Longitudinal Data Analysis: statistical methods for analysing longitudinal changes in health related quality of life which account for deaths and impute for longitudinal missing data.

Steven J Bowe
Master of Medical Statistics,
Bachelor of Education (mathematics)

Thesis submitted for fulfilment of the award of Doctor of Philosophy
Faculty of Health
The University of Newcastle

Centre for Clinical Epidemiology and Biostatistics, October 2010
This thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. I give consent to this copy of my thesis, when deposited in the University Library**, being made available for loan and photocopying subject to the provisions of the Copyright Act 1968.

**Unless an Embargo has been approved for a determined period.
Dedicated to Frankie Bowe
(4-2-1936 to 16-2-2008)
&
Leo Bowe
Born 10-03-2009

In memory of my father who passed away while I was completing this thesis.
Dearly loved and sadly missed.
He always encouraged me to see things differently and he would be very proud of the
“doctor” as he always called me from childhood.
I would also like to dedicate this thesis to my beautiful nephew Leo who has brought so
much joy to my family since the passing of his grandfather.
Acknowledgements

First and foremost, I would like to acknowledge and thank my PhD supervisors, Associate Professor David Sibbritt, Dr Patrick McElduff and Associate Professor Anne Young who have provided invaluable guidance and support along what has been a long journey.

This thesis would not have been possible had it not been for the use of the data from Australian Longitudinal Study on Women’s Health (ALSWH) at the Universities of Newcastle and Queensland. I would like to thank the very helpful and time generous researchers and staff at ALSWH. I would also like to thank Professor Gita Mishra for encouraging me to tackle this topic and thank you to Professor Annette Dobson and Professor Julie Byles for their valuable advice along the way. Big thanks to Dr Adrian Barnett and Dr Hiro Furuya for their valuable input into parts of my research.

Thanks to my colleagues and the staff at The Centre for Clinical Epidemiology and Biostatistics both past and present, who have encouraged me to finish over the past six and half years (part-time). I would especially like to acknowledge Professor Catherine D’Este and Dr Kerry Inder who have both been very supportive.

To my wonderful friends, who have supported and encouraged me through an extremely difficult period in my life.

Finally, I would like to thank my mum Helen and my sisters Julie, Louise and Katie for their encouragement and support through what has been a very tough time for us all.
PUBLICATIONS AND PRESENTATIONS ARISING FROM AND CONTRIBUTING TO THIS THESIS

Published Papers

Bowe SJ, Young AF, Sibbritt DW & Furuya H (2006) Transforming the SF-36 to account for death in longitudinal studies with three year follow-up Med Care 44: 956-959.
This paper is based on the contents in Chapter 4 (see Appendix 8.1 for paper).

The statistical methods used in this paper have contributed to the methods applied in Chapters 6 and 7 (see Appendix 8.1 for paper).

Conference Presentations

This conference presentation is based on the contents in Chapter 5.


Bowe SJ, Young AF, Sibbritt DW, & Furuya H. Transforming the SF-36 to account for death in longitudinal studies: does length of time between surveys and cohorts composition matter? Australasian Epidemiological Association Annual Conference, Newcastle, Australia October 2005 Oral presentation.
# TABLE OF CONTENTS

**LIST OF FIGURES** ........................................................................................................ VI  
**LIST OF TABLES** ........................................................................................................ VIII  
**SYNOPSIS** ................................................................................................................ IX  

## CHAPTER 1: ................................................................................................................ 1  
1.0 INTRODUCTION ........................................................................................................ 3  
1.1 THESIS AIM AND OBJECTIVES .............................................................................. 4  
1.2 OUTLINE OF THESIS ............................................................................................... 5  

## CHAPTER 2: ................................................................................................................ 7  
2.0 INTRODUCTION ........................................................................................................ 10  
2.1 A COMPARISON OF HEALTH RELATED — QUALITY OF LIFE INSTRUMENTS ....... 11  
2.2 HEALTH PROFILES VS HEALTH UTILITY INDEXES ............................................. 13  
2.3 GENERIC MEASURES AND DISEASE SPECIFIC MEASURES .............................. 14  
2.3.1 Examples of Generic Measures ........................................................................... 14  
2.3.2 Examples of Disease-Specific Measures ............................................................ 14  
2.4 A BRIEF HISTORY OF GENERIC HEALTH PROFILES USING SUMMATED RATING SCALES .................................................. 15  
2.4.1 MOS Short Form-36 (SF-36) .............................................................................. 15  
2.4.2 RAND-36 Item Health Survey ............................................................................ 20  
2.4.3 Developments of SF-12 and VR-12 from SF-36 and VR-36 ............................... 21  
2.4.4 The emergence of SF-12 as a possible replacement of SF-36 .............................. 21  
2.4.5 The Veteran RAND-12 (VR-12) ........................................................................ 22  
2.4.6 Short Form-12 version 2 (SF-12v2) .................................................................... 22  
2.4.7 SF-8 Health Survey ............................................................................................ 23  
2.5 WORLD HEALTH ORGANISATION QUALITY OF LIFE (WHOQOL-BREF) ........... 23  
2.6 BRIEF HISTORY OF GENERIC HEALTH UTILITY INDEXES ............................. 24  
2.6.1 Direct and Indirect Utility Techniques .................................................................. 25  
2.6.2 Indirect utility techniques .................................................................................... 26  
2.7 MAPPING FROM THE SF-36 AND SF-12 TO THE SF6D ..................................... 30  
2.8 USAGE OF TRANSLATING NINE GENERIC HRQOL INSTRUMENTS FOR USE IN AFRICA .................................................. 32  
2.9 DEVELOPMENT OF THE AUSTRALIAN STANDARDISED SF-36 ............................ 33  
2.9.1 THE AUSTRALIAN LONGITUDINAL STUDY ON WOMEN’S HEALTH (ALSWH) ........................................................................... 34  
2.9.2 ALSWH PCS and MCS standardised using a formula relevant to Australian population .................................................................................. 35  
2.9.3 Measures ........................................................................................................... 35  
2.9.4 Analysis ............................................................................................................. 36  
2.9.5 Norms of the PCS and MCS scores for older Australian women ........................ 36  
2.10 VALIDITY OF THE SF-12 COMPARED WITH THE SF-36 IN ALSWH PILOT STUDIES .... 38  

## CHAPTER 3: ................................................................................................................ 39  
3.0 INTRODUCTION ........................................................................................................ 41  
3.1 CONCEPTS AND TERMINOLOGY OF MISSING DATA .......................................... 41  
3.1.1 Theories and Concepts ....................................................................................... 41  
3.1.2 Attrition in longitudinal studies ....................................................................... 46  
3.1.3 Dealing with Missing Data ............................................................................... 57  
3.1.4 Developments in Imputation Theory and Methodology ..................................... 59  
3.1.5 Single imputation – longitudinal methods (LVCF) ............................................ 63  
3.1.6 Weighting Methods ........................................................................................... 63  
3.1.7 Multiple imputation ......................................................................................... 64
CHAPTER 6: ................................................................................................................. 165

6.0 INTRODUCTION ........................................................................................................ 167
6.1 BACKGROUND .............................................................................................................. 167
  6.1.1 HAlex Approach to account for death ................................................................... 168
  6.1.2 Estimation of Health Utility Index scores .............................................................. 169
  6.1.3 Pros and Cons of the HALEx Approach ................................................................. 173
  6.1.4 Usage of Other Health Utility Indexes ................................................................. 174
  6.1.5 Using another Health Utility Index- Converting SF-36 to SF-6D ......................... 174
  6.1.6 Further Developments of the SF-6D health utility index ........................................ 175
  6.1.7 Death as a changing Quality of Life value ............................................................ 176
6.2 AIMS AND OBJECTIVES .......................................................................................... 178
6.3 METHODS .................................................................................................................. 179
  6.3.1 Subjects .................................................................................................................. 179
  6.3.2 Measures ............................................................................................................... 179
  6.3.3 Development of the Outcome Variable PCSD ...................................................... 184
  6.3.4 Covariates .............................................................................................................. 186
6.4 STATISTICAL ANALYSIS ......................................................................................... 186
  6.4.1 Developments of SF-6D and PCS death value ..................................................... 186
  6.4.2 Cross-sectional Analysis ....................................................................................... 186
  6.4.3 Longitudinal Data Analysis ................................................................................. 187
6.5 RESULTS ..................................................................................................................... 195
  6.5.1 Descriptive Statistics – SF-6D at Surveys 1-4 ....................................................... 195
  6.5.2 Correlations between PCS and SF-6D at Survey 1-4 ............................................ 196
  6.5.3 Comparing Results from R-squared (R²) and Y-intercepts (α) and beta (B) estimates .............................................................. 197
  6.5.4 Calculation of the PCS Death scores using a Linear Equation at each survey ......... 205
  6.5.5 Considerations of the PCS value for death ............................................................ 205
  6.5.6 Comparisons using general Australian populations aged 65 years older .............. 209
  6.5.7 What PCS value then to use for death? ................................................................. 211
  6.5.8 Cross-Sectional Data Analysis ......................................................................... 214
  6.5.9 Longitudinal Data Analysis ................................................................................. 220
6.6 DISCUSSION .............................................................................................................. 228
  6.6.1 Cross-Sectional comparisons of death values ...................................................... 228
  6.6.2 Comparisons to Trisolini et al.'s PCS value for death of 14 ................................. 228
  6.6.3 The Use of Decreasing PCS Values for Death (7, 5, 5) ........................................ 229
  6.6.4 The Impact of Using of Multiple Imputation ....................................................... 229

CHAPTER 7: ...................................................................................................................... 231

7.0 INTRODUCTION ...................................................................................................... 233
7.1 BACKGROUND ......................................................................................................... 233
  7.1.1 Original and Current HOS Case-Mix methodology .............................................. 235
List of Figures

Figure 2.1  Health Related Quality of Life: Taxonomy ........................................... 12
Figure 2.2  SF-36® Scales Measure Physical and Mental Components of Health ............ 17
Figure 2.3  Direct vs Indirect Techniques according to Brazier et al (1999) and Arnold et al (2009) .... 25
Figure 3.1  Illustration of intermittent missing data and drop-outs .................................. 48
Figure 3.2  Example of the Missing Data Mechanisms ................................................ 53
Figure 4.1  Mean PCS scores at Surveys 1 and 2 by General Health Classification at Survey 3 ........... 112
Figure 4.2  A Comparison of Equation 4.1 and ALSWH Equation .................................. 120
Figure 4.3  Predicted percent Healthy versus observed percent Healthy at Survey 2 .................. 122
Figure 4.4  Observed percent ‘healthy’ (at Survey 3) vs predicted percent ‘healthy’ .................. 125
Figure 5.1  Mean APCT and Reason for Attrition between Survey 2 and 3 by Diabetes Groups .......... 156
Figure 5.2  APCT – Mean Probability of being healthy in 3 years (Deaths excluded), with 95% confidence intervals .................................................................................. 157
Figure 5.3  APCTD – Mean Probability of being healthy in 3 years (Complete case Deaths included), with 95% confidence intervals ........................................................................ 158
Figure 5.4  APCTDI – Mean Probability of being healthy in 3 years (Deaths and imputed values included), with 95% confidence intervals ................................................................ 159
Figure 6.1  Erickson (1998) Original matrix for scoring the HALEX .................................. 171
Figure 6.2  Trisolini et al matrix for scoring the HALEX Utility Scores by Activity Limitation and Perceived Health Utility Scores .............................................................................. 171
Figure 6.3  Regression results from Smith, et al (2005) Conference Presentation .................. 172
Figure 6.4  The SF-36 items and sub-scales used for PCS and MCS ................................... 180
Figure 6.5  Regression results of PCS on SF-6D at survey 1 using linear, logarithmic and polynomial regression models ................................................................. 199
Figure 6.6  Regression results of PCS on SF-6D at survey 2 using linear, logarithm and polynomial regression models .................................................................................. 200
Figure 6.7  Regression results of PCS on SF-6D at survey 3 using linear, logarithm and polynomial regression models .................................................................................. 201
Figure 6.8  Regression results of PCS on SF-6D at survey 4 using linear, logarithm and polynomial regression models .................................................................................. 202
Figure 6.9  Regression results of PCS on SF-6D at survey 1 using the Stevenson (1996) subscale weightings for linear, logarithm and polynomial regression models .................. 210
Figure 6.10 Cross-sectional comparisons of PCS at Survey 1, 2, 3 and 4 ............................ 216
Figure 6.11 Cross-sectional comparisons of PCS9D at Survey 1, 2, 3 and 4 .......................... 218
Figure 6.12 Cross-sectional comparisons of PCS9D at Survey 1, 2, 3 and 4 .......................... 220
Figure 6.13 Longitudinal Analysis – Mean PCS when death=9 but without other imputation of PCS ...... 223
Figure 7.1  Difference in PCS before Multiple Imputation ............................................... 270
Figure 7.2  Difference in PCS after Multiple Imputation (5 datasets) .................................... 271
Figure 7.3  Difference in PCS before Multiple Imputation ............................................... 270
Figure 7.4  ALSWH Mean Probability of being Alive and in Same or Better Health at each Survey by Diabetes Groups without deaths ......................................................... 285
Figure 7.5  ALSWH Mean Probability of being Alive and in Same or Better Health at each Survey by Diabetes Groups with deaths ......................................................... 285
Figure 7.6  Adjusted mean probabilities of being Alive and in Same or Better Health over time (No deaths and no interaction) ................................................................. 289
Figure 7.7  Adjusted mean probabilities of being Alive and in Same or Better Health over time (No deaths but an interaction) ................................................................. 289
Figure 7.8  Adjusted mean probabilities of being Alive and in Same or Better Health over time (Deaths included but no interaction) ................................................................. 291
Figure 7.9  Adjusted mean probabilities of being alive and in same or better health over time (Deaths included and time interaction) ........................................................................................................... 296
Figure 7.10  Mean probability of being alive and in same or better health over time (with deaths included and imputed APASBD no interaction) ................................................................. 298
Figure 7.11  Mean probability of being alive and in same health or better health over time by diabetes groups (Deaths and imputed APASBD and an interaction) ................................................. 299
Figure 7.12  Adjusted mean probability of being alive and in same or better health over time (Deaths and no imputed APASBD and no interaction) .......................................................... 296
Figure 7.13  Adjusted mean probability of being alive and in same or better health over time (Deaths and imputed APASBD and no interaction) ............................................................... 297
Figure 7.14  Mean probability of being alive and better health at each survey by diabetes groups with deaths included .............................................................................................................. 298
Figure 7.15  Mean probability of being alive and same health at each survey by diabetes groups with (deaths included) ........................................................................................................ 299
Figure 7.16  Adjusted probability of being alive and better health at each survey by diabetes groups with deaths included with multiple imputation ......................................................................... 301
Figure 7.17  Mean probability of being alive and better health at each survey by diabetes groups with deaths included and time interaction with MI ........................................................... 301
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>NORMS FOR PCS AND MCS SCORES FOR THREE GROUPS OF AUSTRALIAN WOMEN COMPARED WITH US NORMS FOR FEMALES</td>
<td>37</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>EIGHT STRATEGIES OF HANDLING DEATHS SUGGESTED BY DHEHR ET AL (1995)</td>
<td>73</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>RESPONSE, RETENTION RATES AND DEATHS OF OLDER COHORT</td>
<td>110</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>DESCRIPTIVE STATISTICS – COMPARISONS OF EQUATIONS AT SURVEYS 1-3</td>
<td>119</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>PREDICTED PERCENT HEALTHY VS OBSERVED PERCENT HEALTHY AT SURVEY 2</td>
<td>121</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>OBSERVED PERCENT ‘HEALTHY’ AFTER 3 YEARS (AT SURVEY 3) VS PREDICTED PERCENT ’HEALTH’</td>
<td>124</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>CHANGES IN HEALTH BETWEEN SURVEYS 1 AND 2 FOR DIABETES GROUPS</td>
<td>126</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>COMPARISON OF MEAN SCORES BETWEEN SURVEY 1 AND SURVEY 2</td>
<td>126</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>DIABETES STATUS DEFINED RETROSPECTIVELY FROM 2005</td>
<td>153</td>
</tr>
<tr>
<td>Table 5.2A</td>
<td>OVERALL MORTALITY AND WITHDRAWALS</td>
<td>153</td>
</tr>
<tr>
<td>Table 5.2B</td>
<td>CUMULATIVE NUMBER (%) OF DEATHS BY DIABETES STATUS</td>
<td>154</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>PROBABILITY OF BEING HEALTHY IN 3 YEARS BY ATTRITION CLASSIFICATIONS FOR EXISTING CASES</td>
<td>155</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>COMPARISONS OF REGRESSION COEFFICIENTS FOR PREDICTING PCSDT</td>
<td>161</td>
</tr>
<tr>
<td>Table 6.1</td>
<td>DESCRIPTIVE STATISTICS – SF-6D AT SURVEYS 1,2, 3 AND 4</td>
<td>195</td>
</tr>
<tr>
<td>Table 6.2</td>
<td>CORRELATIONS FOR PCS AND SF-6D AT SURVEYS 1, 2, 3 AND 4</td>
<td>196</td>
</tr>
<tr>
<td>Table 6.3</td>
<td>COMPARISONS OF LINEAR, LOGARITHMIC AND POLYNOMIAL REGRESSION ESTIMATES FOR SURVEY 1</td>
<td>198</td>
</tr>
<tr>
<td>Table 6.4</td>
<td>APPROACHES FOR IMPUTING PCS VALUES FOR DEATH</td>
<td>212</td>
</tr>
<tr>
<td>Table 6.5</td>
<td>DESCRIPTIVE STATISTICS OF APPROACH 1 – COMPARISONS OF PCS AND PCSD AT SURVEYS 1-4</td>
<td>213</td>
</tr>
<tr>
<td>Table 6.6</td>
<td>CROSS-SECTIONAL COMPARISONS OF PCS AT SURVEY 1, 2, 3 AND 4</td>
<td>215</td>
</tr>
<tr>
<td>Table 6.7</td>
<td>CROSS-SECTIONAL COMPARISONS OF PCSD9 (DEATH = 9) AND SURVEYS 1-4</td>
<td>217</td>
</tr>
<tr>
<td>Table 6.8</td>
<td>CROSS-SECTIONAL COMPARISONS OF PCSD7, 5, 5 AT SURVEYS 1-4</td>
<td>219</td>
</tr>
<tr>
<td>Table 6.9</td>
<td>COMPARISONS OF REGRESSION COEFFICIENTS USING WINBUGS AND STATA (PCSD9)</td>
<td>226</td>
</tr>
<tr>
<td>Table 6.10</td>
<td>COMPARISONS OF REGRESSION COEFFICIENTS USING WINBUGS AND STATA (PCSD7-5-5)</td>
<td>227</td>
</tr>
<tr>
<td>Table 7.1</td>
<td>MISSING DATA - PROPORTION OF MISSING COVARIATES AT EACH SURVEY FOR ORIGINAL COHORT</td>
<td>264</td>
</tr>
<tr>
<td>Table 7.2</td>
<td>NO MULTIPLE IMPUTATION (N=7010)</td>
<td>266</td>
</tr>
<tr>
<td>Table 7.3</td>
<td>MULTIPLE IMPUTATION “AVERAGED COEFFICIENTS” (N=8790)</td>
<td>267</td>
</tr>
<tr>
<td>Table 7.4</td>
<td>DIFFERENCE IN PCS BEFORE AND AFTER IMPUTATION</td>
<td>269</td>
</tr>
<tr>
<td>Table 7.5</td>
<td>ALSWH AND ROGERS’ CUT-OFFS (SAME OR BETTER VS WORSE)</td>
<td>270</td>
</tr>
<tr>
<td>Table 7.6</td>
<td>LOGISTIC REGRESSION MODEL FOR SAME OR BETTER NO MULTIPLE IMPUTATION</td>
<td>272</td>
</tr>
<tr>
<td>Table 7.7</td>
<td>LOGISTIC REGRESSION MODEL (SAME OR BETTER) - MI “AVERAGED COEFFICIENTS”</td>
<td>272</td>
</tr>
<tr>
<td>Table 7.8</td>
<td>USING ROGERS’ CUT-OFFS (5.66) MULTIPLE IMPUTATION</td>
<td>274</td>
</tr>
<tr>
<td>Table 7.9</td>
<td>DESCRIPTIVE STATISTICS – ALSWH CUT-OFF (8.65) APASB AT SURVEYS 1-4</td>
<td>276</td>
</tr>
<tr>
<td>Table 7.10</td>
<td>DESCRIPTIVE STATISTICS – ALSWH CUT-OFF (8.65) APASD AT SURVEYS 1-4</td>
<td>277</td>
</tr>
<tr>
<td>Table 7.11</td>
<td>DESCRIPTIVE STATISTICS - ROGERS’ CUT-OFF (5.66) PASS AT SURVEYS 1-4</td>
<td>277</td>
</tr>
<tr>
<td>Table 7.12</td>
<td>DESCRIPTIVE STATISTICS - ROGERS’ CUT-OFF (5.66) PASSB AT SURVEYS 1-4</td>
<td>278</td>
</tr>
<tr>
<td>Table 7.13</td>
<td>ALSWH AND ROGERS’ CUT-OFFS (BETTER VS SAME VS WORSE)</td>
<td>280</td>
</tr>
<tr>
<td>Table 7.14</td>
<td>PROBABILITY OF BEING ALIVE AND BETTER (WITH AND WITHOUT DEATHS) AT NEXT SURVEY</td>
<td>282</td>
</tr>
<tr>
<td>Table 7.15</td>
<td>PROBABILITY OF BEING ALIVE AND SAME (WITH AND WITHOUT DEATHS) AT NEXT SURVEY</td>
<td>282</td>
</tr>
<tr>
<td>Table 7.16</td>
<td>“CHANGE IN HEALTH” APASB (NO DEATHS) FOR DIABETES GROUPS AT SURVEYS 1 TO 4</td>
<td>284</td>
</tr>
<tr>
<td>Table 7.17</td>
<td>“CHANGE IN HEALTH” APASBD (DEATHS INCLUDED) FOR DIABETES AT SURVEYS1 TO 4</td>
<td>286</td>
</tr>
<tr>
<td>Table 7.18</td>
<td>(SAME OR BETTER) RANDOM INTERCEPT MODEL NO DEATHS (NO MI) AND WITH A TIME INTERACTION</td>
<td>288</td>
</tr>
<tr>
<td>Table 7.19</td>
<td>(SAME OR BETTER) RANDOM INTERCEPT MODEL (DEATHS) WITHOUT MULTIPLE IMPUTATION</td>
<td>290</td>
</tr>
<tr>
<td>Table 7.20</td>
<td>(SAME OR BETTER) RANDOM INTERCEPT MODEL (DEATHS) WITH MULTIPLE IMPUTATION OF APASBD</td>
<td>292</td>
</tr>
<tr>
<td>Table 7.21</td>
<td>(SAME OR BETTER) RANDOM SLOPES AND INTERCEPTS (DEATHS) WITHOUT MI</td>
<td>294</td>
</tr>
<tr>
<td>Table 7.22</td>
<td>(SAME OR BETTER) RANDOM INTERCEPTS AND SLOPES (DEATHS) (WITH MI)</td>
<td>295</td>
</tr>
<tr>
<td>Table 7.23</td>
<td>(BETTER) RANDOM INTERCEPT MODEL (DEATHS) WITH AND WITHOUT MULTIPLE IMPUTATION</td>
<td>302</td>
</tr>
</tbody>
</table>
SYNOPSIS

Analysis of data from longitudinal studies is made more complex by the death of study participants over time. Many statistical methods depend on complete case analysis, meaning that data for participants who die are often removed from the analysis and reported and/or sometimes analysed separately. This complete case analysis approach limits longitudinal analysis to survivors, who often begin or remain in better health than those who died, and hence researchers may miss important changes over time in the total cohort.

