (F8).
VARIABLE LABELS randseq Random sequence.
EXECUTE.
SORT CASES BY randseq.
EXECUTE.

DATASET ACTIVATE mortvalidfemale
MATCH FILES /FILE=*
  /TABLE='mortvalidfemale'
  /BY randseq.
EXECUTE.
DATASET CLOSE mortvalidfemale.

DATASET ACTIVATE hospnat.
DATASET COPY  hospnatmale.
DATASET ACTIVATE  hospnatmale.
SELECT IF (gender_h=2).
EXECUTE.
DATASET CLOSE hospnat.

SET SEED=2000000.
COMPUTE randsort=UNIFORM (1).
FORMATS randsort (F8.7).
EXECUTE.
SORT CASES BY randsort (A).
EXECUTE.
IF ($casenum EQ 1) #counter=0.
IF (#counter EQ 225805) #counter=0.
COMPUTE #counter=#counter + 1.
COMPUTE randseq=#counter.
FORMATS randseq (F8).
VARIABLE LABELS randseq Random sequence.
EXECUTE.
SORT CASES BY randseq.
EXECUTE.

DATASET ACTIVATE mortvalidmale.
MATCH FILES /FILE=*
  /TABLE='mortvalidmale'
  /BY randseq.
EXECUTE.
DATASET CLOSE mortvalidmale.

DATASET ACTIVATE mortvalidfemale.
ADD FILES /FILE=*
  /FILE='mortvalidmale'.
EXECUTE.
```

```
DATASET CLOSE mortvalidmale.

SORT CASES by enc_nhi (A) event_start_date_h (A) event_end_date_h (A).
SAVE outfile  'filepath\hosp_nat_fake1.sav'.
DATASET NAME hospnatfake1.
```

Now each record in the hospital dataset has a ‘fake’ age at death in days, based on the (sex-specific or sex- and deprivation-specific) distribution of age at death in the mortality dataset from 1990 to 2006. This is used to calculate a ‘fake’ date of death. Further data preparation is largely similar to steps 1 to 7 and are so described only briefly.

This dataset (hosp_nat_fake1.sav) is merged (deterministically and probabilistically) with the mortality dataset and any individuals who have a ‘fake’ date of death prior to their observed date of death are deleted, as are individuals with a ‘fake’ date of death outside the observation window (1990 to 2006). Duplicate and overlapping hospital stays are resolved, hospital event-specific and aggregate length of stay variabies are created and the data is reduced to one record per individual, before being saved as hosp_nat_fake2.sav.

Before reducing the data to one line per individual for the deprivation analysis one further step is required. Individuals may move from a more deprived area to a less deprived area between hospital events. An area may become more or less deprived over time. Therefore, even in the absence of recording errors, an individual’s deprivation score can change over time and thus vary by hospital event. But only one deprivation value can be recorded once the dataset is reduced to one line per individual. As the analysis is concerned with the time preceding death, it is logical to use the deprivation value from the mortality analysis, this is also likely to be less prone to recording errors than the hospital deprivation value. However as we are applying a ‘fake’ date of death to individuals in the hospital dataset, it is necessary to sometimes rely on hospital deprivation values. In this situation the *last* hospital deprivation value recorded for the individual is used, this is the closest recoded deprivation to their ‘fake’ date of death. This is copied up to all hospital records for the individual before the dataset is reduced to one line per individual.

```
COMPUTE finaldep_h=scale_h.
VARIABLE LABELS finaldep_h last HOSPITAL deprivation score recorded for individual.
FORMATS finaldep_h (F1).
SORT CASES BY fileseq (D).
IF ($casenum=1 OR enc_nhi NE LAG (enc_nhi)) #flag=0.
IF (#flag=1) finaldep_h=LAG (finaldep_h).
DO IF (#flag=0 AND enc_nhi=LAG (enc_nhi)).
COMPUTE finaldep_h=LAG (finaldep_h).
COMPUTE #flag=1.
```

```
END IF.
EXECUTE .
SORT CASES BY fileseq (A).
```

Many individuals will have gender, date of birth and (for the deprivation analysis) deprivation recorded on both mortality and hospital records, while others have only mortality or only hospital records. It is necessary to create prioritised variables for gender, date of birth and (where appropriate) deprivation that are equal to the mortality value where available, and is equal to the hospital value where mortality data is not available.

```
IF gender_m>0 mortality=1.
RECODE mortality (1=1) (ELSE=0).
IF (mortality = 0) dob=dob_date_h.
IF (mortality = 1) dob=dob_m.
IF (mortality = 0) gender=gender_h.
IF (mortality = 1) gender=gender_m.
IF (mortality = 0) scale=finaldep_h.
IF (mortality = 1) scale=scale_m.
VARIABLE LABEL dob 'Prioritised date of birth (mortality if available, otherwise hospital)'.
VARIABLE LABEL gender 'Prioritised gender (mortality if available, otherwise hospital)'.
VARIABLE LABEL scale 'Prioritised deprivation (mortality if available, otherwise hospital)'.
FORMATS dob (EDATE10).
FORMATS gender (F1).
FORMATS scale (F1).
EXECUTE.
```

The datasets hosp_nat_fake2 and hosp_dep_fake2 are now ready to be used in the reference point modelling to calculate the baseline hospital use in months prior to death. The discussion now moves on to data preparation for the person-month logistic regression, and how this differs from the data preparation for reference point modelling.

Step 9 – Data Preparation for Person-Month Logistic Regression

The process followed when preparing the data person-month logistic regression is very similar to preparing the data for reference point modelling until step 4, and alike for steps 6 and 7, but they are not necessarily conducted in the same order. An entire linked dataset M, H and MH (all mortality, all hospital and linked mortality and hospital records) is created for the period 1988 to 2006 (Step 1) and hospital records with no valid enc_nhi deleted (1790 records).

The national filter is then applied as in Step 3 (this analysis is only conducted at a national level due to erratic results for smaller disaggregations). Then overlapping and duplicate hospital stays are resolved as in Step 4, and probabilistic matching is conducted in the same way as show in Step 2, and the dataset is saved as hosp_nat_merged1.sav. This is the point at which data preparation for reference point modelling and person-month logistic regression diverge. Before the aggregated length of stay variables are calculated for the 24 months either side of age x, the date of birth (and thus the date the individual attains age x) needs to be clarified. Some records have only a hospital date of birth, some have only a mortality date of birth, and some have both. The mortality date of birth is considered more reliable than the hospital date of birth, thus a new variable needs to be created that prioritises the mortality date of birth, and is equal to the hospital date of birth where there is no mortality information (dob). There is however a second consideration that needs to be addressed first. Due to the leniencies of the probabilistic matching, it is possible for the hospital date of birth to be different for different hospital events relating to the same individual. For individuals where there is no mortality information and the hospital date of birth is required, it needs to be consistent for all records relating to that individual. The date of birth from the first hospital event for an individual is chosen (arbitrarily) and a new variable (dob_h_adjusted) is created, which copies the first hospital date of birth for an individual down to all records for that individual.

```
COMPUTE dob_h_adjusted=dob_date_h.
IF (morbseq=1) #flag=0.
IF (#flag=1) dob_h_adjusted=lag (dob_h_adjusted).
DO IF (morbseq NE 1 AND #flag=0).
COMPUTE dob_h_adjusted=lag (dob_h_adjusted).
COMPUTE #flag=1.
END IF.
FORMATS dob_h_adjusted (EDATE10).
VARIABLE LABELS dob_h_adjusted Hospital date of birth adjusted to be the same for every event
relating to an individual.
EXECUTE.
```

It is now possible to calculate the prioritised date of birth (dob). The variable that indicates which dataset the information has come from can (dataset) be used to identify where there is only hospital and no mortality information. Where dataset is equal to 1, there is only hospital and no mortality information.

```
COMPUTE dob=dob_m.
FORMATS dob (EDATE10).
IF (dataset=1) dob=dob_h_adjusted.
```

```
EXECUTE.
```

The date of attaining exact age x, that is, the x'th birthday of each individual can now be calculated. In this example, 80 years is age x. When conducting person-month logistic regression for other ages (70, 75, 85) the syntax needs to be rerun from this point for that age.