Many longitudinal studies that aim to measure changes in physical and mental health over time use the SF-36 instrument, a 36-item health questionnaire. However, there is no score on the SF-36 to reflect death. In recent years various methods to assign a score for death have been suggested but in most cases need greater validation and development across a variety of longitudinal studies.

This thesis discusses four methods for analysing longitudinal changes in health related quality of life; three methods for incorporating death into longitudinal studies of elderly populations and another method that attempts to deal with longitudinal missing data which may be missing not at random:

- Method 1 - Transforming the SF-36 Physical Component Summary score (PCS) to a probability of being healthy in three years;
- Method 2 - Converting the Short Form - 36 health survey into the Short Form - 6 Domains (SF-6D) to extrapolate a PCS value for death;
- Method 3 - Adapting the Health Outcomes Study (HOS) case-mix method to predict the probability of being Alive and in Same/Better health in 3 years;
Method 4 - Longitudinal multiple imputation approach using Fully Bayesian methods.

The four methods were applied to data obtained from the Australian Longitudinal Study on Women’s Health (ALSWH). This is a longitudinal, population-based survey that examines the health of three large cohorts of community-dwelling Australian women over a 20 year period (1996-2016) with follow-up surveys every three years. To demonstrate each method, a case study was used to determine whether or not there were statistically significant differences in elderly women with and without diabetes over time. The main focus was the impact of including deaths and other missing data over time. The analyses found that the inclusion of those who died impacted more heavily on those women with diabetes. The difference in predicted health in 3 years for women with and without diabetes became greater over time when deaths were reinstated with a value. Longitudinal multiple imputation of intermittent missing outcome and covariate data reduced the impact of the deaths on both groups over time. Finally, all four methods suggest that ignoring deaths and other missing data can lead to biased results towards study survivors.

The recommendation is that researchers using the SF-36 to study the longitudinal change in quality of life of elderly populations over time need to consider using methods which account for deaths and other missing data. These results can then be compared to analyses which ignore deaths and other missing data. This will result in less biased findings from longitudinal studies of ageing populations.
CHAPTER 1:

Introduction- Outline of Thesis
List of Abbreviations used in Chapter 1

ALSWH       Australian Longitudinal Study on Women’s Health
HR-QOL     Health Related - Quality of Life
MOS SF-36   Medical Outcomes Study Short Form - 36
SF-36       Short Form -36
1.0 INTRODUCTION

The Medical Outcomes Study Short-Form 36 (SF-36) (Ware 1993) is a commonly used tool for measuring changes in health-related quality of life (HR-QOL) over time. It has been well documented that a limitation of the SF-36 is that it only considers morbidity and not mortality (Diehr et al 1995; Diehr et al 2007). When analysing longitudinal data, researchers typically ignore participants who die and analyse complete case data, effectively limiting longitudinal analysis to survivors, so that important changes over time in the total cohort may be missed (Diehr et al 2003). To ignore deaths could suggest that the results are only able to be generalised to those who have survived. There is also a possibility of a failure to detect true differences in changes in HR-QOL over time between groups of people. According to Selim et al (2006), accurate information on health outcomes has become an expectation of regulatory and accreditation agencies and important decisions, such as reimbursement and accreditations, will be based on perceived performances. Recently, a number of approaches to incorporate mortality have been developed by Diehr et al (2007) then adopted and/or adapted in other studies. However, none of these approaches have adequately adjusted for missing values caused by reasons other than death.

Missing data are a very common problem confronting researchers, especially in longitudinal studies. Rubin (1976) first discussed the concept of “missingness” (Rubin 1976) and since then there has been ongoing discussion of the definitions and patterns of missingness. This thesis will attempt to explore possible methods which may be useful to account for deaths in longitudinal studies of elderly populations and also consider what should be done when there are large amounts of missing outcome data. To explore these issues of accounting for deaths and other missing data a case study
will be conducted using current longitudinal data analysis methods which will compare changes in health-related quality of life for older women with and without diabetes.

1.1 THESIS AIM AND OBJECTIVES

- To investigate the statistical methods used to account for death in longitudinal studies.

- To determine whether there is a need to improve current statistical methods and apply and assess new strategies if applicable.

- To investigate the current methods for accounting for missing data in longitudinal studies.

- Apply the current statistical methods to the older cohort of Australian Longitudinal Study on Women’s Health (ALSWH) data and then evaluate the advantages and disadvantages of those methods.

- To examine the impact of diabetes on quality of life among older women - adjusting for deaths by applying the methods developed.
1.2 **Outline of Thesis**

This thesis consists of eight chapters. This chapter provides an outline of the aims and content of this thesis.

**Chapter 2** will outline and describe key health related quality of life measurement instruments which are currently being applied to longitudinal studies. This chapter will also highlight the changes to a HR-QOL instrument known as the Short Form-36 questionnaire. This questionnaire is an integral part of the Australian Longitudinal Study on Women's Health. This chapter will describe the Australian Longitudinal Study on Women's Health and explain how the older cohort of this large study will used to apply the methodology developed throughout this thesis.

**Chapter 3** will include two main sections. The first section will describe important concepts related to missing data and how missing data impact on longitudinal studies. There will be discussion as to current approaches to dealing with missing data. The second section will discuss the results of two literature reviews. The first literature review examines the current approaches to dealing with deaths in longitudinal studies of ageing populations. Greater attention is given to those studies which are using the SF-36 questionnaire. The second literature review will consider methods that are being applied to deal with other missing data in longitudinal studies.

**Chapter 4** focuses on the development of a statistical method to account for deaths in longitudinal studies. This first method involves the transformation of a physical health summary score to a probability of being healthy in 3 years. The method is applied and validated using the ALSWH data from Surveys 1-3.
Chapter 5 extends on the work of Chapter 4. In this chapter a method is developed which allows for multiple imputation of longitudinal missing data using fully Bayesian methods. The random effects models developed allow for the missingness assumptions of ignorable and informative missing data. The models are used in a case study of diabetes to determine the impact of missing outcomes after accounting for deaths.

Chapter 6 involves the development of another method to account for deaths in longitudinal studies of elderly populations. The second method is adapted from a method which is yet to be validated. This approach extrapolates a value for death by considering a relationship between health related quality of life measurements and health utility index scores. The value for death is then imputed and longitudinal analysis is conducted using a case study of diabetes participants in the ALSWH data. Further analysis is conducted which considers the missingness of the outcome.

Chapter 7 adapts a third approach to account for deaths in a longitudinal study. This final method involves the conditional probability of being alive and in same or better health in the future. The approach accounts for deaths by reinstating those who died with a value of zero. The impact of imputing those who have died is evaluated through longitudinal data analysis which also accounts for missing data which maybe missing at random or missing not at random. Also, in this chapter different health status cut-offs for better vs same vs worse are considered to calculate the conditional probability of being alive and in same or better health in the future.

Chapter 8 summarises strengths and limitations of the four methods, makes recommendations about each method and suggests future research and further applications of the methods.
CHAPTER 2:

Background –

The Development of Health Related Quality of Life measurement tools
List of Abbreviations used in Chapter 2

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>Australian Bureau of Statistics</td>
</tr>
<tr>
<td>ALSWH</td>
<td>Australian Longitudinal Study on Women’s Health</td>
</tr>
<tr>
<td>AQLQ</td>
<td>Asthma Quality of Life Questionnaire</td>
</tr>
<tr>
<td>CATI</td>
<td>Computer Assisted Telephone Interview</td>
</tr>
<tr>
<td>DQOL</td>
<td>The Diabetes Quality of Life measure</td>
</tr>
<tr>
<td>DTSQ</td>
<td>Diabetes Treatment Satisfaction Questionnaire now - DTSQc</td>
</tr>
<tr>
<td>EORTC - QLQC30</td>
<td>European Organisation for Research and Treatment of Cancer - Quality of Life Questionnaire C30</td>
</tr>
<tr>
<td>EQ-5D</td>
<td>European Quality of Life- 5 Domains</td>
</tr>
<tr>
<td>EVGGFP</td>
<td>Excellent, Very Good, Good, Fair or Poor</td>
</tr>
<tr>
<td>HAQ</td>
<td>Health Assessment Questionnaire</td>
</tr>
<tr>
<td>HR-QOL</td>
<td>Health Related Quality of Life</td>
</tr>
<tr>
<td>HUI2</td>
<td>Health Utilities Index Mark 2</td>
</tr>
<tr>
<td>HUI3</td>
<td>Health Utilities Index Mark 3</td>
</tr>
<tr>
<td>MCS</td>
<td>Mental Component Summary score (derived from SF-36)</td>
</tr>
<tr>
<td>MOS SF-36</td>
<td>Medical Outcomes Study Short Form 36</td>
</tr>
<tr>
<td>PAID</td>
<td>Problem Areas in Diabetes</td>
</tr>
<tr>
<td>PCS</td>
<td>Physical Component Summary score (derived from SF-36)</td>
</tr>
<tr>
<td>PROQOLID</td>
<td>Patient Reported Outcome and Quality of Life Instruments Database</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality Adjusted Life Years</td>
</tr>
<tr>
<td>QOLID</td>
<td>Quality of Life Instruments Database</td>
</tr>
<tr>
<td>QWB-SA</td>
<td>Quality of Well-Being scale (Self-Administered form)</td>
</tr>
<tr>
<td>RAND SF-36</td>
<td>Research and Development Short Form 36</td>
</tr>
<tr>
<td>RAQoL</td>
<td>Rheumatoid Arthritis Quality of Life Questionnaire</td>
</tr>
<tr>
<td>SF-6D</td>
<td>Short Form 6 Domains</td>
</tr>
<tr>
<td>SF-8</td>
<td>Short Form 8</td>
</tr>
<tr>
<td>SF-12v1</td>
<td>Short Form 12 version 1</td>
</tr>
<tr>
<td>SF-12v2</td>
<td>Short Form 12 version 2</td>
</tr>
<tr>
<td>SF-36v1</td>
<td>Short Form 36 version 1</td>
</tr>
<tr>
<td>SF-36v2</td>
<td>Short Form 36 version 2</td>
</tr>
<tr>
<td>SG</td>
<td>Standard Gamble</td>
</tr>
<tr>
<td>TTO</td>
<td>Time Trade-Off</td>
</tr>
<tr>
<td>VAS</td>
<td>Visual Analog Scale</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>VR-12v1</td>
<td>Veterans RAND Short Form 12 version 1</td>
</tr>
<tr>
<td>VR-36</td>
<td>Veterans RAND Short Form 36</td>
</tr>
<tr>
<td>WHOQOL-BREF</td>
<td>World Health Organisation Quality of Life Brief Assessment</td>
</tr>
<tr>
<td>WHOQOL-100</td>
<td>World Health Organisation Quality of Life -100</td>
</tr>
</tbody>
</table>
2.0 INTRODUCTION

This chapter will discuss the developments of health related quality of life (HR-QOL) instruments and health utility indexes. Not all summary measures will be discussed as it would go beyond the scope of this thesis. Instead this chapter will focus on what Fryback (2003) considered to be the most useful and commonly applied summary measures. This chapter will outline the tools he has suggested with the addition of two Australian versions of a HR-QOL tool. Some of these HR-QOL tools are an introduction and background to the methodology which will be developed throughout the course of this thesis.

Over the past two decades many health measurement tools have been developed and used in many different countries. McDowell and Newell’s (1996) encyclopaedic book on health measures lists 21 instruments in its chapter on “general status and quality of life”. According to Fryback (2003) the Mapi Research Institute’s (Lyn, France) Quality of Life Instruments Database (QOLID) Web page (Feb 2003) listed 73 instruments under the rubric “generic QOL instruments”. Fryback (2003) suggests that the two lists contained only 11 common instruments. As of August 2010 there were 100 instruments listed on the now named Patient Reported Outcome and Quality of Life Instruments Database (PROQOLID) webpage: http://www.qolid.org/proqolid/search__1/generic.

Summary Measures of Health

A summary measure of health is a measure that represents the overall generic health of a person as a single value on a certain scale (Fryback 2003). The aim of health measures is to assess the quality of life of individuals. Schron and Shumaker (1992) defined quality of life as a multi-dimensional concept referring to an individual’s total well being including: psychological, social and physical health status.
The purpose of generic summary measures of quality of life should allow for comparisons across diseases, and/or persons with multiple conditions, and/or heterogeneous sub-populations (Fryback 2003). Also, summary measures of health are generally conceived as summarising morbidity, its impact on well-being, and possibly role and social functioning of an individual (Fryback 2003).

Over time two broad approaches to measuring and assess quality of life have emerged – generic and disease (illness) specific (Rubin & Peyrot 1999; Marra et al 2005). Both generic and disease-specific instruments have been developed and validated over time to enable researchers to use the instruments with confidence (Marra et al 2005). The focus in the following sections will be the generic summary measures but later in Section 2.3.2, examples of disease-specific instruments will be mentioned.

2.1 A COMPARISON OF HEALTH RELATED – QUALITY OF LIFE INSTRUMENTS

This section discusses the two main generic types of instruments for measuring Heath Related Quality of Life (HR-QOL). These two types of instruments are referred to as health profiles and utility-based instruments. This section will briefly discuss the history and development of the health profile: The Medical Outcomes Survey (MOS) SF-36 version 2.0. The MOS-SF36 has developed and changed over time and has even been developed into a health utility index (Brazier’s SF-6D). The most commonly used generic summary health profiles and health utility indexes on Fryback’s list include:

Health Profiles using summated rating scales:

1. The Medical Outcomes Study 36-Item Short Form Health Survey (SF-36)
2. The World Health Organisation WHOQOL-BREF Quality of Life Assessment
Health Indexes (preference based measures):

3. The Quality of Well-Being scale (Self-Administered form) (QWB-SA)
4. The Health Utilities Index Mark 2 and Mark 3 (HUI2/3)
5. The EuroQol EQ-5D (EQ-5D)
6. The SF-6D, a preference-based measure derived from the SF-36.

Figure 2.1 from Tsevat et al (1994) outlines, using a flowchart, the taxonomy of health related quality of life information. Figure 2.1 shows how HR-QOL can be divided into two main areas: Health status and Value or Preference-based. These two areas will be discussed further in the following sections.


Figure 2.1 Health Related Quality of Life: Taxonomy
2.2 Health Profiles vs Health Utility Indexes

Leading experts in the field of quality of life measurement suggest there is a need for a system of health indicators and measures that form a pyramid for measuring population health- a broad base pyramid, with successively higher layers, that provide a summary of detailed health information (Wolfson 2000). There has been discussion by Fryback (2003) and McDowell and Newell (1996) about what should be at the apex of this pyramid. Fryback (2003) argued there should be a summary measure, which is not disease or condition specific but instead summarises all aspects of a person’s health. McDowell and Newell (1996) suggest that the apex should be the realm of “health profiles” and “health indexes”. Fryback (2003) points out that advocates of the health profile contend that the apex is flat with no further aggregation possible or needed. Those who support the health indexes suggest that there must be an end point with a single summary score to represent overall health (Fryback 2003).

Thus there is division as to the role of health dimensions amongst advocates of the health profile and the health utility indexes. Those who support the health profiles emphasise that they provide the diverse aspects of health quality of life. They suggest that the dimensions of health should be kept separate and the measurement is only meaningful within each health domain (McDowell and Newell 1996). Whereas supporters of health indexes advocate that health has several dimensions but argue for a combination of the impressions from each dimension into an overall score (McDowell and Newell 1996).
2.3 GENERIC MEASURES AND DISEASE SPECIFIC MEASURES

2.3.1 Examples of Generic Measures

The earlier list of generic summary measures provided by Fryback (2003) is what he called (although possibly viewed by others as contentious) the “state of the art”. This list will be the focus throughout this thesis. Mostly generic measures will be discussed in Chapter 2, with a brief mention of some of the more commonly used disease-specific measures.

2.3.2 Examples of Disease-Specific Measures

The disease/illness specific quality of life measures are designed to focus on the specific problems posed by an individual disease (Rubin & Peyrot 1999). For example, Marra et al (2005) and de Jong et al (1997) report that the Rheumatoid Arthritis Quality of Life Questionnaire (RAQoL) is the first patient-completed instrument specifically designed for use with Rheumatoid Arthritis patients. Some other examples of disease-specific instruments include: the Diabetes Quality of Life measure (DQOL); the European Organisation for Research and Treatment of Cancer – (EORTC QLQC30 for cancer studies); the Asthma Quality of Life Questionnaire – (AQLQ for asthma); and the Health Assessment Questionnaire (HAQ).

Even though there are many generic and disease-specific HR-QOL measures available to researchers, there have been further developments within each specific disease (Rubin & Peyrot 1999). For example, since the introduction of the DQOL, a number of comprehensive diabetes-specific quality of life measures have been developed. These diabetes specific measures include: the Problems Areas in Diabetes (PAID) survey a measure designed to be used with either Type I or Type II diabetes (Rubin & Peyrot...
1999), as well as the Diabetes Treatment Satisfaction Questionnaire (DTSQ) now changed to DTSQc (Bradley et al 2007).

2.4 A BRIEF HISTORY OF GENERIC HEALTH PROFILES USING SUMMATED RATING SCALES

This section will discuss briefly, the history of health profiles with an emphasis on the Medical Outcome Study (MOS) SF-36. The developments will outline the changes in versions and an Australian version of the SF-36. This outline of developments was necessary to highlight a statistical method for dealing with deaths in Chapter 6, where the SF-36v1 is converted to Brazier’s SF-6D.

2.4.1 MOS Short Form-36 (SF-36)

An interest in short-form health surveys came about from the refusal of participants to complete lengthy health surveys. To reduce the loss to follow-up, Ware et al (1981) began developing a short survey that could be administered in 5 minutes by telephone. The concepts from the short-form survey eventually developed over time into what is now possibly the most well-known and commonly used health profile, the Medical Outcome Study (MOS) Short Form-36 (SF-36) version 1.0 (Kazis et al 2007). By 1997 there were some 700 publications which referenced their experiences with the SF-36 (Ware & Gandek 1998). A medline search by Fryback (2003) for “SF-36” in the title or abstract yielded over 2200 references. As recently as August 2010 a Medline search of MeSH items yielded a 4 fold increase in SF-36 references to 8750 articles. Updates of the usage and version changes to the SF-36 can be found at http://www.sf-36.com.

During the 1990s, researchers developed a number of measurement tools for use in studying the health of populations. Researchers perceived a need to widen the already
available measures of mortality and morbidity which would be used to measure the
benefits of interventions aimed at improving a wide range of dimensions (Brazier et al
1992). These dimensions would include mobility, functioning, mental health and overall
well being.

The SF-36 consists of 36 questions (items) in a self administered questionnaire asking
the respondent about various aspects of health “over the past 4 weeks” (Fryback 2003). The final 36 items chosen came from a bank of 250 items used in the
RAND health insurance experiment for assessing general health. The SF-36 was used
by the Medical Outcome Study (MOS) and is titled MOS SF-36. The MOS SF-36 does
not yield a single summary score. Instead it divides into a profile of eight scores
(vitality, general health perceptions, mental health, bodily pain, social function, role
function as limited by emotional problems, role functions as limited by physical
problems and physical function), which in turn are aggregated into two summary scores
developed using factor analysis, the mental health component score (MCS) and the
physical health component score (PCS). The SF-36 scoring is not based on preference
judgements, instead its scales are psychometric summated rating scales (Ware 1993;
Fryback 2003).

In the early development of the MOS SF-36 version 1 factor analysis confirmed that
two factors accounted for approximately 80-85% of the variance in the eight scales
(Ware & Gandek 1998). From these two factors developed the two component
summary measures known as the Physical Component Summary (PCS) and the
Mental Component Summary (MCS). To achieve these component summary measures
Ware and Gandek (1998) standardised the eight SF-36 sub-scales using means and
standard deviations from the general US population. The two aggregate PCS and MCS
scores were then standardised using a linear t score transformation (Ware & Gandek
The summary scores are standardised to a mean of 50 and standard deviation of 10, with higher scores representing better health.

Figure 2.2 illustrates the structure of the two components of health through the use of a diagram of the domains which combine to form the component summary scores. Figure 2.2 shows how the general health and vitality domains are shared by both the physical and mental components. An example of the mathematical algorithms required for generating the physical and mental summary scores can be found in Appendix 2.1.


Figure 2.2 SF-36® Scales Measure Physical and Mental Components of Health

The SF-36v1 was developed using US norms which were established in 1992 from a national survey of non-institutionalised adults (Ware 1993; Fryback 2003). In 1996 the SF-36v2 was released and has become the standard today (Ware 2000). Later in 1998 a mail-out survey was conducted to establish age and sex-specific norms for the SF-36v2 (Ware 2000; Fryback 2003). The survey involved 6 700 participants with a
response rate of 68%. The ages ranged from 18-96 years, of whom 84% were white and 80% had completed high school (Ware 2000; Fryback 2003).

This United States based instrument has also been modified for other populations within the US population and in other parts of the world. Using a population of United States war veterans the MOS SF-36 was modified and called the Veterans RAND 36 Items Health Survey (VR-36 version 1.0). A subset of 12 items was later used to shorten the VR-36v1 to the Veterans RAND 12 Item Health Survey (VR-12 version 1.0). In recent times the MOS SF-36v2 has been shortened to the SF-12v2 using a similar subset of items.

In 1991 the International Quality of Life Assessment (IQOLA) Project was established to translate the SF-36 from English into other languages (Ware et al 1998). The role of the IQOLA was then to validate, normalise and document the translations as required (Ware & Gandek 1998). According to the QualityMetric™ website the SF-36v2 is currently available in more than 120 language translations and in English-language adapted versions (http://www.qualitymetric.com). In other English speaking countries, adjustments to the phrasing of questions have been made to make them more applicable to their populations. Since 2002 a fee has been charged for using the SF-36v2 and for obtaining the norm-based scoring algorithms. This charge has led some researchers and government users to continue using SF-36v1 (Fryback 2003). According to the SF-36 website (August 2010), there are more than 500 publications that use translations or English language adaptations of the SF-36.

Throughout the 1990s, attempts were made to validate the SF-36. Brazier et al (1992) examined the reliability and validity of the SF-36 in a British population and compared it with the Nottingham Health Profile developed by Hunt et al (1992) and widely used in Britain. Over the decade further validation was conducted with adults of all ages (Ware & Sherbourne 1992; Ware 1993; Ware et al 1994) and the SF-36 has been used extensively in clinical trials (Brazier et al 1998).

The popularity of the SF-36 grew throughout the 1990s and 2000s by many validation studies. The SF-36’s popularity was due mostly to it being brief and a comprehensive measure of generic health status (Ware & Gandek 1998). The brief nature (what was considered at that time) of the SF-36 meant that it “could be supplemented with other generic and disease-specific measures in clinical studies” (Ware & Gandek 1998).

In Australia, the SF-36v1 was applied and validated within Australian populations by McCallum (1995). The SF-36v1 has also been defined and applied to three age-specific populations of Australian women (Mishra and Schofield 1998; Lee et al 2005). The scoring algorithms used by The Australian Longitudinal Study on Women’s Health (ALSWH) have been included in Appendix 2.1. All scoring algorithms currently used for Australian populations can be found at the following website: http://www.alswh.org.au. The development of the scoring algorithms used by ALSWH and the Australian version of the SF-36v1 will be discussed in greater detail later in this chapter and throughout the thesis.
2.4.2 RAND-36 Item Health Survey

The MOS SF-36 was modified by the RAND Corporation and is known as the RAND-36 Item Health Survey (RAND SF-36). The RAND (Research and Development) Corporation’s health insurance experiment required a measure which was sensitive to health differences in a general population, so the RAND SF-36 was developed (Brazier et al 1992; Ware & Sherbourne 1992). The RAND health insurance experiment was a comprehensive evaluation of alternative methods to financing health care in the United States (Brazier et al 1992). The RAND Corporation conducts research on a wide range of topics, beyond health, and began in 1946 as a research project backed by the U.S. Army Air Forces. The RAND Corporation is now an independent, non-profit research institution (http://www.rand.org). According to Hays et al (1993), the RAND 36-Item Health Survey 1.0 includes the same items as those in the MOS SF-36. However, there are differences in how the eight sub-scales and two physical and mental health composite scores were derived (Hays et al 1993).

2.4.2.2 Veteran RAND-36 (VR-36)

There was another modification of the SF-36 for veterans and this version is called the Veteran RAND-36 (VR-36). The VR-36 has different response choices for the two role limitation items (physical and emotional). The role limitation items have been changed from dichotomous scales to a five-point Likert scale (1 = no, none of the time; 2 = yes, a little of the time; yes, 3 = some of the time; 4 = yes, most of the time; 5 = yes, all of the time). This change, according to Kazis et al (2007) increases the instrument’s precision. The VR-36 also includes items to assess change in physical health and emotional problems, rather than the one general change item found in the RAND SF-36 (Kazis et al 2004; Kazis et al 2004b).
The VR-36 has been widely used by the Veterans Health Administration (VHA) in multiple national surveys since 1996 (Kazis et al 2007). The VR-36 is the primary measure of HR-QOL and functional status measures for US Veterans. It is also a secondary measure of disease burden and for case-mix adjustment purposes for comparisons made among veteran networks (Kazis et al 2007).

2.4.3 Developments of SF-12 and VR-12 from SF-36 and VR-36

According to Kazis et al (2007) a shorter instrument, the Short Form-12 (SF-12), was developed from items in the SF-36 in order to further reduce respondent burden. Summary scores from the SF-12 have shown excellent correlation with those from the SF-36 ($r^2 = 0.911$ and 0.918 for PCS and MCS respectively) (Ware et al 1996).

2.4.4 The emergence of SF-12 as a possible replacement of SF-36

The 36 items (of the SF-36) place considerable burden on both the participants and the investigators (Iglesias & Torgerson 2000; Muller-Nordhorn et al 2004; Kazis et al 2007). It also leaves open the possibility of many missing items. Hence, Ware and colleagues decided to develop the substantially shorter questionnaire (Ware et al 1996). Ware et al (1996) suggested the SF-36 is too long for inclusion in some large-scale health measurement and monitoring efforts. So they were able to reduce the number of items from 36 to 12 (Ware et al 1996). Ware and colleagues tested the SF-12 in the general US populations and in the Medical Outcomes Study (MOS) and found that the SF-36 and SF-12 were highly correlated (Ware et al 1996; Muller-Nordhorn et al 2004). There is also the possibility to convert from the previously used SF-36 to the SF-12 and hence obtain PCS-12 and MCS-12 scores. Muller-Nordhorn et al (2004) concluded that
in large studies assessing HR-QOL of patients with heart disease, the SF-12 appears adequate in replacing the SF-36 and thus may reduce the respondent burden and save resources.

2.4.5 The Veteran RAND-12 (VR-12)

The VR-12 was developed as a shortened form of the Veteran RAND-36 (VR-36) (Kazis et al 2006; Kazis et al 2006b). Kazis et al (2007) suggest scores between the two instruments have been shown to correspond well and that the two instruments (VR-12 and VR-36) have demonstrated excellent reliability and validity. A conversion formula has been validated by Kazis et al (2004b) and Jones et al (2001) so that summary scores using PCS and MCS are comparable to the scores of the MOS SF-36 and SF-12 (Kazis et al 2007).

2.4.6 Short Form-12 version 2 (SF-12v2)

The SF-12 Health Survey version 2.0 is a proprietary instrument owned by QualityMetric™ (Kazis et al 2007). This assessment tool includes items all selected from the SF-36. The items include at least one item or more from each of the eight domains in the SF-36. Changes made to the stem and response choices of selected items are controversial and Dr Ware expressed concerns about the comparability of the SF-12v2 scores with the MOS SF-12 and the VR-12 summary scores (Kazis et al 2007).
2.4.7 SF-8 Health Survey

Finally, there is another short-form for surveying adults currently available from QualityMetric™ and is called the SF-8 Health Survey. The SF-8 measures the same eight health domains as the SF-36v2 but uses only eight questions and can produce a PCS-8 and MCS-8 (http://www.qualitymetric.com).

According to the QualityMetric™ website the SF-8 was developed to replicate the SF-36v2 with one question for each health domain. Since the SF-8 uses only one question for each health domain, QualityMetric™ does not recommend using the SF-8 for monitoring the health of individual patients or for smaller studies. However, QualityMetric™ promote the SF-8 as a “perfect way to engage patients on a website”.

2.5 WORLD HEALTH ORGANISATION QUALITY OF LIFE (WHOQOL-BREF)

During the 1990s the World Health Organisation formed the Quality of Life (WHOQOL) group to develop an instrument which could be possibly used across cultures (Fryback 2003). Through the collaboration of 15 field centres which involved lengthy discussion and worldwide focus groups, a 100-item instrument was created via the use of factor analysis. The structure of the WHOQOL-100 focused on twenty-four facets of quality of life which contains four items for each facet, as well as four general items covering subjective overall quality of life and general health, each item uses 5-point Likert response scale (Fryback 2003). The instrument can be located at the following web address: http://www.who.int/evidence/assessment-instruments/qol/ql5.htm.
The WHOQOL-BREF was developed once it was realized that a 100-item instrument would be difficult to implement (Fryback 2003). The WHOQOL-BREF is a 26 item instrument, made up of 24 items selected from the 100 items to cover all 24 facets in the WHOQOL-100 and two general items covering overall quality of life and general health were added (Fryback 2003). In the current approach to scoring the WHOQOL-BREF, four domains have been merged based on factor loading from a confirmatory factor analysis of the 100 items. The four major domains are assessed: physical, psychological, social relationships and environment impacts on quality of life (Fryback 2003). The instrument can be located at the following web address: http://www.who.int/evidence/assessment-instruments/qol/ql6.htm.

2.6 Brief History of Generic Health Utility Indexes

Health utility indexes or preference-based measures yield both single and multi-attribute utility values anchored at zero (death) and 1.0 (perfect health) as a measure of HRQOL (Trisolini et al 2002; Marra et al 2005). The scores from these utility instruments can be integrated into cost-utility analyses as the weightings for quality adjusted life years (QALY) (Marra et al 2005).

Health utilities indexes are used in economic evaluations to integrate survival and HRQOL into a single metric and can be measured either directly or indirectly (Marra et al 2005). Direct utility techniques include the use of standard gamble (SG) or time trade-off (TTO), whereas indirect utility techniques have been developed using multidimensional HR-QOL questionnaires. Examples of indirect techniques include the Health Utilities Index 2 and 3 (Torrance et al 1996; Feeny et al 2002; Feeny et al 2004) the Short-Form 6D (SF-6D), (Brazier et al 1992) and the EuroQoL (EQ-5D) (The EuroQoL Group, 1990).
In the next section, the direct techniques such as, the Standard Gamble (SG) and Time Trade Off (TTO) will only be briefly discussed. Greater focus will be given to the indirect methods, such as the SF-6D which is of greater interest in this thesis.

2.6.1 Direct and Indirect Utility Techniques

Figure 2.3 shows a comparison of the direct and indirect techniques which were discussed by Brazier et al (1999). The direct technique involves mapping preferences directly onto a scale, for example a visual analogue scale. Other direct approaches involve techniques known as Standard Gamble (SG) and Time Trade Off (TTO). Standard Gamble (SG) is a technique which involves the respondents having a choice of alternative outcomes based upon a level of uncertainty. As Brazier et al (2002) suggest, it is a choice about how much risk of “death or some other bad outcome the respondent is willing to accept to avoid living in the certainty of the health state”.

![Comparison of Direct and Indirect Techniques](image)

Source: (Brazier et al (1999) and adapted (and shown in Figure 2.3) by Arnold et al 2009- Figure 1 Direct versus indirect methods of utility elicitation)

Figure 2.3 Direct vs Indirect Techniques according to Brazier et al (1999) and Arnold et al (2009)
The other direct utility technique known as Time Trade Off (TTO) was developed as an alternative to the Standard Gamble. The purpose of the TTO design was to better explain probabilities to the respondents (Brazier et al 2002). “Respondents are asked to consider trading off their length of life for an improvement in health”, (Brazier et al 2002) that is, a shorter length of life but with an improved state of health.

The TTO and SG techniques require a trained interviewer who needs to explain the concepts without biasing the responses or distressing the patient. These techniques can be much more time and resource consuming than the indirect techniques and they do not capture dimensions such as mobility or pain (Arnold et al 2009).

2.6.2 Indirect utility techniques

The following sections will discuss the indirect utility techniques mentioned by Fryback (2003) and summarised again in Fryback et al (2007). The health utilities considered as indirect techniques include: QWB-SA, HUI (marks II and III), EQ-5D and SF-6D.

Each indirect utility involves a multidimensional health state classification (Fryback et al 2007). These health state classifications are usually a set of health domains referred to as “attributes or dimensions, (such as pain) which have pre-defined levels (e.g., “none”, “moderate”, “severe”)” (Fryback et al 2007). Thus the utility indexes (scores) are indirectly derived from the self-reported participant responses to the health domain questionnaire items.
2.6.2.1 Quality of Well Being Scale self-administered form (QWB-SA)

The Quality of Well Being Scale (QWB) evolved from the Index of Well-Being (Fryback 2003). The Index of Well Being is the one the first summary health related quality of life indexes. The QWB was also one of the first instruments to combine multi-attribute theory with psychometric methods for scale constructions (Fryback 2003). The QWB combines symptoms and problems with three scales of functioning – mobility, physical activity, and social activity – to produce an estimate of well-being on a scale from 0 (dead) to 1.0 (Anderson et al 1998). The QWB summary averaged the daily well-being scores for the 6 days immediately prior to the time of the interview (Fryback 2003). A self-administered version of the QWB has also been developed (Fryback et al 2007). The QWB-SA asked about the past 3 days (Fryback et al 2007) and not past 6 days. It has been shown that the QWB-SA is highly correlated with the interviewer administered QWB and retains the psychometric properties (Kaplan 1997). The QWB-SA has been validated and compared to the SF-36 in a population of older adults (Andersen et al 1998; Fryback 2003). While the newer QWB-SA is still somewhat long, the construction and testing of the self-administered form represents a major advance in the development of the QWB instrument (Kazis et al 2007).

2.6.2.2 Health Utilities Index Mark II and III (HUI2) & (HUI3)

The Health Utilities Index Mark III (HUI3) is a multi-attribute system that consists of eight attributes: hearing, vision, speech, ambulation, dexterity, emotion, cognition and pain (Fryback et al 2007). Each attribute consists of multiple levels of functioning varying between normal and severe (Hatoum et al 2004). The questionnaire is available as a self administered form with 15 questions or a computer assisted telephone interview (CATI) with 40 questions. Hence the HUI can be administered as a
self-assessment or as an assessment by proxy (Coons et al 2000; Horsman et al 2003). The questionnaire is distributed by Health Utilities Inc. group and a fee is charged for access to the questionnaire, coding algorithm and procedure manuals. Torrance’s multi-attribute utility theory is based on data from focus groups (Torrance et al 1996). According to Fryback (2003), “the data were used to devise a set of dimensions that collectively are comprehensive in covering important aspects of health”. Torrance et al (1995) developed Time Trade Off as an alternative approach to standard gamble which was used to scale the HUI1. In the later HUI2 and HUI3 standard gamble was used in scoring. There are seven dimensions of health status included in the HUI2, these include: sensation (vision, hearing, speech), mobility, emotion, cognition, self-care, pain, and fertility (Fryback 2003). There are three to five levels per attribute, ranging from a severe problem to no problem/normal. The HUI2 health-status classification system defines 24,000 unique health states (a factorial of the numbers of levels in each attribute) (Barr et al 2001). For the HUI3’s eight attributes there are five or six levels per attribute (Fryback et al 2007). The multiplicative, multi-attribute scoring function is based on preference scores from a random sample of the general adult population in Hamilton, Ontario (Fryback et al 2007).

The Health Utility Index was initially developed for use in children with cancer (Coons et al 2000; Horsman et al 2003; Kazis et al 2007). The validity of the Health Utility Index in adult populations is comparable to that in children (Torrance et al 1996). However, the Health Utility Index has several methodological shortcomings (Kazis et al 2007). Although each dimension has a high number of response choices, the HUI instrument has been shown to have ceiling effects (i.e. the highest attainable value below a set maximum), and the validity of the scoring system has been questioned by Hatoum et al (2004) and McCabe et al (2005) (Kazis et al 2007). Both the HUI2 and HUI3 use a scoring algorithm which allows health states to be scored less than 0
The HUI2 index ranges from -0.03 to 1.0, while the HUI3 index ranges from -0.36 to 1.0 (Fryback et al 2007). Further information about the HUIs can be found at the website: http://www.healthutilities.com/.

2.6.2.3 EuroQol -5D

The Euro-Qol-5D (EQ-5D) was designed as a health index to describe and value Health Related Quality of Life (Brooks et al 1996; Marra et al 2005). The EQ-5D instrument was developed in 1987 by a multi-disciplinary group of researchers from seven centres across five countries. The aim was to develop an instrument which could be self-administered and was not disease-specific. It was intentionally developed to generate an interval health scale index and was concurrently developed in five languages (Fryback 2003). The EQ-5D instrument has five dimensions: mobility, self-care, usual care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels and together defined 243 health states. It has become one the most widely used generic measures of health in Europe (Kazis et al 2007) and has been commonly used in economic evaluations (Brazier et al 2004; Brazier and Roberts 2004). The responses on the EQ-5D can be summarized into a single summary score, known as the EQ-5D Index and the summary score algorithm is based on European valuation sets (Petrou & Hockley 2005; Kazis et al 2007).

The EQ-5D has shown high reliability and low response bias, although there is a problem with ceiling effects, due to the small number of response choices (Kazis et al 2007). In one study, more than 52% of subjects had scores of 1.0 (Petrou & Hockley 2005). However, Kazis et al (2007) suggests that this ceiling effect must be weighed up against the high response rates and low rates of missing data. Further information about the EQ-5D can be found at the website: http://www.euroqol.org.
2.6.2.4 Short Form-6D

During the 1990s, Brazier and his colleagues at the University Of Sheffield (UK) developed a preference-based single index measure of health from the SF-36 (Brazier et al 2004, Hatoum et al 2004). “While the SF-36 includes algorithms for producing summary scores, the SF-36 does not incorporate patient preferences” (Kazis et al 2007). Deriving preference-based summary scores enables the calculation of QALYs (Quality Adjusted Life years) and cost-effectiveness analysis (Brazier et al 2004).

The SF-6D was developed using a multi-attribute system similar to that used for HUI3 (Hatoum et al 2004). The SF-6D is composed of six multilevel dimensions of health, which together define 18000 states of health. Through development, using a UK general population, an algorithm has been generated to convert SF-36 data at the individual level to a preference-based index (SF-6D) (Brazier et al 2004). “With the availability of many SF-36 data sets that could be converted into SF-6D preference based measures, it is anticipated that the application of this measure will continue to grow” (Kopec & Willison 2003; Marra et al 2005).

2.7 Mapping from the SF-36 and SF-12 to the SF6D

The mapping of SF-36 responses to the SF-6D, was a three-step approach (Kazis et al 2007).

Step 1:

Firstly, the SF-36 was reduced to 20 items and mapped to six dimensions: physical functioning, role limitation, social functioning, bodily pain, mental health and vitality (Kazis et al 2007). The role limitation dimension had two choices while the other
dimensions had five or six. In total 9000 health states are mapped by the original version of the SF-6D (Kazis et al 2007).

**Step 2:**

A subset of fifty-nine health states (from the 9000 total health states) was produced to represent all six health dimensions. Then a group of 165 subjects in the UK were asked to value these 59 health states using a Visual Analog Scale (VAS) and Standard Gamble (SG) techniques (Kazis et al 2007). Each person valued a subset of the 59 health states, such that some health states were valued by all 165 subjects.

**Step 3:**

For the 9000 health states a value was estimated using statistical inference techniques. Separate models using Visual Analogue Scores and Standard Gamble values were then developed (Kazis et al 2007). Later developments by Brazier et al (2002) and Brazier et al (2004) saw these 9000 health states double by expanding the choices of the role limitation.