```
COMPUTE bday80=DATESUM (dob, 80, "years", 'closest').
VARIABLE LABELS bday80 "date of birth plus 80 years (mortality record or if not available
adjusted hospital record)".
VARIABLE LEVEL bday80 (SCALE).
FORMATS bday80 (EDATE10).
VARIABLE WIDTH bday80 (10).
EXECUTE.
```

It is necessary to cut the dataset down to include only individuals who attain age x within the time period for which we have data (1st January 1988 to 31st December 2000 and 1st January 2002 to 31st December 2006). However each individual is observed for 672 days before and 672 days after their xth birthday (24 months). So this window is reduced to 672 days after 1st January 1988 to 672 days before 31st December 2000, and 672 days after 1st January 2002 to 672 days before 31st December 2006.

```
SELECT IF ( (bday80 >= DATE.DMY (03,11,1989)
AND bday80 <= DATE.DMY (28,02,1999))
OR (bday80 >= DATE.DMY (04,11,2003)
AND bday80 <= DATE.DMY (27,02,2005))).
EXECUTE.
```

The dates of 24 months prior to, and following age x are now calculated. These are necessary to create the length of stay variables as in Steps 5 and 6. First, the date of 24 months prior to age x is calculated as month0. A scratch variable (#counter) is then used to create 48 dates after the date of month0, increasing sequentially by 28 days.

```
COMPUTE month0=DATESUM (bday80, -672, "days", 'closest').
VARIABLE LEVEL month0 (SCALE).
FORMATS month0 (EDATE10).
VARIABLE WIDTH month0 (10).
EXECUTE.

COMPUTE #counter=0.
FORMATS #counter (F10.0).
DO REPEAT month=month1 to month48.
COMPUTE #counter= (#counter+28).
COMPUTE month=DATESUM (month0,+ (#counter),"days",'closest').
FORMATS month (EDATE10).
END REPEAT.
EXECUTE.
```

It is useful at this stage to check that the date of the xth birthday (bday80) is the same as the date of month 24. The hospital event-specific length of stay variables (losm1 to losm48) are created at this stage in a similar way to Step 5. Given that all the variable names for the month variables (month0 to month48) end numerically it is not necessary to do this separately for the first month. The aggregated length of stay variables (totlos1 to totlos48) are calculated in exactly the same way as Step 6 (not shown here).

```
DO REPEAT los=losm1 to losm48
/month=month0 to month47
/month1=month1 to month48.
IF (enddate_h>=month AND enddate_h<month1 AND event_start_date_h>=month AND
event_start_date_h<month1) los=DATEDIF (enddate_h, event_start_date_h, "days").
IF (enddate_h>=month AND enddate_h<month1 AND event_start_date_h<month) los=DATEDIF
(enddate_h, month, "days").
IF (enddate_h>=month1 AND event_start_date_h<month1 AND event_start_date_h>=month)
los=DATEDIF (month1, event_start_date_h, "days").
IF (enddate_h>=month1 AND event_start_date_h<month) los=28.
RECODE los (0=0.5).
FORMATS los (F8.1).
END REPEAT.
EXECUTE.

DO REPEAT los=losm1 TO losm48.
RECODE los (SYSMIS=0) (ELSE=COPY) INTO los.
END REPEAT.
EXECUTE.
```

Before deleting the month date variables (month1 to month48), they are required for one more thing. An indicator of death in each month (dead) needs to be calculated. This will be the outcome variable in the person-month logistic regression. Once this has been created the event-specific length of stay and month date variables (los1 to los48 and month0 to month48) can be deleted (not shown here) and the data reduced to one line per individual. Due to the deterministic matching of sex, hospital and mortality values of sex will always be the same. Therefore one of the (hospital/mortality) sex variables can be deleted and the other renamed into gender. This one record per individual dataset is saved as hosp_nat_indiv_merged.sav, keeping only essential non hospital event-specific variables.

```
DO REPEAT dead=deadm1 to deadm48
/month=month0 to month47
/month1=month1 to month48.
IF (dod_m>=month AND dod_m<month1) dead=1.
RECODE dead (SYSMIS=0) (ELSE=COPY) INTO dead.
FORMATS dead (F1).
END REPEAT.
EXECUTE.
```

```
RENAME VARIABLES gender_m=gender.

SELECT IF morbseq=1.
EXECUTE.
COMPUTE fileseq=$casenum.
EXECUTE.

DELETE VARIABLES event_id_h dob_date_h end_type_h event_start_date_h morbseq enddate_h
gender_h.

SAVE outfile 'filepath\hosp_nat_indiv_merged.sav'.
DATASET NAME hospnatindivmerged.
```

The dataset can now be transposed from person-event format (with only one 'event', or record, per individual) to person-month format (one record for each of the 48 person-months identified by totlos1 to totlos48), using the variables to cases command.

```
DELETE VARIABLES fileseq.

VARSTOCASES
/MAKE los FROM totlosm1 totlosm2 totlosm3 totlosm4
totlosm5 totlosm6 totlosm7 totlosm8 totlosm9 totlosm10 totlosm11 totlosm12 totlosm13
totlosm14 totlosm15 totlosm16 totlosm17 totlosm18 totlosm19 totlosm20 totlosm21    totlosm22
totlosm23 totlosm24 totlosm25 totlosm26 totlosm27 totlosm28 totlosm29 totlosm30 totlosm31
totlosm32 totlosm33 totlosm34 totlosm35 totlosm36 totlosm37    totlosm38 totlosm39 totlosm40
totlosm41 totlosm42 totlosm43 totlosm44 totlosm45 totlosm46 totlosm47 totlosm48
/MAKE dead FROM deadm1 deadm2 deadm3 deadm4 deadm5 deadm6 deadm7 deadm8 deadm9
deadm10 deadm11 deadm12 deadm13 deadm14 deadm15 deadm16 deadm17 deadm18 deadm19
deadm20 deadm21 deadm22 deadm23 deadm24 deadm25 deadm26 deadm27 deadm28 deadm29
deadm30 deadm31 deadm32 deadm33 deadm34 deadm35 deadm36 deadm37 deadm38 deadm39
deadm40 deadm41 deadm42 deadm43 deadm44
deadm45 deadm46 deadm47 deadm48
/INDEX=month (48)
/KEEP=enc_nhi dod_m dob gender scale_m regyear_m icdgroup_b_m age_grouped_m bday80
dataset
/NULL=KEEP.

COMPUTE fileseq=$casenum.
EXECUTE.
```

A little further data preparation is required before person-month logistic regression can be conducted. In order to identify if the individual is still alive in each person-month, or has already died, a variable (alive) is created.

```
COMPUTE alive=1.
FORMATS alive (f1).
IF (month=1) #flag=0.
IF (#flag=1) alive=0.
DO IF (dead=1 AND #flag=0).
COMPUTE alive=0.
COMPUTE #flag=1.
END IF.
VARIABLE LABELS alive "marker of alive or dead status in each month".
```

```
VALUE LABELS alive 1 alive
0 dead, or death in current month.
EXECUTE.
```

In order to run the analysis for different years, a variable indicating which year the individual turns age x (year80) also needs to be created.

```
COMPUTE year80=XDATE.YEAR (bday80).
VARIABLE LABEL year80 "Year turned 80".
VARIABLE LEVEL year80 (SCALE).
FORMATS year80 (F8.0).
VARIABLE WIDTH year80 (8).
EXECUTE.
```

An average, or expected number of bed days in hospital in each person month by year and gender needs to be created. This is deducted from the observed number of days spent in hospital in each person month to give the ‘excess’ bed days, the difference in length of stay between the observed and expected bed days (losdifference).

```
USE ALL.
COMPUTE filter_$= (alive=1).
VARIABLE LABEL filter_$ 'alive=1 (FILTER)'.
VALUE LABELS filter_$ 0 'Not Selected' 1 'Selected'.
FORMAT filter_$ (f1.0).
FILTER BY filter_$.
EXECUTE.

SORT CASES BY gender month year80.
AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /PRESORTED
  /BREAK=gender month year80
  /meanlos=MEAN (los).

SORT CASES BY fileseq (A).
FILTER OFF.
USE ALL.
DELETE VARIABLES filter_$.
```

The lag of this losdifference variable in previous months is used as a predictor in the regression model (laglosdifference1 to laglosdifference48).

```
COMPUTE losdifference1=los-meanlos.
IF (enc_nhi=lag (enc_nhi)) laglosdifference1=lag (losdifference).
FORMATS laglos (F8.1).
VARIABLE LABELS laglosdifference1 "difference between expected and observed hospital bed
days in previous month".
EXECUTE.