According to Kazis et al (2007) a later version of the SF-6D was developed from the MOS SF-12 which uses 7 of the 12 items. Comparison with the original SF-6D revealed that the newer SF-6D (SF-12) produced similar scores (Brazier et al 2004). Kazis et al (2007) suggest that as a preference-based instrument the SF-6D has been shown to be reliable and valid, although Kopec and Willison (2003) observed floor effects. Kazis et al (2007) did raise some concerns with the SF-6D, such as it has only been used in a few specific United Kingdom populations and the weightings for generating utilities may not be generalisable to US populations.
2.8 USAGE OF TRANSLATING NINE GENERIC HRQOL INSTRUMENTS FOR USE IN AFRICA

Most generic measures of HRQOL have been developed in English, leaving researchers working in other languages with two options; either to develop a new measure and/or to translate an existing measure (Bowden and Fox-Rushby 2003). There has been much concern and criticism of the quality of the translations, hence the need during the 1990s to develop guidelines for translating HRQOL instruments (Guillemin et al 1993, Bullinger et al 1998).

Between 1990 and 1999 there were 58 papers using nine generic HRQOL instruments which had been translated for use in Africa. The SF-36 dominates the literature (40%) followed by the WHOQOL (19%), Dartmouth COOP Charts (14%), EQ-5D (9%), Nottingham Health profile (NHP) (7%), Sickness Impact Profile (SIP) (5%), 15-D (3%) and HUI (3%). No papers were identified for QWB (Bowden and Fox-Rushby 2003).

Initially, Bowden and Fox-Rushby (2003) conducted a literature search in the mid 90s using Medline, EmBase, Psychological Abstracts, Social Science and Psychological Abstracts and ADIS which had no restriction on language. This search generated 1347 papers (Bowden and Fox-Rushby 2003). The proportions mentioned also translate across to similar usage in English speaking countries.

Acquadro et al (2008) conducted a literature review of methods to translate Health-Related Quality of Life questionnaires for use in multi-national clinical trials. They identified 891 references by searching MEDLINE, Embase, and the Mapi Research Trust’s database with “quality-of-life,” “questionnaires,” “health status indicators” matched with “translating,” “translation issues,” “cross-cultural research,” and “cross-
cultural comparison”. They included articles in their review if they proposed, compared or criticized translation methods. They identified 45 articles which made their criteria and concluded that producing high-quality translations is labour-intensive (Acquadro et al 2008).

2.9 **Development of the Australian Standardised SF-36**

Kazis et al (2007) suggest that generic instruments measuring HR-QOL are developed using methods from cognitive and psychology and psychometric theory. These instruments attempt to capture a multidimensional health perspective and are applicable in various patient populations (Kazis et al 2007). The MOS SF-36 was considered as a very useful instrument for the Australian Longitudinal Study on Women’s Health (ALSWH) since the study involves three age cohorts of women. The following section will discuss the development of the Australian Standardised SF-36 by ALSWH.
2.9.1 The Australian Longitudinal Study on Women’s Health (ALSWH)

Originally known as Women’s Health Australia, The Australian Longitudinal Study on Women’s Health (ALSWH) came about as part of the development process of the National Women’s Health Policy during the 1980s (Lee et al 2005). A national consultation with women’s organizations representing more than a million women developed the idea of a national longitudinal study on women’s health. The longitudinal study was to be “premised on a social rather than a narrowly focused medical approach to health” (Lee et al 2005). Studies at the time were suggesting that Australian women were mostly concerned about tiredness, menstrual difficulties, overweight, depression and anxiety (Brown & Redman 1995; Brown & Doran 1996; Lee et al 2005). From these studies and questionnaire development the ALSWH decided to use the SF-36 as the main HR-QOL tool. However, as will be discussed in this chapter there were changes made to the wording of the questionnaire items and ALSWH standardised the SF-36 to three aged cohorts of population-based Australian women.

According to Lee et al (2005), the ALSWH’s main strengths are that it includes women from three different age cohorts within a longitudinal design. The ALSWH is based on a national sample, rather than being defined by a region and participants come from the “widest possible range of geographic, socioeconomic, and personal circumstances”. Apart from the statistical power afforded by the large database, insight can be gained from the qualitative answers that participants can provide at the end of each survey (Lee et al. 2005).
2.9.2 ALSWH PCS and MCS standardised using a formula relevant to Australian population

As previously mentioned, the MOS SF-36 provides an eight scale health profile and two component summary scores representing physical and mental health (Ware et al 1994). The two distinct factors (physical and mental health) were confirmed by factor analysis in general US population. The published norms and scoring procedures are based on data from the US general population. According to Mishra and Schofield (1998), the PCS and MCS scores had been shown by McHorney et al (1993) and Ware et al (1994) to have good discriminant validity in US samples. Age and gender based norms have been reported for the general population by Ware et al (1994) but not for the Australian populations. Australian studies by McCallum (1995) and Schofield and Mishra (1998) found differences in the mean and standard deviations for the eight scales, as well as in the factor structures between Australian and US populations. Thus Mishra and Schofield (1998) decided there was a need to establish Australian age and gender-based norms for PCS and MCS. Mishra and Schofield (1998) argued that the scoring formula used for calculating the summary scores for the US population did not take into account the differences across age and by gender. Complete details of all three age cohorts for the ALSWH project are available in Lee et al (2005) and via the study website: http://www.alswh.org.au/.

2.9.3 Measures

The ALSWH baseline questionnaire included the official Australian version of the SF-36 Health Profile (Ware 1993), as approved by the Medical Outcomes Trust in Boston. The Australian version has a slight modification to the wording of two items but conceptual dimensions remained similar (Mishra and Schofield 1998). There was a need to remove the Americanisms ‘pep’ and ‘blue’ which are not commonly used by
Australians. They were replaced with terms such as ‘life’ and ‘sad’ respectively, while McCallum (1995) replaced them with ‘enthusiasm’ and ‘sad’.

2.9.4 Analysis

Mishra and Schofield (1998) found that their principal component analyses identified two factors in each age group. When they rotated the two principal components into simple orthogonal structures they found that the results of the eigen values for the older cohort behaved differently to the younger and middle age cohorts. Only the first principal component corresponded to an eigen value of greater than 1. Mishra and Schofield (1998) suggest that a single factor solution was plausible which supported earlier finding of the McCallum (1995) Australian Bureau of Statistics (ABS) sample of approximately 2000 adults.

In order to compare the mean PCS and MCS scores of the ALSWH cohorts with the US women (similar age groups), Mishra and her colleagues used a two sample t-test. The PCS and MCS scores for the ALSWH cohorts were calculated separately for each age from weights created using the McCallum (1995) Australian norms, whilst the US women were from Ware et al (1994) (Mishra and Schofield 1998).

2.9.5 Norms of the PCS and MCS scores for older Australian women.

Table 2.1 presents the norms (means, standard deviations and median) for the PCS and MCS scores for the three age groups (Australian and US populations). Note that the US norms are based on roughly comparable age groups. There appears to be some skewness of the distributions, since the median for the PCS and MCS scores are higher than the means for the three age groups, respectively (Mishra and Schofield 1998). According to Table 2.1 there was a statistically significant difference in mean PCS scores for the older cohort of Australian (70-74 years) and US women (65-74
years). While there was no statistically significant difference in mean MCS scores for the older cohort of Australian (70-74 years) and US women (65-74 years).

According to Mishra and Schofield (1998) the summary statistics provided by the ALSWH represent the largest sample available in the Australian context and is also one of the largest age-based samples internationally. Although the summary statistics show similar trends as the US population, such as the increase in mean mental health scores with ageing, Mishra and Schofield (1998) believed there was enough variation to suggest that the two populations should be treated differently. Finally, with the ever increasing usage in the Australian Bureau of Statistics (ABS) National Health Survey, Mishra and Schofield (1998) believe it is important to have relevant population based norms for comparative purposes.

Table 2.1 Norms for PCS and MCS scores for three groups of Australian women compared with US norms for females

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>PCS</th>
<th></th>
<th>MCS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Median</td>
</tr>
<tr>
<td>ALSWH cohort</td>
<td>48.89</td>
<td>8.70</td>
<td>50.95</td>
</tr>
<tr>
<td>US population</td>
<td>53.39</td>
<td>8.74</td>
<td>55.15</td>
</tr>
<tr>
<td>Mean difference</td>
<td>-4.50**</td>
<td>-2.02*</td>
<td></td>
</tr>
<tr>
<td>ALSWH cohort</td>
<td>49.55</td>
<td>9.32</td>
<td>52.15</td>
</tr>
<tr>
<td>US population</td>
<td>48.95</td>
<td>9.64</td>
<td>51.61</td>
</tr>
<tr>
<td>Mean difference</td>
<td>0.60</td>
<td>-3.36**</td>
<td></td>
</tr>
<tr>
<td>ALSWH cohort</td>
<td>51.10</td>
<td>9.10</td>
<td>52.64</td>
</tr>
<tr>
<td>US population</td>
<td>41.02</td>
<td>11.52</td>
<td>42.93</td>
</tr>
<tr>
<td>Mean difference</td>
<td>10.08**</td>
<td>-0.12</td>
<td></td>
</tr>
</tbody>
</table>

* Calculated using Australian weights (McCallum et al 1995)  
** Data from Ware et al (1994)

ALSWH cohort vs US population  
* p value < 0.05 and **p value ≤ 0.0001

2.10 Validity of the SF-12 Compared with the SF-36 in ALSWH Pilot Studies

Schofield and Mishra (1998) saw the emergence of the SF-12 and the potential for widespread uptake. However, in their caution they decided to use pilot study ALSWH data to compare the SF-12 with the SF-36. They found that the eight dimensions of the SF-12 may differ more from the SF-36 than the current norms suggested. Their results also suggested that “the SF-36 offers considerably more scope for studies that track individuals over time since the eight dimension scores derived from the 36-item measure offer greater comparability with large amount of normative data internationally available on the SF-36”. Schofield and Mishra (1998) concluded that the “SF-36 provides a more reliable measure of health status because of its more precise measurement and seems to offer greater capacity to detect change over time”. Thus the SF-36 was chosen as the HR-QOL tool for all three ASLWH cohorts at baseline and future follow-ups.
CHAPTER 3:

Literature Reviews

Concepts of Missing Data and Current Methods to Dealing with Missing Data in Longitudinal Studies
List of Abbreviations used in Chapter 3

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSWH</td>
<td>Australian Longitudinal Study on Women's Health</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation Maximization</td>
</tr>
<tr>
<td>FIML</td>
<td>Full Information Maximum Likelihood</td>
</tr>
<tr>
<td>GEEs</td>
<td>Generalised Estimating Equations</td>
</tr>
<tr>
<td>HR-QOL</td>
<td>Health Related Quality of Life</td>
</tr>
<tr>
<td>LISREL</td>
<td>Linear Structural Relations</td>
</tr>
<tr>
<td>LOCF</td>
<td>Last Observation Carried Forward</td>
</tr>
<tr>
<td>LVCF</td>
<td>Last Value Carried Forward</td>
</tr>
<tr>
<td>MAR</td>
<td>Missing at Random</td>
</tr>
<tr>
<td>MCAR</td>
<td>Missing Completely at Random</td>
</tr>
<tr>
<td>MCMC</td>
<td>Markov chain Monte Carlo</td>
</tr>
<tr>
<td>MCS</td>
<td>Mental Component Summary score</td>
</tr>
<tr>
<td>MDE</td>
<td>Missing Data Estimation</td>
</tr>
<tr>
<td>MI</td>
<td>Multiple Imputation</td>
</tr>
<tr>
<td>MICE</td>
<td>Multivariate Imputation using Chained Equations</td>
</tr>
<tr>
<td>MI-GEE</td>
<td>Multiple Imputation- Generalised Estimating Equations</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MLE</td>
<td>Maximum Likelihood Estimation</td>
</tr>
<tr>
<td>MNAR</td>
<td>Missing Not at Random</td>
</tr>
<tr>
<td>MRE</td>
<td>Modified Regression Estimates</td>
</tr>
<tr>
<td>NMAR</td>
<td>Not Missing at Random</td>
</tr>
<tr>
<td>PCS</td>
<td>Physical Component Summary score</td>
</tr>
<tr>
<td>PCT</td>
<td>Diehr's Probability of being healthy in one year (transformed PCS)</td>
</tr>
<tr>
<td>PHS</td>
<td>Physical Health Status</td>
</tr>
<tr>
<td>RE</td>
<td>Regression Estimates</td>
</tr>
<tr>
<td>REML</td>
<td>Restricted Maximum Likelihood</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic Blood Pressure</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short Form 36</td>
</tr>
<tr>
<td>SG</td>
<td>Standard Gamble</td>
</tr>
<tr>
<td>TTO</td>
<td>Time Trade Off</td>
</tr>
<tr>
<td>WinBUGS</td>
<td>Windows -Bayesian inference using Gibbs Sampling</td>
</tr>
<tr>
<td>WEE</td>
<td>Weighted Estimating Equations</td>
</tr>
<tr>
<td>WGEE</td>
<td>Weighted Generalised Estimating Equations</td>
</tr>
</tbody>
</table>
3.0 INTRODUCTION

This Chapter contains two main sections:

Section 3.1 describes the concepts that are important for analysing longitudinal studies which have missing data.

Section 3.2 involves two literature reviews:

a) a review of the literature on the current methods being considered to account for deaths in health related quality of life longitudinal studies of elderly populations;

b) a review of the literature on the current methods and advancements for analysing longitudinal studies with missing data.

3.1 CONCEPTS AND TERMINOLOGY OF MISSING DATA

The purpose of this section is to introduce the main concepts that will be discussed throughout the thesis. In later chapters where statistical methods will involve the imputation of missing data this section will explain the terminology and introduce certain approaches for dealing with missing data. However, prior to this, the two statistical theories which underpin the methodologies applied throughout the thesis will be defined.

3.1.1 Theories and Concepts

3.1.1.1 Bayesian Vs Frequentist Theory

At this point a distinction and clarification of the two prominent “ideologies” used in statistical inference needs to be mentioned. Statistical inference can be thought of as two distinct but at times overlapping theories known as Frequentist theory and
Bayesian methods. Frequentist methods for analysing longitudinal data are well established in current statistical practices and programs. In recent times there has been a surge in the literature towards using Bayesian modelling techniques. One suggestion by Kynn (2006) is that this could be due to the increased computational power of personal computers and the development of Bayesian statistical programs (Spiegelhalter et al 2003).

The frequentist interpretation of a parameter (β) estimate from a simple linear regression (for example) would be that there is a ‘true’ slope to the line that is contained in the confidence interval centred on the estimated parameter value of ‘b’. That is, the 95% confidence interval will contain the true parameter (β) 95% of the time. While the Bayesian interpretation says that the parameter (β) of interest is 95% likely to lie in the interval. The core difference between the two interpretations comes back to whether one considers the hypothesis to be fixed (frequentist) or the data to be fixed (Bayesian) (Kynn 2006). There are also differences in the theory underlying how models are calculated, but in most cases when there is sufficient data one should obtain similar parameter estimates for the model (Kynn 2006).

This theorem is generally translated to the following:

\[ p(H|D) = \frac{p(D|H) \times p(H)}{p(D)} \]  \hspace{1cm} \text{(equation 3.1)}

\[ \text{posterior} = \text{prior} \times \text{likelihood} / p(D) \]

\[ \text{posterior} \propto \text{prior} \times \text{likelihood} \]

The prior distribution allows one to formally state what is believed before the collection of the data, and the likelihood function suggests how to update beliefs. The powerful part of Bayesian formulation is the information does not need to be lost between subsequent data collection exercises (Spiegelhalter et al 2004; Kynn 2006).
Throughout this thesis, there will be many references to Bayesian theory. Later in Chapter 5 a Fully Bayesian approach to dealing with missing data will be applied. So what are missing data and what are the impacts of missing data on longitudinal studies? The following sections will define missing data and explain current ways for dealing with the problem.

3.1.1.2 What are missing data?

Schafer and Graham (2002) suggest that missing values belong under the heading of a more general concept known as coarsened data. That is, coarsened data includes values which have been grouped, aggregated, rounded, censored or truncated, which results in some loss of information (Heitjan & Rubin 1991; Schafer & Graham 2002). Missing values can occur when no response is recorded, for example, for a particular item in a questionnaire or a measurement such as systolic blood pressure. In these instances a true value may exist but has not been captured in the data collection and is therefore truly unknown. Other times however, missing values can occur as part of the “grand plan” of the data collection. For example, a response to an item in the survey may prompt the responder to skip items because they are not relevant since they answered “No” to the first item. There are times when data are coded to indicate a lack of response, such as: “Don’t Know” or “Refused to respond” (in a telephone survey). This information can be helpful in understanding individual differences in participants. However, the lack of response codes may not be included in the analysis which results in a reduction in sample size and a loss of study power.
3.1.1.3 A Brief Background of Statistical Techniques

At this point a brief background into the two distinct areas for dealing with incomplete data needs to be introduced. The field of incomplete data is fast advancing so the techniques which underpin this field will be briefly mentioned without going too deeply into theoretical discussion.

Historically, until the 1970s missing values were dealt with by editing the dataset. This was common practice until Rubin (1976) published a “framework of inference from incomplete data” (Schafer & Graham 2002). Rubin’s framework is still the basis of current methodology and led to the development of the Expectation-Maximization (EM) algorithm (Dempster et al 1977). The EM algorithm came into prominence in 1977 due to the work of Dempster, Laird and Rubin (Schafer 1997). The EM algorithm has become the general technique for finding maximum-likelihood estimates for parametric models when there were missing data (Schafer & Graham 2002). According to Schafer (1997) the technique caused a revolution in analysis of incomplete data and made it possible to calculate parameter estimates using available data, thus rendering the ad hoc methods like case deletion obsolete. In 1997 Schafer mentioned that the producers of statistical software have been rather slow to incorporate general purpose EM algorithms for incomplete data into products. Since then the implementation of the EM algorithms has gained in popularity and is becoming a standard feature of statistical methods of longitudinal data, such as Stata’s *xtmixed* command which will be considered in Chapters 6 and 7. The EM algorithm is considered a likelihood-based technique, whereas the technique known as Markov Chain Monte Carlo (MCMC) is considered Bayesian. The MCMC technique is generally applied using the Gibbs Sampling or the Metropolis-Hastings Algorithm (Schafer 1997).
According to Schafer (1997), the Markov Chain Monte Carlo is a “body of methods for generating pseudorandom draws from probability distributions via Markov chains. A Markov chain is a sequence of random variables in which the distribution of each element depends on the value of the previous one”. The process of repeatedly simulating steps of the chain will continue as long as the conditions are met until eventually it stabilises to what is known as the common or stationary distribution of interest (Schafer 1997). This MCMC technique, like the EM algorithm has caused a revolution in the field of applied Bayesian inference (Schafer 1997).

The Gibbs Sampling and Metropolis-Hastings algorithm are considered to be methods of MCMC. These are the two most popular methods being used. The Gibbs Sampling will be discussed in greater detail in Chapter 5 where it will be applied using the Bayesian software called WinBUGS (Windows-Bayesian Inference Using Gibbs Sampling).

Further developments by Little and Rubin (1987), Rubin (1987) and Schafer (1997) have seen major advances in dealing with the problems associated with missing data. In 1987 Rubin introduced the idea of multiple imputation (MI) that has become a benchmark in dealing with missing data. There have been major developments during the past two decades in computing power, which has lead to the emergence of the Bayesian MI method. Also, since the 1990s a number of researchers have been developing other techniques which have advanced the field and will be summarised later in this chapter. Thus before discussing the techniques to dealing with missing data it is important to understand the concepts and assumptions which are the framework beneath these methods.
3.1.2 Attrition in longitudinal studies

Missing data is a common problem confronting researchers, especially in longitudinal studies and is often caused by attrition. Attrition is a situation where not all N participants have data on all T measurements (Twisk 2003). When participants have missing data at the end of a study they are referred to as drop-outs. There can be any number of reasons why a participant drops out of a study, the most obvious example being the death of a participant. Data can also be missing intermittently creating “holes” in the dataset. That is, a participant may miss a particular measurement (e.g. survey) and then return to the study at the next follow-up.

Attrition occurs for many reasons, and deaths are considered as a special problematic case of drop-out. Causes for missing data vary depending on the type of study design. For clinical trials and/or intervention studies, Hogan et al (2004) suggest some reasons include: intermittent missed visits by study participants, discontinued participation or loss to follow-up, lack of effectiveness of the treatment or intervention or mortality. Whereas in a cohort study such as the Australian Longitudinal Study on Women’s Health (ALSWH) reasons include: the participant did not attempt the survey; or the participant was unable to be contacted; or the participant was too frail to continue; or the participant had withdrawn or was deceased before the survey. The reasons for attrition in the ALSWH older study group will be discussed and analysed in greater detail in later chapters. The following section will focus on the general differences between intermittent missing data and drop-outs.
3.1.2.1 Intermittent missing and drop-outs

Figure 3.1 is an illustration of intermittent missing data and drop-outs in a panel dataset (Twisk 2003). A panel dataset is another term used to describe a longitudinal (or cohort) dataset. A time point where data are collected has many different names depending on the field of research. A data collection time point may be referred to as a wave, a panel, a unit, a survey or simply a time point. A dataset can be considered as incomplete when participants have missing data (noted as “X”) as shown in Figure 3.1. Each box (in Figure 3.1) represents a cohort at each time point. A row in the box represents the participants and the columns may for example represent items (variables) in the questionnaire. The ‘X” represent the missing items for each participant. In the first row of boxes, participants have missed items intermittently. Some participants shown in the second row of boxes are missing all items by time points 3 and 4. Since all items are missing (X’s) these participants may be considered to have dropped-out of the survey. Also, a participant could be considered to have intermittent missing data and then drop-out. That is, prior to dropping-out of the survey, participants have missed or refused to answer items in the survey or in an earlier survey. To be considered a true drop-out, the participant must no longer participate, it is possible that a participant may miss a complete survey then return at a later time point. Thus the participant has intermittent units and not just intermittent survey items.
Figure 3.1 Illustration of intermittent missing data and drop-outs

Longitudinal studies can generally be classified as clinical trials and cohort studies both of which follow-up participants over T measurements. Although different in study design, both are prone to drop-out. However depending on many factors, the nature of the drop-out can be very different. For example, in the case of the older cohort in the ALSWH, the likely reasons for drop-out are death and illness. Whereas in a clinical trial, the participants may drop-out following adverse events that are caused by the treatment being administered.