IF month>lag (month,2) laglosdifference2=lag (losdifference,2).
IF month>lag (month,3) laglosdifference3=lag (losdifference,3).
IF month>lag (month,4) laglosdifference4=lag (losdifference,4).
```

```
IF month>lag (month,5) laglosdifference5=lag (losdifference,5).
IF month>lag (month,6) laglosdifference6=lag (losdifference,6).
IF month>lag (month,7) laglosdifference7=lag (losdifference,7).
IF month>lag (month,8) laglosdifference8=lag (losdifference,8).
IF month>lag (month,9) laglosdifference9=lag (losdifference,9).
IF month>lag (month,10) laglosdifference10=lag (losdifference,10).
IF month>lag (month,11) laglosdifference11=lag (losdifference,11).
IF month>lag (month,12) laglosdifference12=lag (losdifference,12).
IF month>lag (month,13) laglosdifference13=lag (losdifference,13).
IF month>lag (month,14) laglosdifference14=lag (losdifference,14).
IF month>lag (month,15) laglosdifference15=lag (losdifference,15).
IF month>lag (month,16) laglosdifference16=lag (losdifference,16).
IF month>lag (month,17) laglosdifference17=lag (losdifference,17).
IF month>lag (month,18) laglosdifference18=lag (losdifference,18).
IF month>lag (month,19) laglosdifference19=lag (losdifference,19).
IF month>lag (month,20) laglosdifference20=lag (losdifference,20).
IF month>lag (month,21) laglosdifference21=lag (losdifference,21).
IF month>lag (month,22) laglosdifference22=lag (losdifference,22).
IF month>lag (month,23) laglosdifference23=lag (losdifference,23).
IF month>lag (month,24) laglosdifference24=lag (losdifference,24).
EXECUTE.
```

```
VARIABLE LABELS los "number of days spent in hospital in current month".
VARIABLE LABELS losdifference1 "difference between expected and observed hospital bed days
in current month".
VARIABLE LABELS month "month after turning 80 years old".
VARIABLE LABELS dead "death in current month".
VARIABLE LABELS dob "date of birth (from mortality record, otherwise from first hospital
record)".
VARIABLE LABELS meanlos "expected (average) number of bed days in current month, by year
and sex".
```

This step is repeated to generate a dataset for wach age at death (75, 80, 85). The data processing for the person-month logistic regression is now complete, and the dataset is saved as hosp_nat_indiv_transposed.sav.

```
SAVE outfile 'filepath\hosp_nat_indiv_transposed.sav'.
DATASET NAME hospnatindivtransposed.
```

Having described the data preparation for reference point modelling and person-month logistic regression, the method of obtaining output for these methods is now discussed.

Step 10 – Production of Output for Reference Point Modelling

The production of output for reference point modelling is very simple. The final dataset with all adjustments made (merged_nat_indiv2.sav) is opened, the file is split by gender, and the compare means function is used to generate crosstabulations.

```
Get file 'filepath\merged_nat_indiv2.sav'.
DATASET NAME mergednatindiv2.
DATASET ACTIVATE mergednatindiv2.

SORT CASES  BY gender.
SPLIT FILE SEPARATE BY gender.

MEANS TABLES=totlosm1 totlosm2 totlosm3 totlosm4 totlosm5 totlosm6 totlosm7 totlosm8
totlosm9 totlosm10 totlosm11 totlosm12 totlosm13 totlosm14 totlosm15 totlosm16 totlosm17
totlosm18 totlosm19 totlosm20 totlosm21 totlosm22 totlosm23 totlosm24
BY regyear_m
BY age_grouped
  /CELLS MEAN.

SPLIT FILE OFF.
```

The same process is used to generate tables for the deprivation analysis (adding a 'BY scale_m' in the compare means command). And for the baseline national and area deprivation hospital use in the months prior to 'fake' death date. These tables are copied into Excel in order to generate the graphs presented in the results section of Paper Three.

Step 11 – Production of Output for Person-Month Logistic Regression

Producing output for the person-month logistic regression analysis is rather more involved. Running the logistic regression models is straightforward enough, the relevant dataset (hosp_nat_indiv_transposed.sav) is opened, records for individuals turning age x in a given year (1990) are selected, person-months where the individual is no longer alive, and where the current month is prior to age x are excluded. The dataset is split by gender and a separate logistic regression model run for each previous month, going back to 24. This process is repeated for each year (1990-1998 and 2004) and age at death (75, 80, 85). The syntax below presents an abridged version of the output for age 80 in 1990, and shows how the output for age 80 in 1990 is exported and closed and the data is set up ready for age 80 year 1991.

```
GET FILE "filepath\hosp_nat_indiv_transposed.sav".
DATASET NAME hospnatindivtransposed.

DATASET COPY  year1990.
DATASET ACTIVATE  year1990.
FILTER OFF.
USE ALL.
SELECT IF ( (alive=1 OR dead=1) AND year80=1990 AND month>24).
EXECUTE.
```

```
SORT CASES  BY gender.
SPLIT FILE SEPARATE BY gender.

LOGISTIC REGRESSION VARIABLES dead
  /METHOD=ENTER month laglosdifference1
  /PRINT=CI (95)
  /CRITERIA=PIN (.05) POUT (.10) ITERATE (20) CUT (.5).
.  .  .

LOGISTIC REGRESSION VARIABLES dead
  /METHOD=ENTER month laglosdifference24
  /PRINT=CI (95)
  /CRITERIA=PIN (.05) POUT (.10) ITERATE (20) CUT (.5).

OUTPUT EXPORT
  /CONTENTS  EXPORT=ALL  LAYERS=VISIBLE  MODELVIEWS=VISIBLE
  /XLS  DOCUMENTFILE= 'filepath\nat1990.xls'
     OPERATION=CREATEFILE
     LOCATION=LASTCOLUMN  NOTESCAPTIONS=NO.

OUTPUT NAME nat1990.
EXECUTE.
OUTPUT CLOSE nat1990.
DATASET CLOSE year1990.
DATASET ACTIVATE hospnatindivtransposed.

DATASET COPY year1991.
```

It is evident that a considerable quantity of output is produced, from which only a few figures are required from each model. This would be incredibly time consuming to do manually. The solution was to automatically export the output to MS Excel files (as shown), enabling the user to leave SPSS producing output for a given age and all years overnight; and to use Visual Basic (VB) code to extract the required figures and use these figures to automatically generate the graphs presented in the results section of Paper Three.

The VB code required to do this is presented below. First the code needs to fetch the required statistics (upper confidence interval, odds ratio, lower confidence interval, and number of deaths observed) from the logistic regression output on Sheet1 and paste them into Sheet2.

```
Sub data_parser ()
'03 Sept 10 by Steve Pilbeam
'Step 1, a quick check to ensure that the data import hasn't exceeded Excel's row limit (about 65,000
lines)

LastRow = Sheets ("Sheet1").UsedRange.Rows.Count

If LastRow > 55000 And LastRow < 64000 Then
   MsgBox "Your imported data contains " & LastRow & " rows of data, which is getting near the
data limit for Excel. The macro will try to run anyway."
End If
```

```
If LastRow > 64000 Then
   MsgBox "Your imported data contains " & LastRow & " rows of data, which has probably
exceeded the data limit for Excel. The macro will try to run anyway."
End If
'Step 2. Create a new Sheet2 to ensure it is empty and free from formatting

Dim wsSheet2 As Worksheet
   On Error Resume Next
   Set wsSheet2 = Worksheets ("Sheet2")
   On Error GoTo 0

Application.DisplayAlerts = False

   If Not wsSheet2 Is Nothing Then
      Sheets ("Sheet2").Delete

   End If
Sheets.Add.Name = "Sheet2"

Application.DisplayAlerts = True

Range ("Sheet2!A1").Value = "Gender"
Range ("Sheet2!B1").Value = "laglosdifference"
Range ("Sheet2!C1").Value = "Sig."
Range ("Sheet2!D1").Value = "ExpB ()"
Range ("Sheet2!E1").Value = "Lower"
Range ("Sheet2!F1").Value = "Upper"
Range ("Sheet2!G1").Value = "Death in current month"


'Step 3. Go through the data in Sheet1 and identify the right bits. This will assume that the data was
pasted into cell A1

Dim i_increment As Integer
Dim s_gender As String
Dim i_gender As Integer
i_increment = 2
i_gender = 1

For n = 1 To LastRow

   If Left (Sheets ("Sheet1").Cells (n, 2).Value, 16) = "laglosdifference" Then

   Select Case i_gender
   Case 1
   s_gender = "Female"
   Case -1
   s_gender = "Male"
   End Select

      Sheets ("Sheet2").Cells (i_increment, 1).Value = s_gender
      Sheets ("Sheet2").Cells (i_increment, 2).Value = Sheets ("Sheet1").Cells (n, 2).Value
      Sheets ("Sheet2").Cells (i_increment, 3).Value = Sheets ("Sheet1").Cells (n, 7).Value
      Sheets ("Sheet2").Cells (i_increment, 4).Value = Sheets ("Sheet1").Cells (n, 8).Value
      Sheets ("Sheet2").Cells (i_increment, 5).Value = Sheets ("Sheet1").Cells (n, 9).Value
      Sheets ("Sheet2").Cells (i_increment, 6).Value = Sheets ("Sheet1").Cells (n, 10).Value

      i_increment = i_increment + 1
      i_gender = i_gender * -1
```

```
   End If

Next n
' now adding the death in current month for Block 0, Step 0

i_increment = 2
For n = 1 To LastRow

If Sheets ("Sheet1").Cells (n, 1).Value = "Step 0" Then
   If Sheets ("Sheet1").Cells (n, 2).Value = "death in current month" Then

      Sheets ("Sheet2").Cells (i_increment, 7).Value = Sheets ("Sheet1").Cells (n + 1, 4).Value

      i_increment = i_increment + 1

   End If
End If

Next n

End Sub
```

A split file is used when generating the output in SPSS. This results in the output being in a female, male, female, male order. The values that have been copied to Sheet2 are still in this order. It is necessary to sort the data by sex (column 2).

```
Sub Sort ()

' Macro recorded 16/11/2010 by Ngaire Coombs

   Range ("A1:G49").Sort Key1:=Range ("A2"), Order1:=xlAscending, Header:= _
      xlGuess, OrderCustom:=1, MatchCase:=False, Orientation:=xlTopToBottom, _
      DataOption1:=xlSortNormal
End Sub
```

No available graph templates (line/area/bar) in Excel were suitable for the type of graph required. In order to produce a graph with a vertical line for each model, with a shaded area between the lower confidence interval and 1 on the y axis, it was necessary to use VB code. A graph was generated by plotting every point separately, creating an area graph below the lower confidence intervals, and overlaying a white rectangle below 1 on the y axis. The highly complex code required to achieve this is presented below.

```
Sub my_graph ()

'Prerequisites: data is in Sheet2 in the range A2:H24 i.e. it has a header row
'and Sheet3 does not exist with anything worth keeping
'Steve Pilbeam 07/10/10

Application.DisplayAlerts = False
On Error Resume Next
```

```
Sheets ("Sheet3").Delete
Application.DisplayAlerts = True
On Error GoTo 0
Sheets.Add ().Name = "Sheet3"
   Sheets ("Sheet3").Columns (1).ClearContents
   Sheets ("Sheet3").Columns (2).ClearContents

For n = 2 To 25
Sheets ("Sheet2").Cells (n, 8).Value = n - 1
Next n
For n = 26 To 49
Sheets ("Sheet2").Cells (n, 8).Value = n - 24
Next n
''''''
```

```
skip_n:
Range ("Sheet3!C10").Value = "Month"
Range ("Sheet3!D10").Value = "Lower"
Range ("Sheet3!D8").Value = "1st intercept"
Range ("Sheet3!D9").Value = x_3

Dim i_scale As Long
If Range ("Sheet3!A2").Value = 0 Then Range ("Sheet3!A2").Value = 1
Range ("Sheet3!A1").Value = "Scale Factor"
i_scale = Range ("Sheet3!A2").Value

'initial offset values

i_x_off = 250
i_y_off = 100



'base size (can therefore change relative scaling)

i_x_base = 520
i_y_base = 170

'square coords are y-reversed from top left corner

'Horizontals

Sheets ("Sheet3").Shapes.AddLine (i_x_off, i_y_off, i_x_off + (i_x_base * i_scale), i_y_off).Select

Sheets ("Sheet3").Shapes.AddLine (i_x_off, i_y_off + (i_y_base * i_scale / 2), i_x_off + (i_x_base
* i_scale), i_y_off + (i_y_base * i_scale / 2)).Select

Sheets ("Sheet3").Shapes.AddLine (i_x_off, i_y_off + (i_y_base * i_scale), i_x_off + (i_x_base *
i_scale), i_y_off + (i_y_base * i_scale)).Select

'Verticals

Sheets ("Sheet3").Shapes.AddLine (i_x_off, i_y_off, i_x_off, i_y_off + (i_y_base * i_scale)).Select

Sheets ("Sheet3").Shapes.AddLine (i_x_off + (i_x_base * i_scale), i_y_off, i_x_off + (i_x_base *
i_scale), i_y_off + (i_y_base * i_scale)).Select


''''''
```

```
ActiveSheet.Shapes.AddTextbox (msoTextOrientationHorizontal, i_x_off + (i_gap * i_scale * n) -
8, i_y_off + (i_y_base * i_scale) + (i_y_base / 45), 30, 30).Select

s_label = 24 - n
Selection.Characters.Text = s_label
   With Selection.Font
      .Name = "Arial"
      .FontStyle = "Regular"
      .Size = 14
      .Strikethrough = False
      .Superscript = False
      .Subscript = False
      .OutlineFont = False
      .Shadow = False
      .Underline = xlUnderlineStyleNone
      .ColorIndex = xlAutomatic
   End With
   Selection.ShapeRange.Fill.Visible = msoFalse
   Selection.ShapeRange.Line.Visible = msoFalse
Next n

'y_axis marker lines

i_gap_y = i_y_base / 10

'marker lines

For n = 0 To 10

Sheets ("Sheet3").Shapes.AddLine (i_x_off + (i_x_base * i_scale), i_y_off + (i_gap_y * n), i_x_off
+ (i_x_base * i_scale) + 5, i_y_off + (i_gap_y * n)).Select

'marker  numbers

ActiveSheet.Shapes.AddTextbox (msoTextOrientationHorizontal, i_x_off + (i_x_base * i_scale) +
7, i_y_off + (i_gap_y * n) - 11, 40, 25).Select

s_label = 1.5 - 0.1 * n

Selection.Characters.Text = s_label
   With Selection.Font
      .Name = "Arial"
      .FontStyle = "Regular"
      .Size = 14
      .Strikethrough = False
      .Superscript = False
      .Subscript = False
      .OutlineFont = False
      .Shadow = False
      .Underline = xlUnderlineStyleNone
      .ColorIndex = xlAutomatic
   End With
   Selection.ShapeRange.Fill.Visible = msoFalse
   Selection.ShapeRange.Line.Visible = msoFalse
Next n

''''''
```

```
For n = 11 To 34

i_h = Sheets ("Sheet3").Cells (n, 3).Value - 1

i_v = Sheets ("Sheet3").Cells (n, 4).Value
If i_v > 1.5 Then i_v = 1.5
If i_v < 0.5 Then i_v = 0.5

i_v = 1 - i_v

i_v = i_v / 0.5

i_v = (i_y_base * i_scale * i_v / 2)

Sheets ("Sheet3").Cells (n, 5).Value = i_x_off + (i_x_base * i_scale) - (i_h * i_gap)
Sheets ("Sheet3").Cells (n, 6).Value = i_y_off + (i_y_base * i_scale / 2) + i_v

Next n

'Add top solid rectangle

   With Sheets ("Sheet3").Shapes.BuildFreeform (msoEditingAuto, i_x_off, i_y_off)

   .AddNodes msoSegmentLine, msoEditingAuto, i_x_off + (i_x_base * i_scale), i_y_off
   .AddNodes msoSegmentLine, msoEditingAuto, i_x_off + (i_x_base * i_scale), i_y_off +
   (i_y_base * i_scale / 2)
   .AddNodes msoSegmentLine, msoEditingAuto, i_x_off, i_y_off + (i_y_base * i_scale / 2)
   .AddNodes msoSegmentLine, msoEditingAuto, i_x_off, i_y_off

   .ConvertToShape.Select

   End With

   Selection.ShapeRange.Fill.Solid
   Selection.ShapeRange.Fill.ForeColor.SchemeColor = 9

'Add the polygon

   With Sheets ("Sheet3").Shapes.BuildFreeform (msoEditingAuto, i_x_off + (i_x_base * i_scale),
   i_y_off + (i_y_base * i_scale))