Although a distinction has been made between intermittent missing data and drop-out, a participant in a longitudinal study can have intermittent missing data then drop-out of the study. Missing data are highly likely in clinical trial since participants may either discontinue their prescribed treatment (non-compliance) or cease to be evaluated (drop-out) or both (Carpenter et al 2002).
As Figure 3.1 (Twisk 2003) illustrates, missing data in longitudinal studies can have numerous patterns. In the following sections the concepts of missing data patterns and “missingness” will be defined and discussed.

3.1.2.2 Missing data patterns – monotone and non-monotone missing

In order to further address reasons for drop-out, the distinctions between the missing data patterns need to be discussed. Missing data patterns are generally referred to as monotone missing and non-monotone missing. An example of a monotone missing data pattern is when participants drop-out because of death or other reasons such as frailty or illness. Figure 3.1 illustrates this monotone missing pattern in the drop-outs. Non-monotone missing patterns occur when a participant’s data are missing intermittently. Non-monotonic missing patterns could be due to any number of reasons, such as the study design or refusal and/or failure to answer all items in the survey. Schafer and Graham (2002) use the term arbitrary patterns to describe non-monotonic patterns.

In Chapter 4, an attempt to deal with the impact of the monotone missing patterns due to deaths will be considered. While in Chapter 5 the implications of drop-out for other reasons than death and non-monotone missing patterns will also be considered. Before going any further with discussion, it is at this point the concept of “missingness” needs to be considered.

3.1.2.3 The Distribution of Missingness

Before attempting to describe what has been termed the missingness mechanism and differentiating between the assumptions of missingness (Rubin 1976), there needs to be discussion about the typology developed by Rubin.
For any dataset there exists a distribution $R$ (missingness) which identifies what is known and what is missing (Schafer & Graham 2002). This $R$ distribution can take on many forms (usually binary) but it can depend on the complex nature of missingness patterns. This $R$ distribution is called the missingness mechanism (Schafer & Graham 2002) and is treated as a set of random variables having a joint probability distribution. The distribution of $R$ as a single binary variable indicates whether $Y$ is observed ($R=1$) or is missing ($R=0$) (Schafer & Graham 2002).

In simple terms this $R$ distribution can be thought of as a mathematical tool to describe patterns of missing values and to explore possible relationships between the missingness and the values of the missing items (Schafer & Graham 2002). Schafer & Graham (2002) point out that the $R$ distribution is not used to suggest causality so they refer to the probability distribution of $R$ as the distribution of missingness.

Since missingness may be related to the data, Rubin classified the distributions for $R$ according to the relationships between missingness and the data. Thus Rubin (1976) developed his widely used typology for these distributions (Schafer & Graham 2002). The generic notation Rubin (1976) used is as follows: Let $Y_{com}$ denote the complete data. The complete data is then partitioned into two parts: observed ($Y_{obs}$) and missing ($Y_{miss}$). This is denoted as: $Y_{com} = (Y_{obs}, Y_{miss})$. This notation will be extended in the next section when discussing relationships between missingness and the data.

### 3.1.2.4 The Concept of Missingness

The concept of “missingness” discussed by Rubin (1976) has generated ongoing discussion about the definitions and patterns of missingness. Rubin suggested that missing data takes three forms: missing completely at random (MCAR): missing at random (MAR) and missing not at random (MNAR). Missing at random is where
missing data are dependent on observed data but not on unobserved data, or in other words, given the observed data, the unobserved data are random. Missing completely at random is a special case of MAR where missing data are independent of both observed and unobserved data. Missing not at random is where missing data are dependent on unobserved data (Little & Rubin 1987; Twisk 2003). The term missing not at random has also been referred to as not missing at random (NMAR).

The following equations are often used to explain MAR, MCAR and MNAR:

Missing at Random (MAR) allows the probabilities of missingness to depend on observed data \((Y_{obs})\) but not on missing data \((Y_{miss})\).

\[
P(R|Y_{com}) = P(R|Y_{obs}) \quad \text{(equation 3.2)}
\]

MCAR is when the distribution of missingness does not depend on the observed data

\[
P(R|Y_{com}) = P(R) \quad \text{(equation 3.3)}
\]

If equation 3.2 is not upheld and the distribution of missingness in fact depends on \(Y_{miss}\) then the missing data are considered as missing not at random (MNAR) (Schafer and Graham 2002).

Since first being described by Rubin (1976) the missingness mechanism has been discussed at length in the statistical methodology literature (Little & Rubin 1987; Rubin 1987; Laird 1988; Robins et al 1995; Schafer 1997; Diggle et al 2002; Schafer & Graham 2002; Horton & Kleinman 2007). Missing at random (MAR) and missing not at random (MNAR) are concepts which create a lot of confusion and can be difficult to understand how they actually differ. Schafer and Graham (2002) recognized that the misunderstanding arises from the notion about the meaning of random.
“To a statistician, random suggests a process that is probabilistic rather than deterministic. In that sense, MCAR, MAR, and MNAR are all random, because they all posit probability distributions for $R$. To a psychologist, random may suggest a process that is unpredictable and extraneous to variables in the current study (e.g., tossing a coin or rolling a die), a notion that agrees more closely with MCAR than with MAR. In retrospect, Rubin’s (1976) choice of terminology seems a bit unfortunate, but these terms are now firmly established in the statistical literature and are unlikely to change” (Schafer & Graham 2002).

Also, determining whether missing data are MCAR, MAR or MNAR is difficult to truly assess. Schafer and Graham (2002) provide a very helpful explanation by an example. Their example is shown below and the data used in the example are provided in Figure 3.2.
<table>
<thead>
<tr>
<th>January X</th>
<th>February Y</th>
<th>Complete</th>
<th>MCAR</th>
<th>MAR</th>
<th>MNAR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>169</strong></td>
<td>148</td>
<td>148</td>
<td>148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>123</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>132</td>
<td>149</td>
<td>-</td>
<td>-</td>
<td></td>
<td>149</td>
</tr>
<tr>
<td><strong>160</strong></td>
<td>169</td>
<td>-</td>
<td>169</td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>138</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>116</td>
<td>102</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>125</td>
<td>88</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>112</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>133</td>
<td>150</td>
<td>-</td>
<td>-</td>
<td></td>
<td>150</td>
</tr>
<tr>
<td>94</td>
<td>113</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>109</td>
<td>96</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>109</td>
<td>78</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>106</td>
<td>148</td>
<td>-</td>
<td>-</td>
<td></td>
<td>148</td>
</tr>
<tr>
<td><strong>176</strong></td>
<td>137</td>
<td></td>
<td>137</td>
<td></td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>155</td>
<td>-</td>
<td>-</td>
<td></td>
<td>155</td>
</tr>
<tr>
<td>131</td>
<td>131</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>130</td>
<td>101</td>
<td>101</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td><strong>145</strong></td>
<td>155</td>
<td>-</td>
<td>155</td>
<td>155</td>
<td></td>
</tr>
<tr>
<td>136</td>
<td>140</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td><strong>146</strong></td>
<td>134</td>
<td>-</td>
<td>134</td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>129</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>97</td>
<td>85</td>
<td>85</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>134</td>
<td>124</td>
<td>124</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td><strong>153</strong></td>
<td>112</td>
<td>-</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>118</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>137</td>
<td>122</td>
<td>122</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>101</td>
<td>119</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>103</td>
<td>106</td>
<td>106</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>78</td>
<td>74</td>
<td>74</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td><strong>151</strong></td>
<td>113</td>
<td></td>
<td>113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td>125.7 (23.0)</td>
<td>121.9 (24.7)</td>
<td>108.6 (25.1)</td>
<td>138.3 (21.1)</td>
<td>153.4 (7.5)</td>
<td></td>
</tr>
</tbody>
</table>

Source: Missing Data: Our View of the State of the Art
Schafer JL and Graham JW (2002) adapted from Table 1 Simulated Blood Pressure Measurements (N = 30 Participants) in January (X) and February (Y) With Missing Values Imposed by Three Different Methods

Figure 3.2 Example of the Missing Data Mechanisms

Suppose the systolic blood pressures (SBP) of 30 participants are recorded in January (X). Some of them have a second reading in February (Y), but others do not.
To explain the missingness concepts Schafer and Graham (2002) used the following scenarios:

- **Missing completely at random (MCAR)** – 7 participants are randomly selected from original sample to return for a follow-up SBP measurement in February.

- **Missing at random (MAR)** – participants return in February because their January measurement exceeded 140 (X>140), a level used for diagnosing hypertension.

- **Missing not at random (MNAR)** – all participants return in February, however those recorded in February were only those whose measurements exceeded 140 (Y>140).

In certain situations such as planned missingness, MAR is known to hold (Schafer & Graham 2002). Data are missing because there was never any intention for them to be collected. For example, some longitudinal studies have in the past used multiple questionnaires containing different subsets of items (Schafer & Graham 2002). Normally in planned missingness MCAR is the case but some instances like the example from Schafer and Graham (2002), data can be missing at the post-test if the pre-test was not above a cut-off value.

In Figure 3.2, there are a number of points which are worth noting. Firstly moving from MCAR to MAR to MNAR, the observed values become very select and compared to the population (complete Y’s column) the means and standard deviations differ. In Figure 3.2, it can be seen clearly that the sample means for MCAR, MAR and MNAR are increasing, while the respective standard deviations are
decreasing. Secondly, the proportion of missing data is relatively high in all three examples and thus the choice of missingness will exert influence over the results.

Finally, on the definition of Missing Not At Random, Post et al (2010) provide an example of MNAR in the case of a randomized controlled trial. They focus on the term drop-out to discuss the missing data (drop-out) mechanisms and they define MNAR as a situation when the probability of drop-out is dependent on earlier observed measures and also on the unobserved current and future responses measures (Post et al 2010). For example, a participant has a sudden drop in HRQOL scores and decides not to fill in the questionnaire and drops out of the randomized controlled trial. The reason for the sudden decrease in HRQOL scores could be that the participant has experienced a relapse of a tumour, or that the participant is actually dying. In this case the drop-out (relapse or death) might be informative (Post et al 2010). In the next section the issue of ignorable and informative will be discussed.

3.1.2.5 Ignorable and Informative (Non-ignorable) Missing Data

For a correct interpretation of the results of longitudinal data analysis two issues must be considered. Firstly, it is important to investigate whether or not missing data on the outcome variable Y at a certain time point are dependent on the values of the outcome variable observed one (or more) time points earlier (Twisk 2003). Secondly, it is important to determine whether or not the predictor X variables are related to the missingness of the outcome variable Y data. Hence it is more important to make the distinction between what is considered as ignorable and non-ignorable missing data (Twisk 2003). Ignorable missing data are considered as missing data that are not dependent on earlier observations and predictor variables. Another term used for ignorable is “non-informative”. Non-ignorable missing is when missing data are dependent on earlier observations or predictor variables. Another term for non-
ignorable is informative missing data. Diggle and Kenward (1994) referred to MAR as non-informative or ignorable drop-out, while MNAR was called informative.

Twisk (2003) recommends an investigation when the amount of missing data is high for repeated measures, to determine whether the missing data are informative or non-informative. The technique suggested by Twisk involves logistic regression analysis and will be discussed and applied to ALSWH data in Chapter 5.

3.1.2.6 Implications of Missingness

Statistical methods are usually driven by the assumption that the data are randomly sampled from a population distribution $P(Y_{\text{com}}; \theta)$, where $\theta$ represents unknown parameters (Schafer & Graham 2002). There are two very different interpretations that can be taken and are known as Likelihood and Bayesian approaches.

In the likelihood interpretation, the realised value of $Y_{\text{com}}$ is substituted into $P(Y_{\text{com}}; \theta)$ and the function for $\theta$ summarises the data’s evidence about parameters (Schafer & Graham 2002). When $Y_{\text{com}}$ has missing data, there is the inclination to ignore and focus on what is observed $Y_{\text{obs}}$. The following distribution can be obtained using calculus as a definite integral of $P(Y_{\text{com}}; \theta)$ with respect to $Y_{\text{miss}}$,

$$ P(Y_{\text{obs}}; \theta) = \int P(Y_{\text{com}}; \theta) \, dY_{\text{miss}} $$

(Rubin 1976) realised a problem with equation 3.4 in relation to missing data. Rubin found that it was not always true that equation 3.4 is the correct distribution for $Y_{\text{obs}}$ and the correct likelihood $\theta$ based on $Y_{\text{obs}}$. For equation 3.4 to be the correct sample distribution, the missing data should be MCAR. For equation 3.4 to be the correct likelihood the missing data can be MAR (Schafer & Graham 2002). The case for MNAR will be discussed in Chapter 5 where an application using data will be considered.
Finally, on missingness there is a situation where a missing observation causes it to leave the universe of interest or in other words be of no relevance to the participant (Schafer & Graham 2002). An example could be where a participant is asked how well do you get along with your siblings and the participant does not have siblings. In this case the hypothetical missing data type can be considered as MAR (Schafer & Graham 2002).

In a further extension of this discussion, Schafer and Graham (2002) suggest that when participants die in longitudinal studies then outcome measures of physical or mental status only have meaning for those still alive. So participants who die are outside the universe of interest thus those with missing data due to death may often be regarded as MAR. “However, if the measures are spaced far apart in time - for example, if they are taken annually - then death of a participant may provide indirect evidence of an unmeasured steep decline in the outcome prior to death and response trajectories estimated under an MAR may be somewhat optimistic” (Schafer & Graham 2002). In this particular situation, Schafer and Graham (2002) suggest that joint modelling of the outcome and death events should be considered, as first put forward by Hogan and Laird (1997).

3.1.3 Dealing with Missing Data

In the following sections, the issue of what to do about the missing data once identified as possibly ignorable or informative will be discussed. Generally, there are a number of alternatives, some of which may be considered inappropriate however they will still be mentioned in this section since they are being applied in the literature:

i. remove participants with missing data (Complete Case analysis);

ii. analyse with the available measurements (Available Data Analysis);

iii. impute for the missing data then analyse complete data sets.
3.1.3.1 Complete case analysis

One of the original approaches to missing data was simply to delete a participant (case or unit) whose data was incomplete. This approach has many names; case deletion, listwise deletion and complete case analysis and is often the default in many statistical programs (Schafer & Graham 2002). The main reason why case deletion is used is for simplicity. When only a small proportion of data are missing (5% or less, Schafer 1997) and the assumption of missingness is considered MCAR and the cases being discarded do not heavily impact on findings then this approach is considered appropriate (Schafer 1997). However, MCAR is rarely the case and so complete case analysis is not a preferred approach, especially since the advancements in statistical software which use more sophisticated techniques (Twisk 2003). Thus when missing data are not MCAR complete case analyses may provide biased results.

3.1.3.2 Analysis of Available Data

The current approaches allow for analysis of available data, that is, the dataset can be unbalanced. Current statistical methods of longitudinal data no longer require a complete case data structure. Two methods that are being applied are Generalised Estimating Equations (GEEs) and random coefficient analysis methods (random effects). The GEEs approach can be applied to continuous, discrete and binary outcomes. This method developed by Zeger and Liang (1986) is considered a semi-parametric method which models marginal or population averaged relationships between an outcome and covariates (predictors). GEEs assume that the missing data are missing completely at random (MCAR). Therefore using GEEs to analyse an incomplete dataset where the assumption may be missing at random would produce a biased working correlation structure. Therefore the regression coefficients that are calculated would also be biased (Twisk 2003).
Random (effects) coefficient models can be used to model longitudinal data where there is interest in subject-specific relationships. Thus random coefficient models can consider subject-specific growth curves where the intercepts and slopes for subjects can be random. Random coefficient models can also have just random intercepts. In most statistical software the random coefficient models use Maximum Likelihood Estimation (MLE) and Restricted Maximum Likelihood (REML). These modelling approaches will be discussed and applied to real data in later Chapters. Using a random coefficient model when cases are incomplete assumes that the missing data mechanism is missing at random (MAR) (Little 1995, Twisk 2003).

3.1.4 Developments in Imputation Theory and Methodology

Over the past two decades the imputation theory and methodology in the literature has grown along with statistical software (Carrigan et al 2007). Raghunathan (2004) and Ibrahim et al (2005) suggest that there are several theoretical approaches to imputation and they identify four classes: Weighted Estimating Equations (WEE), Multiple Imputation (MI), Likelihood-based formulations and Fully Bayesian (Carrigan et al 2007). The following sections will firstly discuss the simple imputation approaches that are losing favour. There will then be discussion with a brief overview of the four classes mentioned by Raghunathan (2004) and Ibrahim et al (2005). Finally, in this section the advantages and disadvantages of these methods will be mentioned and whether they may be applied to the ALSWH data.

3.1.4.1 Single Imputation

Rather than use case deletion, a single imputation method involves the replacement of a missing value in a dataset. This dataset may be cross-sectional or longitudinal. Thus, for a single imputation of missing data the methods can be classified as cross-sectional and longitudinal imputation methods (Twisk 2003). Examples of the more
popular cross-sectional imputation are ‘mean substitution’, ‘hot-deck’ and ‘cross-sectional linear regression’ methods (Twisk 2003). For longitudinal imputation methods the most regularly (possibly misused) approach is the ‘last value (observation) carried forward’ (LVCF) and a longitudinal linear regression method (Twisk 2003). Not all cross-sectional single imputation will be discussed in this Section. The focus will be: mean substitution, hot deck imputation and linear regression. Schafer and Graham (2002) mentioned that the main problem with single imputation approaches is “there is no simple way to reflect missing data uncertainty”. They go on to suggest “the problem of undercoverage is solved by multiple imputation” which will be discussed later.

3.1.4.2 Averaging available items

A common problem with questionnaires is that items are often missing. Usually the items are combined in some way to form a scale of measurement, for example, anxiety or depression or physical health. An approach which is often employed by researchers when items are missing is to use averaging. The available items are averaged and under certain criteria (e.g. must not have more than 2 items missing) the missing item is replaced by the average of available items. This approach has been used to calculate the eight subscales of the SF-36 health survey when items are missing. Attempts by Rogers et al (2004) to improve on this approach will be mentioned later. The concern of using the very simple averaging approach is that it may introduce bias under MCAR. Also, the variance of the scale can increase and thus the scale can become less reliable, although there are concerns it is still considered a reasonable approach to deal with missing items (Schafer & Graham 2002).
3.1.4.3 Mean substitution

The use of averaging of available items is a simple form of the mean substitution approach. Most mean substitution approaches are relatively simple to apply and are often used in the literature, but mean substitution can produce ‘strange’ misleading results (Twisk & de Vente 2002; Twisk 2003). In the following section a few of the mean substitution methods will be discussed. Schafer and Graham (2002) classify the methods into the following categories:

i) unconditional mean substitution – Hot deck imputation;

ii) conditional mean substitution – Linear regression.

3.1.4.4 Hot Deck Imputation

The hot deck approach is used to impute values for missing data from the observed data. In a simple hot deck, a random draw from the observed values is used to replace the missing values for non-responders. The approach partly explains the understating of uncertainty but the approach still distorts measures of associations such as correlations (Schafer & Graham 2002).

3.1.4.5 Linear regression

In this approach a linear regression is used to fit a model for observed values Y against possible observed covariates. The predicted values for Y are used to replace missing values for Y with similar characteristics to the observed data. This approach is called a conditional mean imputation because the imputation of the predicted value for Y is conditional on X. In a standard linear regression predicted Y’s may also have an additional residual error added. The use of logistic regression is also possible for binary dependent variables. The main issue with either method (hot deck or linear regression) is they can produce biased estimates for many parameters whatever type of missingness (MCAR or MAR) is considered (Schafer & Graham 2002).
3.1.4.6 SF-36- Approaches to imputing missing PCS and MCS

The Missing Data Estimates (MDE) approach of imputation was developed and applied to the SF-36v2 (Rogers et al 2004b). The MDE method is based on extensions to Item Response Theory for dealing with multivariate concepts. The MDE has been modified on a few occasions but the details of the method are the proprietary of QualityMetric™ and the documentation on the MDE approach is limited (Rogers et al 2004).

Another imputation approach known as the Regression Estimates (RE) approach has been developed and can be used with SF-36v2. The Regression Estimates (RE) approach produces “regression estimates that are based on breaking each item down into a set of indicator variables for the various responses”. These dummy variables are then used as independent variables in a linear regression with PCS (and MCS) for available items (Rogers et al 2004b).