      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E11").Value, Range
      ("Sheet3!F11").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E12").Value, Range
      ("Sheet3!F12").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E13").Value, Range
      ("Sheet3!F13").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E14").Value, Range
      ("Sheet3!F14").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E15").Value, Range
      ("Sheet3!F15").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E16").Value, Range
      ("Sheet3!F16").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E17").Value, Range
      ("Sheet3!F17").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E18").Value, Range
      ("Sheet3!F18").Value
      .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E19").Value, Range
      ("Sheet3!F19").Value
```

```
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E20").Value, Range
        ("Sheet3!F20").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E21").Value, Range
        ("Sheet3!F21").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E22").Value, Range
        ("Sheet3!F22").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E23").Value, Range
        ("Sheet3!F23").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E24").Value,
        Range ("Sheet3!F24").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E25").Value, Range
        ("Sheet3!F25").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E26").Value, Range
        ("Sheet3!F26").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E27").Value, Range
        ("Sheet3!F27").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E28").Value, Range
        ("Sheet3!F28").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E29").Value, Range
        ("Sheet3!F29").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E30").Value, Range
        ("Sheet3!F30").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E31").Value, Range
        ("Sheet3!F31").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E32").Value, Range
        ("Sheet3!F32").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E33").Value, Range
        ("Sheet3!F33").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E34").Value, Range
        ("Sheet3!F34").Value
        .AddNodes msoSegmentLine, msoEditingAuto, Range ("Sheet3!E34").Value, i_y_off +
        (i_y_base * i_scale)

        .ConvertToShape.Select

    End With

    Selection.ShapeRange.Fill.Solid
    Selection.ShapeRange.Fill.ForeColor.SchemeColor = 22

'Add lower solid rectangle

    With Sheets ("Sheet3").Shapes.BuildFreeform (msoEditingAuto, i_x_off + (i_x_base * i_scale),
    i_y_off + (i_y_base * i_scale / 2))

    .AddNodes msoSegmentLine, msoEditingAuto, i_x_off, i_y_off + (i_y_base * i_scale / 2)
    .AddNodes msoSegmentLine, msoEditingAuto, i_x_off, i_y_off + (i_y_base * i_scale)
    .AddNodes msoSegmentLine, msoEditingAuto, i_x_off + (i_x_base * i_scale), i_y_off +
    (i_y_base * i_scale)

    .ConvertToShape.Select

    End With

    Selection.ShapeRange.Fill.Solid
    Selection.ShapeRange.Fill.ForeColor.SchemeColor = 9

 'link the series

For n = 1 To 24
```

```
'lower

i_v_1 = Sheets ("Sheet3").Cells (10 + ( (n - 1) * 3) + 1, 2)
If i_v_1 < 0.5 Then i_v_1 = 0.5

i_v_1 = 1 - i_v_1

i_v_1 = i_v_1 / 0.5

i_v_1 = (i_y_base * i_scale * i_v_1 / 2)

'upper

i_v_2 = Sheets ("Sheet3").Cells (10 + ( (n - 1) * 3) + 3, 2)
If i_v_2 > 1.5 Then i_v_2 = 1.5

i_v_2 = 1 - i_v_2

i_v_2 = i_v_2 / 0.5

i_v_2 = (i_y_base * i_scale * i_v_2 / 2)

 Sheets ("Sheet3").Shapes.AddLine (i_x_off + (i_x_base * i_scale) - ( (n - 1) * i_gap), i_y_off +
(i_y_base * i_scale / 2) + i_v_1, i_x_off + (i_x_base * i_scale) - ( (n - 1) * i_gap), i_y_off +
(i_y_base * i_scale / 2) + i_v_2).Select

Next n

 'put in all points

For n = 11 To 82

 i_h = Sheets ("Sheet3").Cells (n, 1) - 1
i_v = Sheets ("Sheet3").Cells (n, 2)
If i_v > 1.5 Then GoTo skip_point
If i_v < 0.5 Then GoTo skip_point
i_v = 1 - i_v

i_v = i_v / 0.5

i_v = (i_y_base * i_scale * i_v / 2)

 If n / 3 <> Int (n / 3) Then

Sheets ("Sheet3").Shapes.AddLine (i_x_off + (i_x_base * i_scale) - (i_h * i_gap) - (i_x_base / 200),
i_y_off + (i_y_base * i_scale / 2) + i_v, i_x_off + (i_x_base * i_scale) - (i_h * i_gap) + (i_x_base /
200), i_y_off + (i_y_base * i_scale / 2) + i_v).Select

Selection.ShapeRange.Line.Weight = 1.5

Else

ActiveSheet.Shapes.AddShape (msoShapeOval, i_x_off + (i_x_base * i_scale) - (i_h * i_gap) - 3,
i_y_off + (i_y_base * i_scale / 2) + i_v - 3, 6, 6).Select

End If

skip_point:
Next n
```

```
Range ("Sheet3!E11:F34").ClearContents

'group the objects
   Dim ws As Worksheet
   Dim sh As Shape

   Set ws = Worksheets ("Sheet3")

   i_line = 1

   For Each sh In ws.Shapes

      sh.Select False

   Next sh

   Selection.ShapeRange.Group.Select

End Sub
```

The graph is inserted on Sheet 3, and is drawn for the first 24 rows after the title row on Sheet2 (the figures 'fetched' from the output on Sheet1, which have been sorted by gender). It is therefore drawn initially for females. In order to draw a graph for males, a separate (shorter) piece of code is needed to switch the data for males and females in Sheet 2, resulting in the data for males being present in the first 24 rows after the title row.

```
Sub switch_m_f ()

'Steve Pilbeam 07/10/10

Range ("Sheet2!I2:P25").Value = Range ("Sheet2!A2:H25").Value
Range ("Sheet2!A2:H25").Value = Range ("Sheet2!A26:H49").Value
Range ("Sheet2!A26:H49").Value = Range ("Sheet2!I2:P25").Value
Range ("Sheet2!I2:P25").ClearContents