Rogers et al (2004b) modified the regression estimates (RE) approach. Their approach known as the Modified Regression Estimates (MRE) is an improvement on the earlier methodology. They found with the earlier approach that the regression estimates are pulled towards the mean and may produce bias if extended to other populations (Rogers et al 2004b). The modification they suggest to correct for this regression-to-the-mean effect is:

\[ Y_{\text{modified}} = \text{(average)} + \frac{(Y_{\text{regression}} - \text{average})}{R} \]  

(equation 3.5)

where R is the square root of \( R^2 \) from the linear regression model

Kazis et al (2007) suggest that the use of the MRE approach is able to “recover” two thirds of the missing cases for PCS which would still have been missing after the Missing Data Estimates (MDE) approach. They suggest that both approaches are an improvement that may reduce bias and provide more variation than the traditional approach applied using the SF-36v1 (Kazis et al 2007).
Traditionally, to calculate the SF-36 scales, missing data were imputed according to the recommended algorithm in the SF-36 Users’ manual (Ware & Kosinski 2001), “a person specific estimate for any missing item is substituted when the respondent answered at least 50% of the items of a subscale”. The Physical and Mental Component summary scales are then calculated from the subscales and are set as missing if the respondent is missing any one of the eight SF-36 subscales. This traditional approach has been used for the data in the ALSWH SF-36v1. Since the ALSWH uses SF-36v1 the MRE approach will not be applied to these data. Other imputation approaches discussed later in this chapter will be considered and applied in Chapter 6 and 7 to account for missing PCS data.

3.1.5 Single imputation – longitudinal methods (LVCF)

The use of single imputation has also been used in longitudinal studies and a popular approach is the last value carried forward method (LVCF) or also called the last observation carried forward (LOCF). In this approach the last observed value recorded for a particular item or measurement is carried forward to the next wave or panel. The main concern about using single imputation is that one single replacement of the missing item (value) does not truly reflect the uncertainty which needs to be considered about the missing data (Schafer & Graham 2002). If the sudden decline in health was the reason for dropping out of a longitudinal study then the LVCF approach would not reflect this sudden decline in health.

3.1.6 Weighting Methods

Weighting methods is an approach to account for missing predictor data (Robins et al 1995; Xie & Palik 1997; Horton & Lipstiz 1999, Horton & Lipsitz 2001; Horton & Kleinman 2007). This approach models the probability of missing, then the inverse of these probabilities are used as weights for the complete case data (Horton &
Kleinman 2007). This approach attempts to weight the remaining cases so that the
distribution is more closely to that of a full population. Reweighting is useful when
monotone missing data patterns arise but it is more complicated to apply when
missing patterns are arbitrary (Schafer & Graham 2002).

Weighted estimating equations (WEE) weight the complete case data to compensate
for the missing cases (Carrigan et al 2007). There have been improvements made to
this approach; however, the true variance in the data can be underestimated when
the variance is unadjusted. The implementation of the approach uses model-specific
algorithms and is not standard in statistical packages (Carrigan et al 2007). The WEE
is an extension of the semi-parametric GEEs and can accommodate for missing
values as MAR or possibly MNAR provided a missingness model is correctly
specified (Schafer & Graham 2002). However, the WGEEs semi-parametric
estimates are more likely to be less efficient and less powerful than Maximum
Likelihood or Bayesian estimates under a specific parametric model (Schafer &
Graham 2002).

3.1.7 Multiple Imputation

Multiple imputation, developed by Rubin (1987) replaces each missing value with a
set of values that represent the uncertainty about the unknown value to impute
(Rubin 1976; Little & Rubin 1987; Rubin 1987; Laird 1988; Robins et al 1995; Rubin
1996; Schafer 1997; Allison 2000; Yang 2000; Diggle et al 2002; Twisk 2003). The
multiple imputation method generates multiple values for every missing value. These
multiple values (say M=5) are derived from a selected statistical method (such as
linear regression) and are stored in M complete datasets. The M complete datasets
can be combined to form a summary statistic (Twisk 2003). The estimate of the
summary statistic is calculated by averaging the M imputations.
Methods of imputation of missing data are well documented (Schafer & Graham 2002; Horton & Kleinman 2007), and multiple imputation has become the recommended approach to replace missing data (Rubin 1976; Schafer & Graham 2002; Klebanoff & Cole 2008). There are different approaches to implementing MI and one particular approach that will be discussed and applied in later chapters is the use of Chained Equations (van Buuren et al 1999; Raghunathan et al 2001; van Burren 2006). However, the current multiple imputation techniques do have limitations (Schafer & Graham 2002). These limitations include: the use of cross-sectional imputation and the assumption of missing at random (MAR) (Schafer 1997; Schafer & Graham 2002). These two limitations will be addressed in Chapters 5, 6 and 7 since the analyses of the ALSWH data may need to consider the assumption of MNAR and longitudinal missing data.

3.1.8 Likelihood-based Approaches

Maximum Likelihood (ML) is an alternative approach to using Multiple Imputation (MI) which also assumes missing at random. The regression parameters are affected by missing predictor variables and information can be retrieved by estimating the distribution of covariates using an expectation-maximization (EM) algorithm developed by Dempster et al (1977). When an “observation is missing, then multiple entries are created in an augmented dataset for each possible value of the missing covariates and a probability of observing that value is estimated given the observed data and current parameter estimates” (Horton & Kleinman 2007).

A major difference between Maximum Likelihood and Multiple Imputation is that likelihood methods deal with missing data during the modelling process while MI considers missing data prior to the modelling (Schafer & Graham 2002). There are number of disadvantages of Maximum Likelihood approaches to imputation according to Carrigan et al (2007), such as they are “often intractable” in the popular statistical
software and their “implementation relies upon strict assumptions about patterns of missingness that are frequently violated in complex survey data”.

### 3.1.9 Bayesian Approaches

Multiple imputation was derived from Bayesian theory, that is sampling from the posterior distribution of interest (Horton & Kleinman 2007). Fully Bayesian approaches are being applied and Ibrahim et al (2005) described the close relationship between Bayesian approach and ML and MI methods (Horton & Kleinman 2007). The method is flexible but requires specific coding of prior distributions and model relationships (Horton & Kleinman 2007). According to Carrigan et al (2007), “Fully Bayesian (FB) models extend Multiple Imputation methodology by jointly simulating the distributions of variables with missing data as well as unknown parameters in a regression equation”. Fully Bayesian techniques like Maximum Likelihood techniques depend on fully specified models, but differ in that parameter estimates are generated using Bayesian simulation rather than likelihood-based approximations (Carrigan et al 2007).

The appeal of the Fully Bayesian technique in this thesis is that the analysis and imputation models are fully and simultaneously specified (Carrigan et al 2007). Carpenter et al (2002) and Carpenter and Kenward (2005) have presented examples of missing data models (MNAR and MAR) using WinBUGS. These examples will be explored and discussed in greater detail in later chapters. In Chapter 5, a Fully-Bayesian approach that allows for the MNAR assumption and longitudinal multiple imputation (Carpenter & Kenward 2005; Carrigan et al 2007) will be developed and applied.
3.1.10 Pros and Cons of imputation methods

The purpose of this overview was to frame the missing data problem and briefly describe the imputation methods which are currently being applied. Analysing complete case data can produce biases, such that responders to surveys may be different to non-responders. Imputation is a way of compensating for the missing data. Single imputation methods are viewed as not going far enough because there is often a lack of uncertainty about the unknown values. Multiple imputation is an improvement because a set of values can be produced for the unknown value. For this reason multiple imputation will be used in later chapters. There has not been a great deal of work on imputation in longitudinal studies (Carrigan et al 2007). A number of possibilities will be explored later in Chapters 5 and 6. Four theoretical approaches to imputation suggested by Raghunathan (2004) and Ibrahim et al (2005) were mentioned. Some approaches such as WEE are not easy to implement and are not part of the commonly used statistical software packages. There is also the need to consider the missingness assumptions. ML approaches to imputation and MI usually assume MAR or MCAR. In a longitudinal setting there is the possibly of non-ignorable (informative) missing and thus imputation approaches may be limited when MAR cannot be assumed. Of the imputation approaches mentioned earlier, Bayesian methods offer an opportunity for further exploration in longitudinal studies. The following will briefly mention current methods to deal with non-ignorable missing data. They will also be discussed later in Section 3.2.2 of the literature review.

3.1.11 Non-ignorable missing data methods

When the missingness law cannot be simplified and is considered non-ignorable, more complex models are required which incorporate the missing data mechanism in the analysis (Kaciroti et al 2008). There is an extensive literature regarding regression models with non-ignorable (informative) missing outcomes (Horton &
Fitzmaurice 2002). Researchers have addressed informative non-response which can be an issue for clinical trials. In this non-ignorable setting, correct specification of the missingness law must be given to obtain consistent regression parameters (Horton & Kleinman 2007). Non-ignorable issues and model types have been addressed and discussed by Diggle & Kenward (1994), Little (1994), Little (1995), Kenward (1998) and Verbeke & Molenberghs (2000). In the second part of the following literature review the methods being developed and applied for non-ignorable (informative) missing data will be discussed. There will be discussion as to which approach will be considered to deal with the ALSWH missing data, later in Chapters 5, and 7.

3.2 Literature Reviews

In the first part of this chapter important missing data concepts and methodology were outlined as an introduction to the following literature reviews. This next part of the chapter consists of two main sub-sections. The first sub-section is a comprehensive review of the literature on the current methods to account for deaths in health related quality of life longitudinal studies of elderly populations. The second sub-section is a review of the literature on the current methods and advancements for analysing longitudinal studies with missing data.

Chapter 2 provided a background of all current and most commonly used Health Related Quality of Life (HR-QOL) measurement tools currently being used. The background also discussed the most common Health Utility Indexes being administered. The first review of the literature will focus on the HR-QOL tool known as the SF-36 Health Survey (Ware 1993). The SF-36 Health Survey is used by the ALSWH and this study is the major focus of this thesis.
Over the course of the study, ALSWH researchers have noticed a number of issues with the SF-36 that need addressing. For this thesis the issue of deciding how to analyse the data with such a high frequency of deaths occurring in the older cohort (aged in 1996: 70-75 years) will be considered. The next part of this chapter will focus on the developments in relation to how other researchers are attempting to account for deaths in longitudinal studies using the SF-36 health survey. As the three ALSWH cohorts continue to age the issue of accounting for deaths is becoming of greater concern, especially for the middle aged (aged in 1996: 45-50 years) and the older cohort.

### 3.2.1 A Review of the Literature – Accounting for Deaths

Diehr et al (1995) suggest that “accounting for deaths in longitudinal health surveys can be viewed as a problem of missing data”. However, most statistical methods mentioned in Section 3.1 assume that the missing data mechanism is non-informative, that is ignorable. This assumption is not likely to be the case “because death is usually highly related to health” (Diehr et al 1995).

In this following sub-section the review will consider methods that are accounting for deaths. The main interests in this body of work are methods that specifically deal with ways to analyse longitudinal studies of ageing cohorts when deaths have occurred. Some methods will be briefly discussed but are not part of the intended direction of this thesis. As will be pointed out there has been greater attention to censored (by death) data in clinical trials.

#### 3.2.1.1 Statistical Methods in HR-QOL research for the treatment of Deaths

A review of the literature from 1960 until September 2010 was conducted. There was no limit placed on the dates of the searches. Although the SF-36 was not developed until the 1990s, there were earlier HR-QOL tools available. The electronic search of
literature was conducted using Medline, PubMed and Embase databases. The searches of journal article titles and MeSH terms used in words such as: “SF-36”, “HR-QOL”, “death” or “mortality”, “longitudinal studies”. The MeSH terms were combined using Boolean logic terms: “and”, “or”. The review was extended to include both cohort studies and clinical trials and even unpublished papers. The search was not restricted to articles published in English.

From the comprehensive review of the literature, there were 8887 articles which mentioned the term “SF-36”, whereas a combined search for “HRQOL and SF-36” resulted in 1181 articles. Refining the searches to a combination of specific terms found: 41 articles for “SF-36” and (“death” or “mortality”) and “longitudinal”. Of these 41 articles, twelve articles mentioned a method (i.e. other than survival analysis) for how they accounted for deaths in their longitudinal study, and those twelve articles are mostly from the same authors (i.e. Diehr and colleagues).

Very little work in statistical methodology has addressed missing data when deaths occur during follow-up (Kurland et al 2009). Kurland et al (2009) cited the following papers which discussed attempts to account for deaths: Revicki et al (2001), Diehr et al (2003); Pauler et al (2003), Dufouil et al (2004), Kurland & Heagerty (2005), Rubin (2006), Harel et al (2007) and Egleston et al (2007). Of these attempts only the work by Revicki et al (2001) and Diehr et al (2003) was directly applied to the SF-36. Kurland et al (2009) examined these approaches and “how models can arise from the statistical distribution of longitudinal data and survival information”. They also gave guidance on appropriate analysis techniques. Some of the articles (Diehr et al 2003 and Revicki et al. 2001) suggested by Kurland et al (2009) are of great interest in this thesis and will be discussed later in Section 3.2.1.3. The other authors mentioned will be briefly discussed later in this chapter but are not part of the intended direction of this thesis.
The issues associated with correctly estimating a causal effect of a treatment (in a clinical trial) on an outcome when data are censored by death was discussed by Rubin (2006). It is an important issue for randomized clinical trials and does deal with the issue of drop-outs, however, the causal effect of a treatment on an outcome is not the direction of this thesis. In passing, Rubin (2006) does make the comment that “comparing quality of life when it is observed and dropping people who died although popular in some settings, is simply wrong in general”.

In another approach based upon earlier work by Rubin (2000) involving survivors’ average causal effect (SACE), Egleston et al (2007) proposed a sensitivity analysis. This involved a procedure for drawing inference about the estimand (SACE) in the context of observational studies with missing outcomes among observed survivors (Egleston et al 2007). This methodology within a sensitivity analysis framework involves an understanding of the concept of principal stratification. As such, the concept of principal stratification and the sensitivity analysis is not of interest and thus the approach will not be explored in this thesis.

The majority of work which accounts for deaths in longitudinal studies can be accounted to Diehr and colleagues. Table 3.1 is a summary of eight strategies which Diehr and colleagues developed and presented in 1995. Many of the earlier strategies suggested by Diehr et al (1995) and Diehr et al (2001) will be discussed in this chapter. Table 3.1 will be explained in greater detail in Section 3.2.1.5.

Two later approaches by Diehr and colleagues not shown in Table 3.1 are the use of trajectories of health over time to account for deaths (Diehr et al 2003b; Diehr et al 2007). When considering the self-perceived quality of life near the end of life plotting the trajectories of health related variables over time may be useful. Often when
graphs are used the data only include survivors and they can confuse patterns of mortality with change in health (Diehr et al. 2003b). Diehr et al. (2003b) proposed two approaches that use graphs to incorporate deaths. The first approach has been applied to the SF-36. Death is incorporated as a category into the perceived health status - EVGGFP variable and then stacked bar graphs are generated over each time point. The proportions of participants, in each category, are shown over the course of the study.
<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
<th>Scores for each category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Remove deaths - Available case analysis</td>
<td>Uses all participants at baseline and removes all deaths at follow-up. The two sample sizes are different.</td>
<td>Dead</td>
</tr>
<tr>
<td>2. Remove deaths - Complete case analysis</td>
<td>Remove all participants who died from both baseline and follow-up. The sample sizes are equal and all alive.</td>
<td>NA</td>
</tr>
<tr>
<td>3. Healthy (yes/ no)</td>
<td>Recode scores from T0 and T1 as 100 if healthy 0 otherwise (deaths included)</td>
<td>0</td>
</tr>
<tr>
<td>4. Better/same/worse</td>
<td>Recode data as better, same, worse for changes from T0 to T1 +1 =better, 0 =same, -1 = decline (includes death)</td>
<td>0</td>
</tr>
<tr>
<td>5. Death arbitrary</td>
<td>Deaths were given a value of 0 for all variables representing worst health status</td>
<td>0</td>
</tr>
<tr>
<td>6. Prob (healthy) - Probability of being healthy in 2 yrs</td>
<td>Recodes values to approximate the prob. of being healthy in future. Regression used to get probabilities of being healthy at T1</td>
<td>0</td>
</tr>
<tr>
<td>7. Prob (Alive) - Probability of living 2 more years</td>
<td>Recodes values to approximate the probability of being Alive in future Regression used to get probabilities of being Alive. Emphasis on death</td>
<td>0</td>
</tr>
<tr>
<td>8. Alive (yes/no)</td>
<td>Recode scores from T0 and T1 as 100 if alive 0 if dead</td>
<td>0</td>
</tr>
</tbody>
</table>
The second graphical approach measures time before death. Diehr et al (2003b) suggest this graphical approach “accounts for death and is useful for descriptive studies of the health related variable before an event”. The method allows for death to be incorporated into the time variable by measuring time backward from the date of death. Diehr et al (2003b) provide an example of how graphing the mean Mini-Mental State Examination scores decreased in the years before death for both men and women.

The trajectory approach is more of a descriptive method and not meant to be used a predictive method. Also, the date of death must be known to determine a person’s place on the curve. Diehr et al (2007) have modified and applied the approaches but have recommended that the approaches should be used to support analyses in longitudinal studies in which deaths have occurred (Diehr et al 2003b).

Revicki et al (2001) conducted a simulation study to evaluate the bias associated with four different imputation methods. They were estimating physical health status (PHS) scores missing due to death. One imputation approach they considered was the last value carried forward which was described earlier in this chapter. A second approach they called arbitrary substitution had been suggested by Diehr et al (1995). This method imputed missing PHS scores by assigning the previous value minus a decrement to account for the fact that the current value was missing due to death. A third approach was called the Within-Subject modelling approach. This approach used ordinary least squares regression as its basis (Revicki et al 2001). The fourth approach involved an empirical Bayes approach. Revicki et al (2001) concluded that the Bayes approach to imputation was the most appropriate. Also, the Revicki et al (2001) simulation model approach has been highlight by Kazis et al (2007).
3.2.1.2 Kazis et al’s White Paper with recommendations for Medicare Health Outcomes Survey (HOS)

A white paper was commissioned by (for their internal use) the Centers for Medicare and Medicare Services (CMS) in the US. The aims of the white paper were to consider the current approaches used for monitoring and comparing plans for the Medicare Advantage program in the Health Outcomes Survey (HOS) and to make recommendations for future research (Kazis et al 2007). In this unpublished internal report, special consideration was given to several areas of measurement of HRQOL including scoring, mortality, case-mix, treatment of missing values and norms (Kazis et al 2007). Permission to discuss their findings was granted by Professor Lewis Kazis through personal email correspondence (January 8, 2008).

3.2.1.3 Current approaches to account for deaths in HR-QOL research

There are at least four main strategies for account for deaths that have been suggested by Kazis et al (2007) and some have appeared in the literature (Fihn et al 2004; Strandberg et al 2004; Trisolini et al 2005; Diehr et al 2007; Selim et al 2007). All strategies are very useful for HR-QOL research especially since they all focus on the SF-36. However, in all strategies little has been done to extend and apply the approaches to longitudinal studies of the elderly. Kazis et al (2007) provide an overall summary of the current approaches to accounting for death and their recommendations will be discussed in Section 3.3. The main strategies/methods for accounting for deaths mentioned by Kazis et al (2007) were:

1. Health Outcome Survey Approaches to deaths (Better/Same Vs Worse/Death).
2. Diehr et al's approaches – variable transformations to incorporate deaths.
3. Erickson’s HALex approach.
These strategies will be considered in greater detail and adapted in Chapters 4, 6 and 7. A brief explanation of each method will be discussed in the following section.

3.2.1.3 i) Transforming Self-Rated Health to Include Death

Diehr et al (2001) suggested a number of strategies to account for deaths. All eight strategies will be discussed in Section 3.2.1.5. Two of the suggested approaches have been discussed by Kazis et al (2007) and were considered as being quite reasonable.

One approach relates to the self-rated health question, which asks participants to rate their health as: Excellent, Very Good, Good, Fair or Poor (EVGGFP). Thus the self-rated health question is then dichotomized to: “Healthy” – Excellent, Very Good and Good; “Not healthy” – Fair, Poor and Dead. In this method the follow-up survey is dichotomized and then those who have died before the follow-up survey are included in the “not healthy” category.

The second approach again considers the self-rated health (EVGGFP) and transforms the actual observed state (EVGGFP) into a probability of being healthy in the future (PHF). Diehr and colleagues developed a number of strategies to assign values to each category of self-rated health. The basis of the probability of being healthy in the future is estimated from the regression analysis and each category of self-rated health was assigned a value (probability). The value was assigned as estimated probability of being healthy in the future (PHF). Thus the participant who was self-rated as excellent would be coded as a probability score of 95. That is, the probability of being healthy in the future (in one or two years) for an individual rated excellent was 95%, whereas, a person who had died would receive a value of zero. Other self-rated health values considered were as follows: Excellent (95), Very good (90), Good (80), Fair (30), Poor...
(15), Dead (0). So in this approach there are a set of probabilities and the probability of being in better health for a participant who dies is zero.

3.2.1.3 ii) Transformation of the SF-36 PCS

The PHF approach by Diehr et al (2001) has also been developed for the SF-36 PCS and MCS. The transformation using PCS is known as the Probability of being healthy in one year (PCT). This approach used a logistic regression where the binary dependent variable was the “healthy” variable at follow-up and the independent variable was the baseline PCS. Thus if a participant dies before the follow-up survey one year later, the probability of being healthy in one year is zero. This approach is of great interest here and will be explained and explored further in Chapter 4.

3.2.1.3 iii) HALex Approach

The HALex questionnaire developed by Erickson (1998) has been used to evaluate Activities of Daily Living (ADL). Trisolini et al (2005) regressed PCS scores on the health utility values and extrapolated the scores corresponding to a utility value of zero (death). Their analysis produced a score for death equivalent to 14 points for the PCS. For those respondents who died before the follow-up survey they are imputed with a value of 14. In Chapter 6 this approach will be discussed in greater detail and adapted to a different health utility index such as the SF-6D to develop a PCS value for death. The reasons for this adaptation will be discussed in Chapter 6.

3.2.1.3 iv) HOS case-mix methodology

The Health Outcome Study (HOS) case-mix methodology is considered the conditional probability of being Alive with a PCS the Same or Better (Rogers et al 2004). The HOS approach considers the probability of being alive and in same or better health in 2 years. The approach has defined Same or Better health using a cut-off (5.66) (Rogers
et al 2004) value for a change in Physical Component Summary score over two time points. This approach uses a conditional probability by combining the probability of being Alive in two years with the probability of being in Same or Better health in two years.

An alternative approach (Selim et al 2007) used a different structure to the case-mix method (Rogers et al 2004). The alternative approach used a different method to estimating the probability of being Alive than the HOS approach. In both approaches the probability of death is calculated from logistic regression models and then changed to the probability of being Alive by subtracting the probability from 1. Both the Rogers et al (2004) and Selim et al (2007) methods and a description of the term case-mix will be discussed in greater detail in Chapter 7.

3.2.1.3 v) Model Simulation techniques

The fourth approach is that developed Revicki et al (2001) who conducted a simulation study where they varied mortality rates from 0% to 30% and change in Physical Health Status (PHS) scores for a two-group clinical trial with follow-up over 18 months. As mentioned earlier, they used four imputation methods: last value carried forward, arbitrary substitution, empirical Bayes, and within-subject modelling. They concluded that the different imputation methods produced comparable results when there were few missing data. The empirical Bayes approach most closely estimated true population differences and change in PHS regardless of missing data rates. Thus a Bayesian approach to impute for missing data will be considered in Chapter 5.
3.2.1.4 Approaches to Account for Deaths using the SF-36 Survey

As mentioned in Chapter 2, the Medical Outcomes Study Short-Form 36 (SF-36) (Ware 1993) is a commonly used tool for measuring changes in health-related quality of life (HR-QOL) over time. The SF-36 health survey is a 36 item scale which covers the major aspects of health and has been validated with adults of all ages (Ware & Sherbourne 1992; Ware 1993; Ware et al 1994) and has been used extensively in clinical trials (Brazier et al 1998). A limitation of the SF-36 is that it only considers morbidity and not mortality.

The application of the SF-36 to differing populations is quite extensive throughout the HR-QOL literature. However, very few journal articles report on how researchers have dealt with data from participants who died. Many of the reported studies have not been of elderly populations where death is a common occurrence. The analysis of data from longitudinal studies is made more complex by the death of study participants over time. Thus the following sections discuss the attempts that are currently being applied to longitudinal data specifically using the SF-36 health survey with ageing populations.

When analysing longitudinal data, researchers typically ignore participants who die and analyse complete case data, effectively limiting longitudinal analysis to survivors, so that important changes over time in the total cohort may be missed (Diehr et al 2003). To ignore deaths could suggest that the results are only able to be generalized to those who have survived. There is also a possibility of a failure to detect true differences in changes in HR-QOL over time between groups of people. A number of strategies to incorporate mortality developed by Diehr et al (2007) have been adopted and adapted in other studies (Fihn et al 2004; Ostbye & Taylor 2004; Strandberg et al 2004; Selim et al 2006; Selim et al 2007b). The following section will discuss the developments of the earlier eight strategies (shown in Table 3.1) by Diehr and her colleagues.
3.2.1.5 Diehr et al’s eight strategies to account for deaths

Diehr et al (1995) proposed eight strategies for incorporating deaths in the analysis of health status measures. They considered alternative approaches for assigning numeric values to deaths so they could be included in the health status analyses. These eight strategies were developed using general health status measure (EVGGFP) and the Sickness Impact Profile (SIP) and Quality of Well Being (QWB) measurement tools. Since the main focus is the SF-36, only the general health status measure (EVGGFP) will be discussed in detail. These eight strategies were refined and 2 approaches have been since adapted and adopted as the most useful ways to account for deaths. They were discussed earlier. In this section all eight original strategies are presented to show the progression of Diehr and her colleagues’ research. Table 3.1 presents the scoring used in each strategy with respect to the health status measure.

The first strategy was to use all participants’ data at baseline but omit the deaths at follow-up. This meant that the sample sizes were different at baseline and follow-up. This form of analysis is commonly used and is known as “available data analysis”. In the second strategy Diehr and colleagues removed those who died from both baseline and follow-up. This approach is known as complete case analysis. In both strategies the health status measure was coded by Diehr et al (1995) as: Excellent (100), Very Good (80), Good (60), Fair (40) and Poor (20). Since death is often considered as being a worse state than Poor health, Strategy 5 was constructed. This strategy considers death to be an extreme score which is lower than Poor health and is valued as a zero (Diehr et al 1995).

In strategy 3, deaths were included in the analysis by considering those participants who died as “not healthy”. A participant was scored at baseline and follow-up as 100 if
they were “healthy” (Excellent, Very Good or Good) or 0 if “not healthy” (Fair, Poor or Dead). Strategy 8 is similar to Strategy 3 since it uses values of 100 and 0, but in this strategy participants are considered as “alive or dead”. In this case participants are coded as 100 if they reported Excellent, Very Good, Good, Fair or Poor at follow-up and zero for those who had died. Strategy 4, recoded the data as Better, Same or Worse for changes from baseline to follow-up. That is, changes in general health status category (EVGGFP) were coded as +1 = Better or -1 = Worse (including death) while no change in categories were coded as 0 = Same.

Strategy 7 was similar to the approach used in the Medical Outcomes Study. After personal communications with Dr Ware and Dr Rogers, Diehr et al (1995) mentioned that they adapted the MOS approach which was the “probability of living two or more years”. Using data from a health promotion and disease free study, they found that 98% of the participants who reported being in excellent health at the baseline survey were still alive at the follow-up survey. Thus, those in excellent health were re-coded with a value of 98%, and so on for very good, good, fair and poor, while those who died were scored as 0. They found that 76% of participants who reported Poor health at baseline were still alive at follow-up (Diehr et al 1995). Strategy 6 emphasizes “Health” rather than “Being Alive” but uses a similar approach to Strategy 7. In this strategy Healthy is considered as: Excellent, Very good and Good at follow-up survey. Again using the health promotion and disease free study they found that 96% of participants who reported excellent health at baseline were considered “healthy” at follow-up. Likewise 19% of participants who reported poor health at baseline were considered “healthy” at follow-up.

As mentioned earlier, these strategies have also been developed using HRQOL measures such as the SIP and QWB. Using these tools, Diehr and colleagues focused
on using estimated probabilities which were considered as the following equations 3.6 to 3.9. The equations were derived from logistic regressions of SIP (QWB) on Alive (yes/no) or on Healthy (yes/no).

For the Sickness Impact Profile (SIP), the estimated probabilities in percents are as follows:

\[
\begin{align*}
\text{Prob (Healthy)} &= 53.89 + 0.5875 \times \text{SIP} \quad \text{(equation 3.6)} \\
\text{Prob (Alive)} &= 89.04 - 0.227 \times \text{SIP} \quad \text{(equation 3.7)}
\end{align*}
\]

For Quality of Well Being (QWB) the estimated probabilities in percents are as follows:

\[
\begin{align*}
\text{Prob (Healthy)} &= -139 -5.036 \times \text{QWB} + 0.02746 \times \text{QWB}^2 \quad \text{(equation 3.8)} \\
\text{Prob (Alive)} &= 37.1 +1.497 \times \text{QWB} - 0.00916 \times \text{QWB}^2 \quad \text{(equation 3.9)}
\end{align*}
\]

These transformations eventually evolved to other transformations, such as transformations of the PCS and MCS and will be discussed in Chapter 4.
3.2.2 A Review of the Literature – Missing Data

The literature review in Section 3.2.1 focused on: the SF-36 and accounting for deaths. A further review was conducted with a wider focus to include missing data and imputation approaches. Searches of journal article titles and MeSH terms such as: “missing data”, “multiple imputation”, “death” or “mortality”, “longitudinal studies”, “informative”, “ignorable missing”, “MAR” and “MNAR” were considered. The MeSH terms were combined using Boolean logic terms “and”, “or”. The resulting searches of Medline, Embase and PubMed found 13833 articles when the search term was “missing data”. When the PubMed searches were refined to “missing data” and “longitudinal” 706 articles were found. Searching for “multiple imputation” found 586 articles. There were 14 articles found using the terms “MAR” and “MNAR”. Combining the terms “missing data” and “nonignorable” found 52 articles. A search for “missing data” and “dropout” found 196 articles. Whereas using the terms “missing data” and “dropouts” and (“death” or “mortality”) uncovered 22 articles. It became obvious that there is a prominent list of authors in the field of “missing data”. Some of these authors are: Rubin, Little, Diggle and Schafer, all of whom (and others) are discussed throughout this chapter.

Generally the issues related to deaths and other drop-outs are considered throughout the literature and statistical methods are suggested as ways to consider the implications of informative (non-ignorable) missing data. The majority of attempts to deal with informative missing data have been in a randomised clinical trial setting. In the following section there will be discussion of the advancements in dealing with the issue of informative missing data in longitudinal observational studies. The discussion will be supported by examples of where the methods have been applied. The main purpose of the following section is to provide an overview of current methods.
According to Horton and Switzer (2005) there are few reports of how researchers account for drop-outs and missing data. The literature search mentioned above highlights the lack of detail in articles. In 2004 to 2005 Horton and Switzer (2005) reviewed the statistical methods reported in articles of The New England Journal of Medicine, a widely read and highly respected and cited journal. They found from the n=331 articles in that time period, only 26 (8%) papers mentioned a method of dealing with missing data. On further investigation they found that of those 26 articles, 12 had used some form of the Last Value Carried Forward approach, 13 used a mean substitution approach and 2 articles conducted a sensitivity analysis where they replaced missing data with worst case scenario (Horton & Switzer 2005). They also found that two papers had reported using multiple imputation (Smith et al 2004; van de Beek et al 2004). Smith et al (2004) mentioned that they used the multiple multivariate imputation (MICE) approach of van Burren et al (1999). However, neither paper provided sufficient information so that the analysis could be replicated (Horton & Switzer 2005). Also, neither paper addressed the list of required information suggested by Burton and Altman (2004) to replicate analyses with a reportedly large proportion of missing data (Horton & Switzer 2005). Horton and Switzer (2005) concluded that there is still a wide gap between statistical methodologies and methods being applied in practice.

Spratt et al (2010) discussed the strategies for multiple imputation in longitudinal studies. They noted the increased use of MI in the literature but also highlighted the lack of guidelines for end-users of MI. They suggest that when data are missing, careful preliminary analysis should be undertaken to identify the scope for MI to reduce bias and improve efficiency in building the imputation model. Spratt et al (2010) suggest that the preliminary analysis should be reported regardless of whether MI was used in the final analysis.
3.2.2.1 Dealing with Attrition in Longitudinal studies of elderly populations - current discussion

The concept of attrition in longitudinal studies was discussed earlier in Section 3.1.2. A review of the literature was conducted to determine whether longitudinal studies of elderly populations were dealing with issues related to attrition. There have been a number of papers discussing attrition and in particular and with great relevance to this thesis, Young et al (2006) discussed their findings with respect to the ALSWH data. They reported that there were differences in ALSWH responders and non-responders by age and that the types of attrition by age were significantly different (Young et al 2006). Young et al (2006) also suggest that their experiences with longitudinal studies of differing age cohorts may help guide other researchers embarking on a longitudinal study.

Other recent papers which considered the importance of dealing with attrition in elderly populations include: Chang et al (2009); Brilleman et al (2010); Stimpson & Ray (2010) and Vega et al (2010). Brilleman et al (2010) considered the impact of attrition on the older cohort ALSWH data. They concluded that non-death attrition is potentially a greater source of bias in longitudinal studies than death. Stimpson & Ray (2010) considered the attrition of older Mexican American survey respondents. They concluded that withdrawing is reversible and considering only one event of non-participation could give misleading results.

Chang et al (2009) discussed how to minimize attrition bias with reference to a longitudinal study of depressive symptoms in an elderly cohort. They concluded that ignoring drop-outs can yield biased results. They used shared parameter models and suggested they are a powerful and flexible approach for analysing longitudinal data while minimizing bias due to non-random attrition (Chang et al 2009). According to
Yang & Shoptaw (2008) the shared parameter model is useful from the viewpoint of causation. This type of modelling is not a direction in which this body of work is heading and thus will not be discussed further.

Vega et al (2010) discussed several factors that influenced attrition in a population-based elderly cohort with neurological disorders in a Spanish study. They concluded that an understanding of the participants who are at high risk of drop-out may help in the planning of future prospective studies (Vega et al 2010).

The main focus in this thesis is one particular type of attrition (i.e., death). However, it is important to consider all reasons for attrition especially in regards to the missingness mechanisms. The attrition rates and the reasons for attrition of the older ALSWH cohort will be examined and discussed in Chapter 5.

3.2.2.2 Advancements in dealing with the issue of informative (non-ignorable) missing data in longitudinal studies

In this section the focus will be on advancements in dealing with the issue of non-ignorable missing data in longitudinal studies. Dufouil et al (2004) discussed the problem of informative drop-out. There are two main reasons why participants are lost to follow-up: death or drop-out. Participants may drop-out for a number of reasons including declining health. Dufouil et al (2004) acknowledge that likelihood methods give correct results when missing observations can be assumed to be MAR or MCAR and that such approaches require a full probability model. However, they argued that these two types of incomplete follow-up may need to be treated differently in the analysis (Dufouil et al 2004). They proposed to use marginal modelling (GEEs) with an extension to the GEEs methodology to allow inverse probability weights (Dufouil et al 2004). However, they suggest a limitation to this approach at the time was the available
statistical software. An alternative to the extended marginal modelling was to use random effects models to model subject specific trajectories (Dufouil et al 2004). This approach will be considered further in Chapter 5.

Historically, there have been two likelihood-based approaches to analysing data with informative drop-out. These approaches differ with respect to the way the joint distribution of responses and drop-out process is factorised (Curran et al 2002; Michiels et al 2002; Post et al 2010). The approaches are the pattern mixture approach (Pauler et al 2003; Hedeker & Gibbons 2006) and the selection model approach (Diggle & Kenward 1994; Little 1994; Little 1995; Kenward 1998).

The framework for modelling drop out that will be outlined below is adapted from that used by Carpenter et al (2002). They suggest that the complete information on a participant can be considered as a combination of observed and unobserved (missing), components which will be described by superscript obs and miss, respectively.

Suppose $i$ is an element of 1, ..., $N$ individual subjects and that each subject has their responses measured at times $t_{ij}$, $j$ is an element 1 ..., $J$.

Set $t_i=0$

Let the column vector $y_i = (y_{i1}, \ldots, y_{ij})^T$ represent the complete set of primary responses for an individual. Further, let the column vector $x_i = (x_{i1}, \ldots, x_{ij})^T$ denote baseline covariates measured on the $i$th individual. Then the data are represented by $(y_i, x_i)$.

Now let $M_i$ be an indicator variable being 1 if measurement $j$ on individual $i$ is missing and zero otherwise. Hence the dropout vector considered here will be $M_i = (M_{i1}, M_{i2}, \ldots M_{ij})$ which will consist of a set of 0’s a followed by 1’s. Thus the complete data of an individual $(y_i, x_i)$ can be written as $((y_{iobs}, y_{iMiss}), M_i, x_i)$ where the combined length of the vector is $(y_{iobs}, y_{iMiss})$ is $J$ (Carpenter et al 2002).
Let \( P(y, M; \theta, \varphi, x) \) represent the joint likelihood (probability) of the data, given the baseline covariates, at particular values of the parameter vectors.

Where:

\( \theta \) represents a subset of parameters relating to the outcome \( y \) to the covariates \( x \),

\( \varphi \) represents a subset of parameters relating dropout \( M \) to the same covariates \( x \).

Then, using the notation above:

\[
P(y, M; \theta, \varphi, x) = P(y_{\text{obs}}, y_{\text{miss}}, M; \theta, \varphi, x)
\]

(equation 3.10)

Since this likelihood depends on missing data, it cannot be maximised in the usual way. A way around this problem is to calculate from the joint distribution (Equation 3.10), the distribution of \( (y_{\text{obs}}, M, \) ) by integrating out, \( y_{\text{miss}} \) to give:

\[
P(y_{\text{obs}}, M; \theta, \varphi, x) = \int P(y_{\text{obs}}, y_{\text{miss}}, M; \theta, \varphi, x) \, dy_{\text{miss}}
\]

(Equation 3.11)

It is well stated in the literature that this integration is usually very difficult to calculate and ideally would be avoided according to Carpenter et al. (2002). Carpenter et al. (2002) also suggest that if the data are missing for a particular class of reasons, then the integral (Equation 3.11) can be avoided. Taking this further, if we consider the \( i \)th individual alone, then we can consider the joint likelihood (probability) as the product of the likelihoods for individuals and can be written in two forms:

\[
P(y_{i,\text{obs}}, y_{i,\text{miss}}, M_i; \theta, \varphi, x_i) = P(M_i | y_{i,\text{obs}}, y_{i,\text{miss}}; \varphi, x_i) \, P(y_{i,\text{obs}}, y_{i,\text{miss}}; \theta, x_i)
\]

(Equation 3.12)

or as

\[
P(y_{i,\text{obs}}, y_{i,\text{miss}}, M_i; \theta^*, \varphi^*, x) = P(y_{i,\text{obs}}, y_{i,\text{miss}} | M_i; \theta^*, x_i) \, P(M_i; \varphi^*, x_i)
\]

(Equation 3.13)
In a clinical trial scenario such that used by Carpenter et al (2002), Equation 3.12 suggests that the probability of drop-out is dependent on the response of variable, $y$; if a trial patient was to drop-out it would depend on treatment response. That is, patients are selected for drop-out by their response. Thus Equation 3.12 is known as a selection model. Conversely, Equation 3.13 has the probability distribution on the response depending on drop-out status, so that different patterns of response can be proposed for participants who drop out and who continue. This is what is known as pattern mixture models.

Carpenter et al (2002) suggest that both models have their advantages but decided to focus on selection models. Their reasoning was that a selection model was more natural for modelling drop out in their asthma clinical trial. Carpenter et al (2002), suggest that since drop-out is often plausibly explained by steady decline in a patient’s condition to a level at which they did not wish to participate in the trial any more. It could be argued that in a study such as the ALSWH (elderly cohort), that the drop-outs are related to a steady decline in patient’s health due to age and comorbidities as a result of ageing. This point will be considered again in Chapter 5 when the choice of imputation method is decided. Thus, the key term in Equation 3.12 is $P(M_i | y_i^{\text{obs}}, y_i^{\text{miss}}, \phi, x_i)$. This term is what defines the drop-out mechanism and whether the integral in equation 3.11 can be avoided (Carpenter et al 2002).

The main idea behind selection models is that a complete-data model is used for the outcome of interest (Horton & Fitzmaurice 2002). The probability of non-response is then modelled conditionally on the possibly unobserved outcomes. (Horton & Fitzmaurice 2002). Examples in the literature where these likelihood-based models have been applied to real and simulated data include Diggle et al (1994), and Molenberghs et al (1997). Rotnitzky et al (1998) and Scharfstein et al (1999) used
semi-parametric inference methods. Full-likelihood-based extensions of earlier work by Wu & Bailey (1988) can also be found in work conducted by Schluchter (1992) and De Gruttola & Tu (1994). Wu and Bailey (1988) used a method which jointly modelled random effects and drop-out (Horton & Fitzmaurice 2002). One primary attraction of the selection model is that they directly model the marginal distribution of the longitudinal outcomes (Fitzmaurice & Laird 2000). A disadvantage of selection models is that they are computationally intractable and can be difficult to implement using much of the existing statistical software (Fitzmaurice & Laird 2000).

Pattern-mixture models are considered a favourable approach for handling non-ignorable drop-out in longitudinal studies. The primary focus of these models is to make statistical inferences concerning marginal expectation of longitudinal outcomes. This is attained by averaging over the distribution of the drop-out times (Fitzmaurice & Laird 2000). The development of pattern-mixture models includes works by Wu and Bailey (1988), Follmann & Wu (1995) and Hogan & Laird et al (1997) which extended the generalised linear model to handle non-ignorable drop-outs when longitudinal outcomes are continuous and assumed to be multivariate normal (Horton & Fitzmaurice 2002). Fitzmaurice & Laird (2000) extended this work to include categorical, continuous and count data by using generalised estimating equations (GEEs) (Horton & Fitzmaurice 2002). One proposed an approach using maximum-likelihood-based mixture models for incomplete bi-variate outcomes when the non-response mechanism for some missing observations may be non-ignorable.

Pauler et al (2003) discussed the use of pattern-mixture models to differentiate the pattern of missingness which includes deaths. They examined exploratory data techniques for longitudinal HR-QOL data with non-random missingness due to drop-out and censorship by death. They proposed “a pattern mixture model for longitudinal
quality of life, time of drop-out and survival, which allows for straight forward implementation of sensitivity analyses and explicit comparisons to the raw data". Kurland and Heagerty (2005) developed a taxonomy for the accommodation of both drop-out and death and describe estimation for binary longitudinal data that applies selection weights to estimating equations with independence working correlation. They found that “simulation studies and an analysis of monthly disability status illustrate potential bias in regression methods that do not explicitly condition on survival” (Kurland & Heagerty 2005).