End Sub
```

It would not have been possible to create the graphs presented for the person-month logistic regression without the use of VB by a skilled user. Assistance from Steve Pilbeam in facilitating the presentation of these results is gratefully acknowledged.

Conclusion

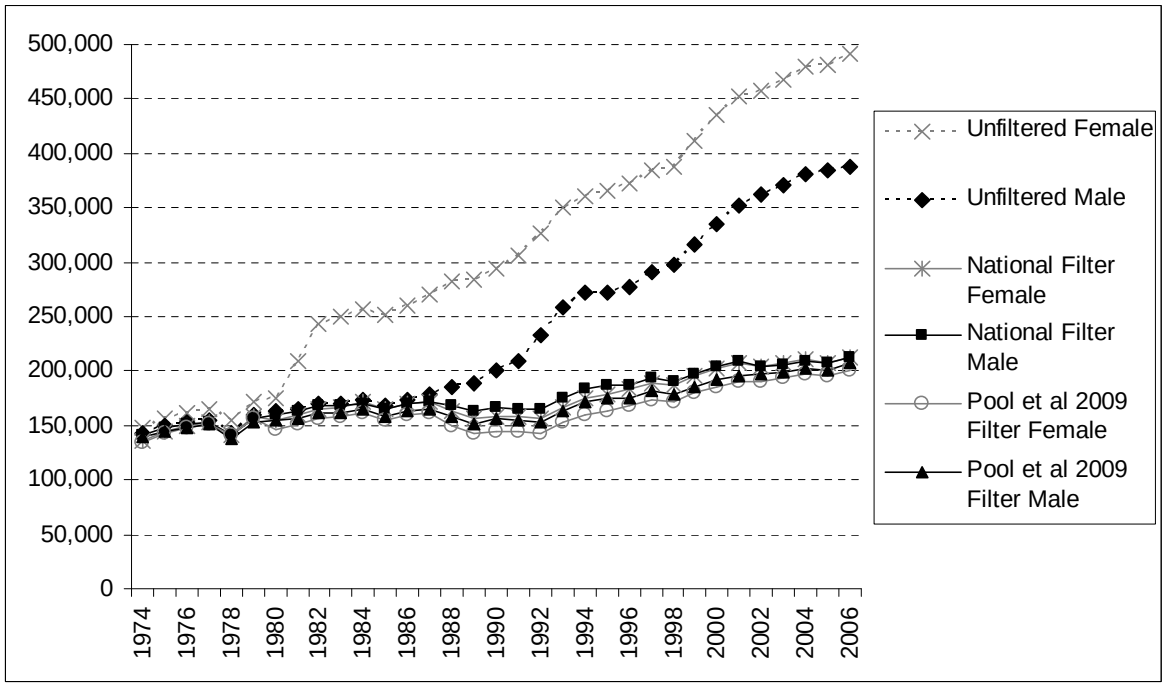
The challenges faced in the data preparation for Paper Three are very different to the challenges faced in Papers One and Two. Individual record linkage and longitudinal

methods are very powerful tools, but require extensive preparation, both in order to prepare the datasets for analysis, and to obtain and present results.

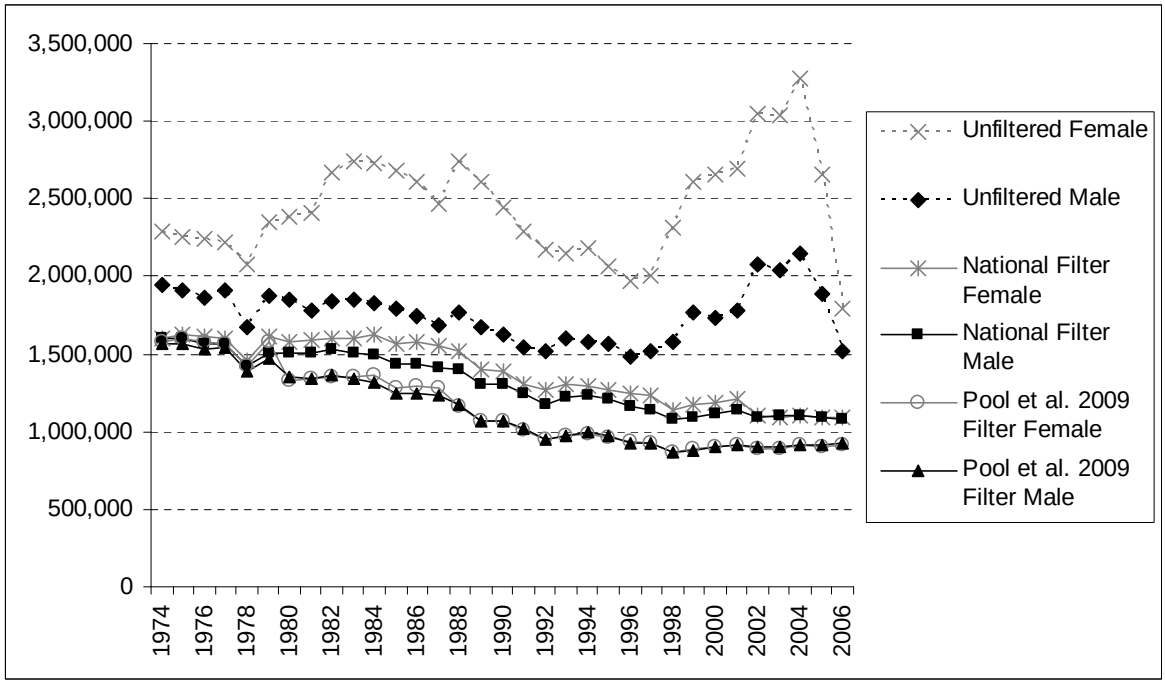
APPENDIXES

APPENDIX FIGURES

Appendix Figure 1: Number of Hospital Discharges by Sex, Unfiltered and Filtered, Comparison to Filter used by Pool *et al.* 2009

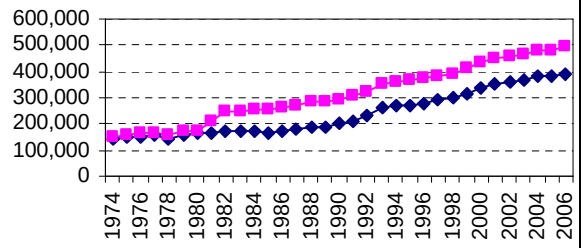


Appendix Figure 2: Number of Hospital Bed Days by Sex, Unfiltered and Filtered, Comparison to Filter used by Pool *et al.* 2009

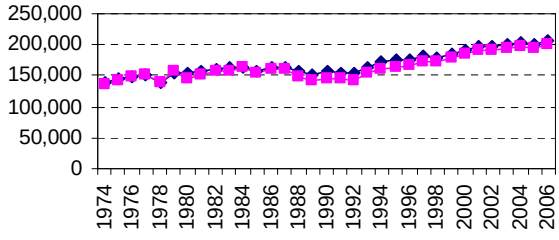


Appendix Figure 3: Number of Hospital Discharges by Filter Variable and Sex

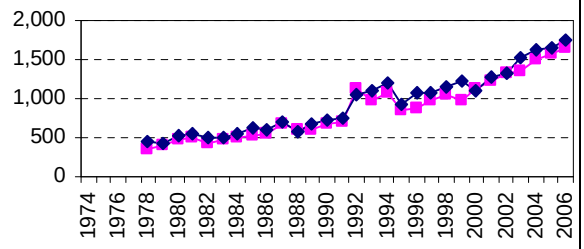
Total



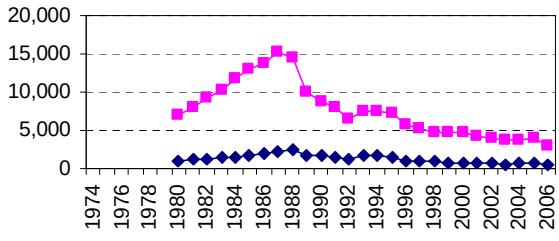
0 - Unfiltered



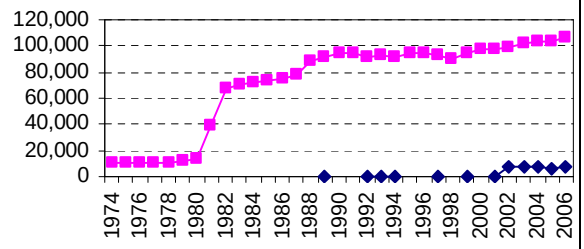
1 - Non-NZ residents



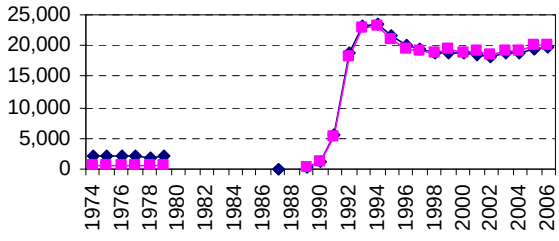
2 - Boarders



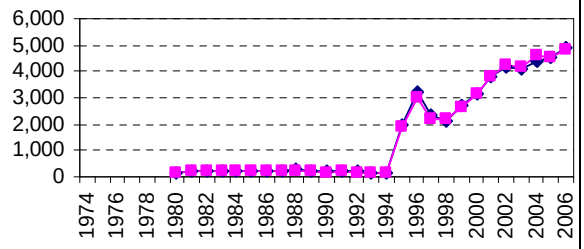
3 - Obstetrics/pregnancy Related



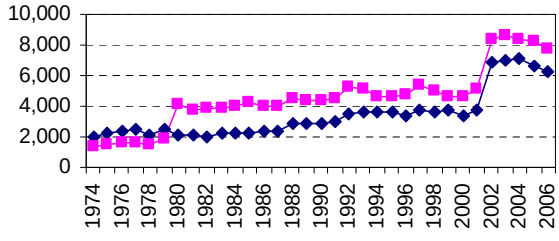
4 - Well babies



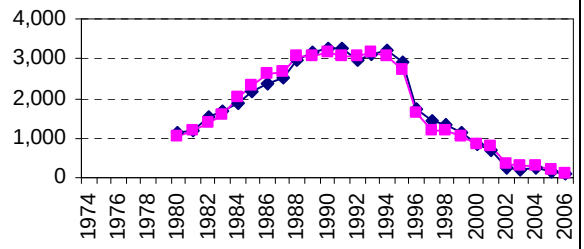
5 - Baby boarder



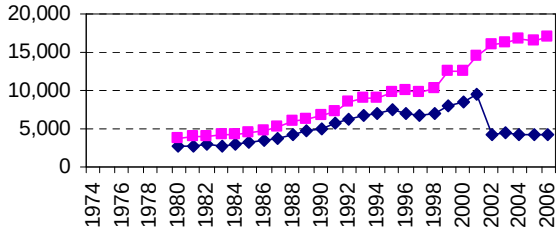
6 - Mental health



7 – Respite

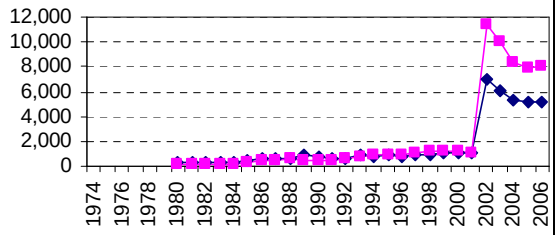


8 - Rehabilitation

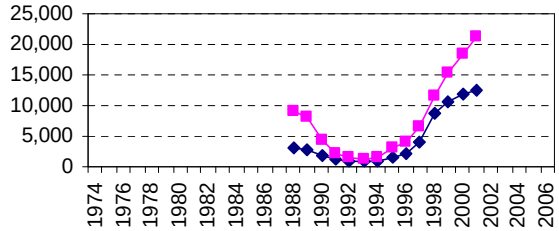


Appendix Figure 3 Continued…

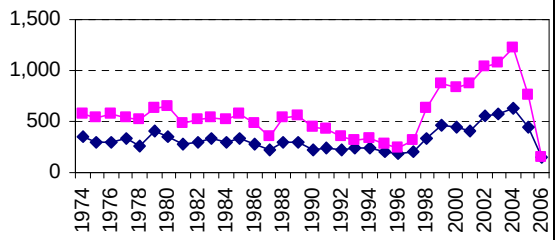
9 – Disability Support Services



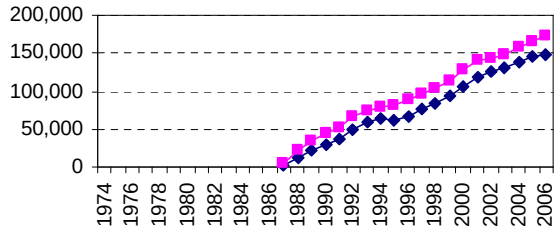
10 – Non-CHE



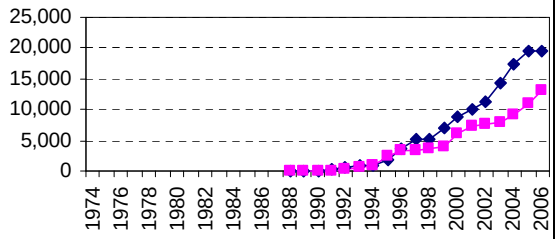
11 – Long stay (>365 days)



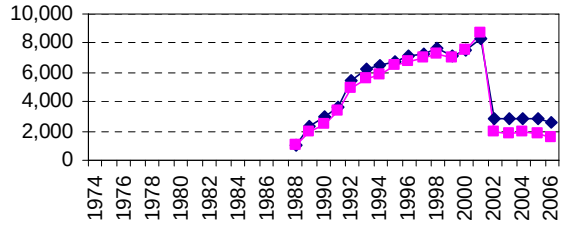
12 - Day patients



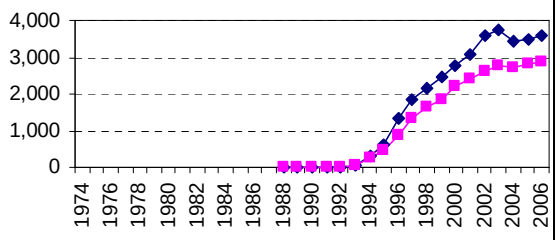
13 - Same day chemotherapy/radiotherapy/renal dialysis



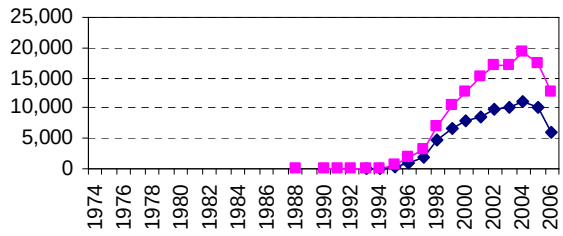
14 - Same day gastroscopies/colonoscopies/cystoscopies



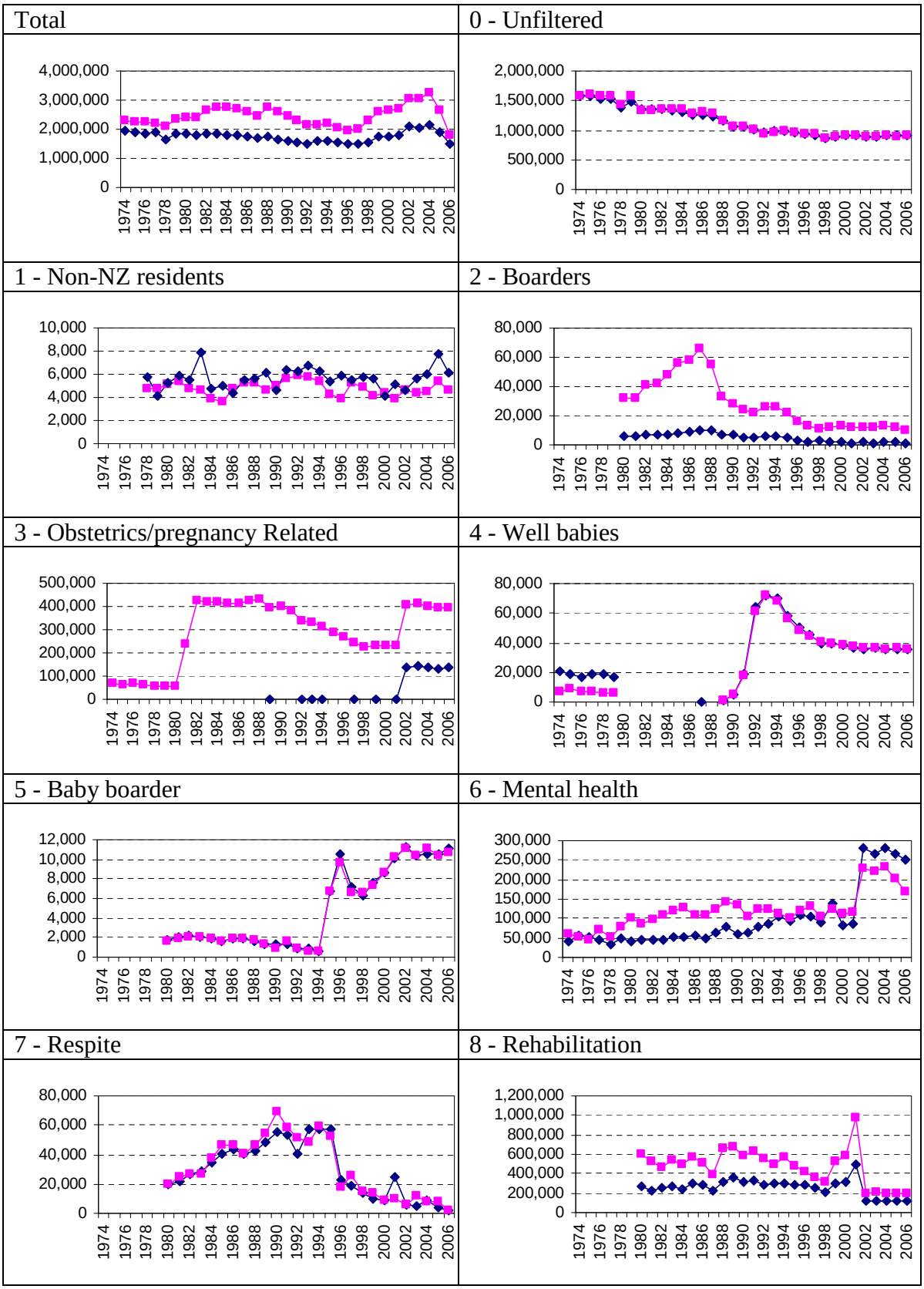
15 - Same day blood transfusions



16 – Publicly funded hospital events with no procedure

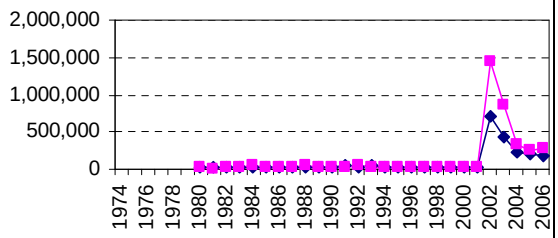


Appendix Figure 4: Number of Hospital Bed Days by Filter Variable and Sex

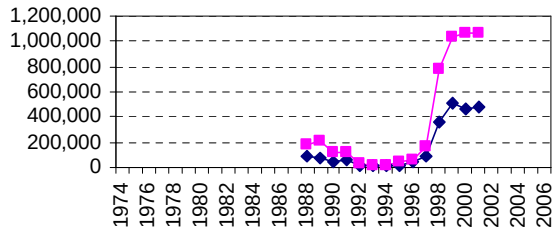


Appendix Figure 4 Continued…

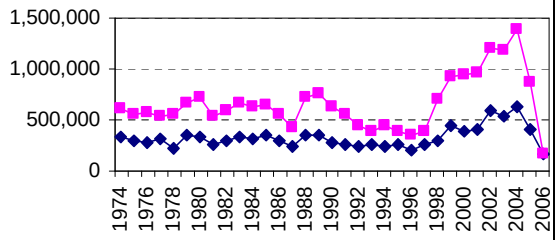
9 – Disability Support Services



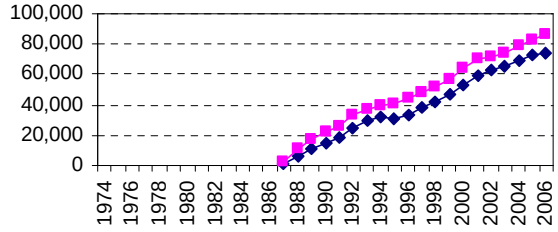
10 – Non-CHE



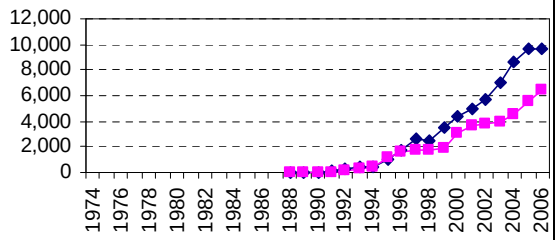
11 – Long stay (>365 days)



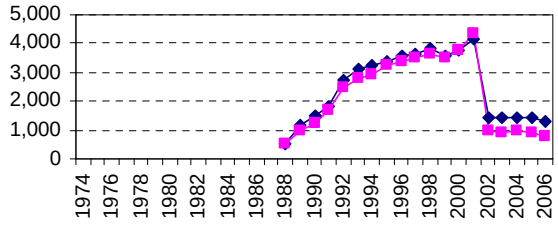
12 - Day patients



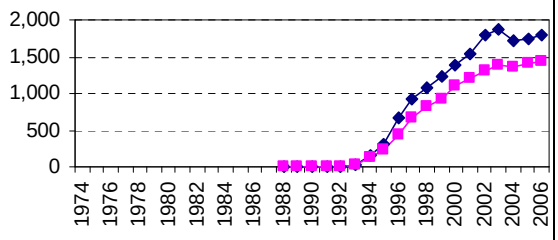
13 - Same day chemotherapy/radiotherapy/renal dialysis



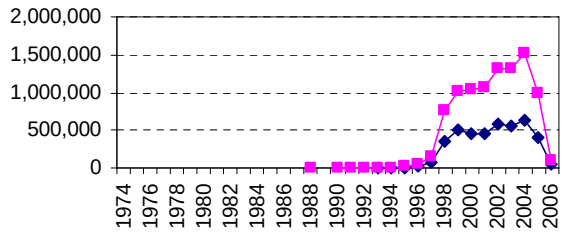
14 - Same day gastroscopies/colonoscopies/cystoscopies

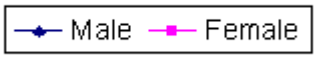


15 - Same day blood transfusions



16 – Publicly funded hospital events with no procedure





APPENDIXES

APPENDIX TABLES

Appendix Table 1: Standard Population Age Structure*

Age	Population
0-4	281,600
5-9	265,400
10-14	260,000
15-19	281,600
20-24	281,800
25-29	286,000
30-34	288,600
35-39	259,800
40-44	238,000
45-49	206,400
50-54	169,400
55-59	148,600
60-64	141,200
65-69	129,600
70-74	103,800
75-79	77,400
80-84	52,400
85+	34,000
All ages	3,505,600

*1991 *de facto* total New Zealand population, interpolated from 1990 and 1995.
Source: United Nations Population Division.

APPENDIXES

Appendix Table 2: Availability of Variables in the Mortality dataset by Year

		2006	2005	2004	2003	2002	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988
Original Data	regyear_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	enc_nhi	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	dob_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	age_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	cob_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	yrsinNZ_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	occupation_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicgp_m	y	y	y	y	y	y	y	y	y		y	y	y	y									
	gender_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg1_m	y	y	y	y	y	y	y	y	y						y	y	y	y	y	y	y	y	y
	ethnicg2_m	y	y	y	y	y	y	y	y	y						y	y	y	y	y	y	y	y	y
	ethnicg3_m	y	y	y	y	y	y	y	y	y						y	y	y	y	y	y	y	y	y
	dom_cd_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	dod_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	icd01_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	icd02_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	icd03_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	death_certifier_m	y	y	y	y	y	y	y	y	y														
	death_info_src_m	y	y	y	y	y	y	y	y	y														
NZ*dep*	Score_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	Scale_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y
	CAU_m	y	y	y	y	y	y	y	y	y		y	y	y	y	y	y	y	y	y	y	y	y	y

Appendix Table 2 Continued…

		1987	1986	1985	1984	1983	1982	1981	1980	1979	1978	1977	1976	1975	1974
Original Data	regyear_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	enc_nhi	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	dob_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	age_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	cob_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	yrsinNZ_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	occupation_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicgp_m														
	gender_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg1_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg2_m														
	ethnicg3_m														
	dom_cd_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	dod_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	icd01_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	icd02_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	icd03_m	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	death_certifier_m														
	death_info_src_m														
NZ*dep*	Score_m														
	Scale_m														
	CAU_m														

Appendix Table 3: Availability of Variables in the Hospital dataset by Year

		2008*	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988
Original Data	year_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_id_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	enc_nhi	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	dob_date_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	dob_year_h																					
	NZ_res_h	y	y	y	y	y	y	y				y	y	y	y	y	y	y	y	y	y	y
	age_dis_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicgp_h	y	y	y	y	y	y	y	y	y	y											
	gender_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg1_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg2_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg3_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	dom_code_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	nhi_dod_h	y	y	y	y	y	y	y														
	adm_src_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	adm_type_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	end_type_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_type_h	y	y	y	y	y	y	y					y	y	y	y	y	y	y	y	y	y
	event_end_date_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_start_date_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	fac_type_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_lvd_h	y	y	y	y	y	y	y														
	los_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	diag01_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	diag02_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	diag03_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	op01_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	op02_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	op03_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ecode01_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ecode02_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ecode03_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	accflag01_h	y	y	y	y	y	y	y														
Filter Variables	mh_flag_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	dss_flag_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	respite_flag_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	rehab_flag_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ungroupable_flag_h	y	y	y	y	y	y	y				y	y	y	y	y	y	y	y	y	y	y
	boarder_flag_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	well_babies_flag_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	short_stay_flag_h	y	y	y	y	y	y	y														
	non_che_flag_h								y	y	y	y	y	y	y	y	y	y	y	y	y	y
	starship_flag_h																					
NZ*dep*	score_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	scale_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	CAU_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y

*data only available up to 30th June 2008

Appendix Table 3 Continued…

		1987	1986	1985	1984	1983	1982	1981	1980	1979	1978	1977	1976	1975	1974
Original Data	year_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_id_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	enc_nhi														
	dob_date_h														
	dob_year_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	NZ_res_h	y	y	y	y	y	y	y	y						
	age_dis_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicgp_h														
	gender_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg1_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ethnicg2_h														
	ethnicg3_h														
	dom_code_h	y	y	y	y	y	y	y	y	y	y				
	nhi_dod_h														
	adm_src_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	adm_type_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	end_type_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_type_h														
	event_end_date_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	event_start_date_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	fac_type_h														
	event_lvd_h														
	los_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	diag01_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	diag02_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	diag03_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	op01_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	op02_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	op03_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ecode01_h	y	y	y	y	y	y	y	y	y	y	y	y	y	y
	ecode02_h														
	ecode03_h														
	accflag01_h														
Filter Variables	mh_flag_h	y	y	y	y	y	y	y	y						
	dss_flag_h	y	y	y	y	y	y	y	y						
	respite_flag_h	y	y	y	y	y	y	y	y						
	rehab_flag_h	y	y	y	y	y	y	y	y						
	ungroupable_flag_h	y	y	y	y	y	y	y	y						
	boarder_flag_h	y	y	y	y	y	y	y	y						
	well_babies_flag_h														
	short_stay_flag_h														
	non_che_flag_h														
	starship_flag_h														
NZ*dep*	score_h														
	scale_h														
	CAU_h														

*data only available up to 30th June 2008