There have been a number of publications in recent years that focus on analysing longitudinal quality of life measures with informative drop-out. These studies mostly consider the use of a pattern mixture approach. Recently, Post et al (2010) discussed the use of a pattern mixture approach to analyse data from a trial comparing the chemotherapy dosage for patients with breast cancer. They defined the patterns based on events related to HRQOL, such as death and relapse. They used a random effects model with additional explanatory variables indicating different follow-up (drop-out) patterns of patients. The model was fitted using maximum likelihood. Post et al (2010) used the approach because they felt it was a “relatively simple extension of the random effects models assuming MAR”. The pattern mixture approach is considered an acceptable approach to dealing with informative drop-out whether using maximum likelihood or Fully Bayesian approaches (Carpenter et al 2002).

In recent times there have been a number of advancements for analysing longitudinal data where there is non-monotone MNAR missingness. For binary outcomes, Parzen et al (2006) developed a pseudo-likelihood method for missing outcomes and missing categorical time-dependent covariates. According to Shardell et al (2008) recent work on monotone missing data patterns has been explicitly differentiated between data

Shen et al (2007) have developed an approach which allows both non-ignorable non-monotone MNAR missingness and death (Shardell et al 2008). Their method developed a mixed effects model for structurally missing outcomes in a two phase design. However, the methods mentioned by Shardell et al (2008) do not allow for the missingness patterns of the covariate and outcome to differ, and “only one method formally considered the missing-data mechanism of the covariate when the outcome data are incomplete” (Shardell et al 2008).

Shardell et al (2008) proposed methods using WGEEs for longitudinal studies with death and non-monotone missing time-dependent covariates and outcomes. They suggest that the most challenging aspects of their method are “estimating standard errors from the bootstrap and the limited utility of off-the-shelf statistical packages in efficiently incorporating observations with missing covariates”. In future research Shardell et al (2008) plan to explore alternative methods, such as: using Bayesian analysis and likelihood-based analysis with pattern-mixture and selection models. Beunckens et al (2008) focused on binary repeated measures and suggested “that multiple imputation (MI) can be used to pre-process incomplete data, after which GEEs could then be applied (MI-GEEs)”. They compared this approach to using WGEEs and suggested that MI-GEEs maybe less biased when MAR is assumed.
Harel et al (2007) suggest that when complete information is available on time to death of participants then a single-stage MI procedure or full information maximum likelihood (FIML) may be appropriate. The full-information maximum likelihood (FIML) procedure can be adapted to perform the confirmatory factor analysis (CFA) modelling (Wang & Russel 2005). The FIML procedure contains an EM algorithm that handles multiple imputations of missing values (Schafer 1997; Wang & Russel 2005). However, it too has the limitations of cross-sectional multiple imputation and the assumption of missing at random (MAR). Harel et al (2007) suggest a two-stage MI procedure “which includes time-to-death as a predictor in the models with incomplete follow-up data”. This two-stage MI procedure can be used in Structural Equation Modelling analysis approaches.

Markov transition models are considered as a quasi-likelihood (QL) approach to regression analysis with time series data (Zeger & Qaqish 1988). A class of Markov models, referred to by Cox (1981) as "observation-driven" models in which the conditional means and variances given the past are explicit functions of past outcomes (Zeger & Qaqish 1988). The class includes autoregressive and Markov chain models for continuous and categorical observations as well as models for counts and continuous outcomes with constant coefficient of variation (Zeger & Qaqish 1988). Apart from times series analysis, Transition Markov models have been used to analyse longitudinal data in clinical trials (Kaciroti et al 2008; Yang et al 2007).

An extension of the Markov transition model was proposed by Albert and Follmann (2003), to handle non-ignorable missing values in a binary longitudinal data set (Li et al 2007). When the missing data mechanism can be assumed to be ignorable, standard Markov transition models can be applied to observed data to draw likelihood-based inference on transition probabilities between outcome events (Yang et al 2007). When longitudinal data have non-ignorable missingness mechanisms, random-effects Markov
transition models can be used to model the joint distribution of the binary data matrix and the matrix of missingness indicators (Yang et al 2007). Categorizing missingness patterns into those for occasional or “intermittent” missingness and those for monotonic missingness or “missingness due to drop-out”, is known as the random-effects Markov transition model. Markov transition models provide a novel re-conceptualization of treatment outcomes, offering both intuitive statistical values and relevant clinical insights (Yang et al 2007). The Markov transition models have also been used by Li et al (2007) and Kaciroti et al (2008) for count (Poisson) outcomes. Kaciroti et al (2008) evaluated “the intervention effect over time on y, the number of hospitalizations”, so they propose a “transition Markov model of first order with random intercept, similar to the model that was proposed by Zeger and Qaqish (1988)”. In the clinical trials setting this approach would be very useful. These types of models will not be discussed any further since the focus is on longitudinal cohort studies and not the evaluation of an intervention effect over time.

The above methods have been presented to demonstrate the ongoing efforts to deal with informative missing data. The majority of approaches have been developed due to issues surrounding drop-outs in clinical trials. Further work by Yang and Shoptaw (2008) has lead to the development a set of imputation-based strategies, for handling intermittent and drop-out missing data in clinical trials that are potentially non-ignorable. Yang and Shoptaw (2008) noted that “selection, pattern-mixture, and shared-parameter models are generalized versions of standard longitudinal models (i.e. marginal models using GEEs, linear mixed-effects models, and transition models).

Finally, although the pattern-mixture model appears to be a more favourable approach to account for non-ignorable missing data, the approach will not be considered in this body of work. In Chapter 5, longitudinal methods will use a Fully Bayesian approach
which involves a selection model, even though Fitzmaurice and Laird (2000) suggest otherwise. Encouraged by the Fully Bayesian method by Carrigan et al (2007) and the use of a selection model by Carpenter and Kenward (2005) both approaches will be adapted and applied in Chapter 5.
3.2.2.3 Software with Multiple Imputation procedures

Multiple Imputation procedures have become more common in recent times in the mainstream statistical software packages such as: **SAS 9.1** (SAS Institute Inc. 2003), **Stata 11.1** (Stata Corp. 2009), **S-PLUS** (Insightful Corp. 2003) and **R** (R development Core team 2007) (Carrigan et al 2007). These MI procedures rely on the assumption that data are multivariate normal or can be approximated by a multivariate normal distribution (Carrigan et al 2007; Schafer 1997).

Developments with chained regression equations has led to a number of improvements and additional commands in programs such as MICE in SPLUS (van Buuren & Oudshoorn 1999), Ice in Stata (Royston 2005), and IVEware for SAS (Raghunathan et al 2002). The chained regression equations also allow for categorical variables to be imputed much more simply than previously (Carrigan et al 2007).

Most statistical programs using MI methods have limitations, especially in regards to “incorporating longitudinal information into the imputation methodology” (Carrigan et al 2007). Software packages such as WinBUGS (Spiegelhalter et al 2003) and MLwiN (Rasbash et al 2005) both use a Fully Bayesian framework. These Fully Bayesian techniques “are most suited to longitudinal imputation, as they can incorporate hierarchical structure into the modelling process” (Carrigan et al 2007). They can also deal with much more efficiently with categorical data. Congdon (2001); Cowles (2004); Woodworth (2004) and Carpenter & Kenward (2005) provide useful overviews to WinBUGS and introductory examples of Fully Bayesian imputation with missing data (Carrigan et al 2007).
3.3 DISCUSSION

3.3.1 Treatment of Deaths in SF-36 by Kazis et al 2007

A major focus of this thesis is the SF-36 health survey. The approaches suggested by Kazis et al (2007) are the recommended approaches to account for deaths when using the SF-36. Some of the approaches will be discussed in greater detail and adapted in later chapters. Kazis et al (2007) suggest that, “in older populations the distribution of functionality and health status is intimately tied to mortality and any approach leaning towards relegating or outrightly excluding deaths is likely to bias the results, by giving weight to survivors”.

The Kazis et al (2007) white paper made the following recommendations in the next paragraph for the Medicare Health Outcomes Survey. Although their recommendations were quite specifically directed at the Health Outcomes Survey they suggest that the methods can have wider applications for those using the SF-36 health survey.

The Kazis et al. (2007) panel recommends that HOS continues to use the “HOS approach” but suggested reconsidering the “definition of “better” or “same” based on the standard error of measurement (SEM) rather than on some clinically meaningful standard” (Kazis et al. 2007). They recommended the continued use of case-mix methodology. This case-mix methodology and the alternative case-mix approach suggested in the white paper will be discussed in greater detail in Chapter 7.

The panel also suggested that the “current reporting method by HOS does not make full use of the richness of the data” and recommended that some additional statistics be calculated, possibly a method similar to the SF-6D. “This would provide a single
summary statistic that has a value for death and is of great interest to economists and others”. These recommendations will be adapted and the methodology will be applied in Chapters 6 and 7.

The Kazis et al. (2007) panel also discussed the approaches to incorporate death into the scoring of PCS and the MCS separately. The HALex approach (Trisolini et al 2005) was considered useful but had not been validated. The approaches by Diehr and colleagues such as, the Probability of being healthy in the future (PHF) might be used. The panel noted that the PHF transformations for PCS and MCS have been published and referred to as PCTD (PCS summary measure, transformed with death set to zero) or MCTD. This transformation approach will be adapted in Chapter 4 and the method will be validated using the ALSWH data. The concept of using the HALex approach will be considered but will need to be adapted to another health utility index. This will be explored and explained in detail in Chapter 6.

3.3.2 Treatment of missing SF-36 items

In Section 3.1.4.6 the imputation approach by Rogers et al (2004) known as Modified Regression Estimates (MRE) was discussed. The Kazis et al. (2007) panel recommended the continued use of MRE and suggested that it is a “superior method to the half rule approach and marginally better than the MDE method” (Kazis et al 2007). The ALSWH data has used the SF-36v1 health survey and for this thesis, the focus will consider MI approaches to impute the missing PCS rather than the MRE approach.

3.3.3 Use of simulation models - Suggestion from Kazis et al (2007)

The white paper panel also recommended the use of simulation models for the treatment of deaths. Revicki et al (2001) suggest using simulation modelling to impute projected outcomes for persons as if they had lived, not died. The use of imputation
techniques were suggested to simulate changes in health status as it relates to mortality and the Bayesian approach was termed the most favourable and robust application (Kazis et al 2007). This thesis will consider using Fully Bayesian and MI approaches to the imputation of missing data (see Chapters 5, 6, and 7).

3.3.4 Methods for dealing with informative (non-ignorable) missing data

According to Kazis et al (2007), “the simulation of deaths with corresponding functional status has sound theoretical basis such that a calculated study can reliably inform the extent to which the mortality is affecting HR-QOL indicators”. A downside of the use of simulations in older populations, “where death can be intimately tied with health status”, is that “pattern mixture models accounting for missing data can show weak validations” (Pauler et al 2003). In addition, Kazis et al (2007) suggests “it is well understood that pattern mixture models cannot account for death without implicitly imputing values for the persons who have died”. In the following chapters, the recommendations put forward by the Kazis et al (2007) white paper will be considered. The suggestions for the treatment of deaths and missing data will be adapted and applied to the ALSWH data.

Finally, many of the approaches mentioned for dealing with missing data have been for clinical trials. The main focus in Chapter 5 will be to determine an approach to account for missing outcome data (in a cohort study) once deaths have been reinstated with some value. To model the probability of drop-out, there has been a trend towards pattern mixture models (Curran et al 2002; Michiels et al 2002; Post et al 2010). In Chapter 5, the use of longitudinal multiple imputation will focus on intermittent missing outcomes and other drop-outs and will consider the selection model approach to model the assumptions of MAR and MNAR. In Chapter 5 the approach considers a random effects model with random intercepts and coefficients using a Fully Bayesian approach.
in a longitudinal cohort study setting. This method will be explained in detail in Chapter 5 and this approach will extend examples by Carrigan et al (2007), Carpenter and Kenward (2005), Carpenter (2009) and Carpenter (2009b) for MNAR and MAR to a longitudinal analysis using a selection model in WinBUGS.
CHAPTER 4:

Transforming the SF-36 to account for deaths in longitudinal studies with a three year follow-up period
List of Abbreviations used in Chapter 4

ACQUIP  Ambulatory Care Quality Improvement Project
ALSWH  Australian Longitudinal Study on Women’s Health
APCT  (Australian) Probability of being healthy in three years
APCTD  (Australian) Probability of being healthy in three years (includes deaths)
CMS  Centres for Medicare and Medicard Services
EVGGFP  Excellent, Very Good, Good, Fair, Poor
MCS  Mental Component Summary score (derived from SF-36)
PCS  Physical Component Summary score (derived from SF-36)
PCT  Probability of being healthy in two years
PCTD  Probability of being healthy in two years (includes deaths)
PHF  Probability of being health in the Future
QALY  Quality Adjusted Life Years
SF-6D  Short Form 6 Domains
SF-36  Short Form 36
SF-36v1  Short Form 36 Version 1
SF-36v2  Short Form 36 Version 2 1996
4.0 **INTRODUCTION**

This chapter will discuss the development of an approach to account for deaths in the analysis of data from longitudinal studies of cohorts. The Australian Longitudinal Study on Women’s Health (ALSWH) data set used in this chapter was introduced in Chapter 2. Surveys 1, 2 and 3 will be examined.

4.1 **BACKGROUND**

The analysis of data from longitudinal studies is made more complex by the death of study participants over time. Many statistical methods depend on complete case analysis, meaning that data for participants who die are often removed from the analysis and reported separately. As Diehr et al (2003) suggest, this approach limits longitudinal analysis to survivors who often begin or remain in better health than those who die and hence researchers may miss important changes over time in the total cohort or produce misleading results.

Many longitudinal studies that aim to measure changes in physical and mental health over time use the Medical Outcomes Study Short-Form 36 (SF-36). The SF-36 is a 36 item scale which covers the major aspects of health and has been validated with adults of all ages (Ware & Sherbourne 1992; Ware 1993; Ware et al 1994) and has been used extensively in longitudinal studies, particularly in clinical trials (Brazier et al 1998). As discussed in Chapter 2, a limitation of the SF-36 is that it only considers morbidity and not mortality. As also discussed in Chapter 2, some other health related quality of life scales do consider both morbidity and mortality (Anderson et al 1998). Using the SF-36 Version 2 (1996), Brazier et al (1998) developed the SF-6D to calculate Quality Adjusted Life Years (QALY) for use by health economists. The SF-6D has scores...
ranging from 0-1, with a value of zero for death and one for healthy and has been regarded as very useful for researchers interested in the effectiveness and cost-effectiveness of health care interventions (Brazier et al 1998). However, the SF-6D does not provide as much information on aspects of the health-related quality of life as the SF-36. The focus of this chapter is trying to determine how to account for deaths in longitudinal studies which have used the SF-36 to measure quality of life.

4.1.1 Two Approaches for incorporating deaths into Health status

As previously mentioned in Section 3.2.1.5, a number of strategies have evolved from Diehr and colleagues which directly attempt to incorporate death into the SF-36 (Diehr et al 1995; Diehr et al 2001; Diehr et al 2001b; Diehr et al 2003; Diehr et al 2003b; Diehr et al 2005). Their eight original strategies for dealing with death in longitudinal studies of health-related quality of life are summarised and presented in Table 3.1 in Section 3.2.1.5. Two of these adapted strategies were discussed by Kazis et al (2007) and will be considered and adapted in this chapter.

4.1.2 Approach 1 - “Healthy” Versus “Not healthy”

Diehr et al (2001) suggested two approaches to accounting for deaths in HR-QOL measures. The first approach simply divides health into two states “healthy” and “not healthy”, where those who died are considered not healthy. This approach has been applied to the self-rated health question, which asks participants to rate their health as: Excellent, Very Good, Good, Fair or Poor.

Thus the self-rated health question is dichotomised to:

“Healthy” – Excellent, Very Good and Good.

“Not healthy” – Fair, Poor and Dead.
4.1.3 Approach 2 - Transformation of Self-rated health (EVGGFP) into a Probability of being healthy in the future (PHF)

As previously mentioned in Section 3.2.1.3(i), the second approach again considers the self-rated health (EVGGFP) and transforms the actual observed state (EVGGFP) into a probability of being healthy in the future (PHF). Diehr and colleagues developed a number of strategies for dealing with deaths which were discussed earlier in Section 3.2.1.5. They decided that set values would be used for an estimated probability of being healthy in the future (PHF). Thus the participant who was self-rated as excellent would be coded as a probability score of 95. That is, the probability of being healthy in the future for an individual rated excellent was 95%, whereas, a person who had died would receive a value of zero. Other self-rated health values considered were as follows: Excellent (95), Very good (90), Good (80), Fair (30), Poor (15), Dead (0).

There are advantages and disadvantages of both approaches: healthy vs not healthy and the probability of being healthy in the future. The dichotomising of EVGGFP into healthy vs not healthy is easy to interpret, however there is a loss of detail about the original EVGGFP scale. In terms of the probability of being healthy in the future approach, this is also easily interpretable and provides a scale 0-100. It is conceptually reasonable but it is not a preference/utility score so it may be a little harder to explain than the dichotomous transformation approach (Kazis et al 2007). A combination of both approaches lead to a transformation of the Physical Component Summary score (PCS). The PCS transformation approach will be explored in greater detail in this chapter.
4.1.4 An approach for the Transformation of PCS and MCS

From the earlier work by Diehr and colleagues came the transformations of PCS and MCS. Eventually they found a more appropriate and useful strategy was a modification of strategy 6, whereby the SF-36 Physical Component Summary score (PCS) is transformed to incorporate death into the scale. The transformed PCS results in a score which is the estimated probability of being healthy one year later. It is this approach to dealing with death which is of great interest here in Chapter 4. The method was mentioned earlier in Section 3.2.1 and will later be adapted and applied in Sections 4.3.6 and 4.5.2.

However, before deciding upon the final PCS and MCS transformations a number of approaches were considered. Diehr et al (2001) discussed three PCS transformation approaches which involved: (1) Probability of being Alive; (2) Probability of being Healthy (Excellent/Very Good/Good) and (3) Probability of having a PCS in the top 75% of the reference population values for men aged 65 and older. In this case the mean PCS was 33.48 for this reference population (Diehr et al 2001). The three approaches were also applied to the MCS and the eight SF-36 sub scales. A table from Diehr et al (2001) is presented in Appendix 4.1 (Table A4.1), to provide more information on these transformation equations. The transformation of PCS involving the use of the probability of being “healthy”, will be discussed further in Section 4.3.6.

The methodology developed by Diehr and colleagues was derived from and applied to a number of datasets, in particular to a large study of United States Veterans (Diehr et al 1995; Diehr et al 2001; Diehr et al 2003; Diehr et al 2005). This Ambulatory Care Quality Improvement Project (ACQUIP) study cohort of veterans consisted mostly of men (97%) who were 65 years and older and chronically ill. Diehr et al (2003)
suggested that a limitation of their transformation approach was “that study data were primarily from veterans, who were quite different in health from the general populations. Experience with these transformations in more populations is needed”. Fihn et al (2004) applied this method in their trial that evaluated the effects of a sustained program of audit/feedback on patient health satisfaction. They found no significant differences in the intervention and control groups after adjusting for deaths (Fihn et al 2004). Strandberg et al (2004) applied the method when assessing the effects of midlife alcohol consumption on mortality and quality of life in old age (Strandberg et al 2004). Quality of life was significantly worse in men with high alcohol consumption after accounting for deaths during follow-up using the method developed by Diehr and colleagues (Strandberg et al 2004). It is not clear whether either study used Equation 4.1 (see Section 4.3.6) or generated their own equation from their data.

The Australian Longitudinal Study on Women’s Health provides an opportunity to apply the methodology of Diehr et al to a new cohort and study design. This longitudinal study examines the health of three large cohorts of community-dwelling Australian women with follow up Surveys which are spaced approximately every three years. At the beginning of these longitudinal studies in 1996, the three cohorts were named as the following:

- Young Women aged 18-23 years at baseline (N=14247)
- Middle Aged Women aged 47-52 years at baseline (N=13715)
- Older Women aged 69-74 years at baseline (N= 12432)

The cohort of greatest interest in this thesis is the older women cohort which is often referred to by researchers, as the older cohort. The older cohort is surveyed at three year intervals. The project was designed to explore factors that influence health among women who are broadly representative of the Australian population (Brown et al 1998).
A more comprehensive explanation of the three cohorts can be found at the Study’s website: http://www.alswh.org.au/.

4.2 AIMS AND OBJECTIVES

There are four main aims and objectives for this chapter (i.e. not ALSWH):

1. To create the three-year transformation equation for the SF-36 scores using Survey 1 and Survey 2 ALSWH data, using the method developed by Diehr et al (2003).

2. To validate the transformation using ALSWH data from Survey 2 and Survey 3.

3. To determine the generalisability of the method and in particular whether the method appears sound when a time interval of three years is used.

4. To apply the transformation equation to determine the effect of deaths on the health related quality of life scores for a case study of women with and without diabetes.

4.3 METHOD

4.3.1 Sample

Women in the ALSWH sample were randomly selected from the Australian national health insurance database (Medicare) which includes almost all citizens and permanent residents, regardless of age or income (Brown et al 1998). Further details of the study were presented in Chapter 2.
4.3.2 Data

ALSWH data were collected by mailed self-complete surveys, consisting of a range of self-reported data including health, demographics, social and psychological variables. The SF-36 is typically reported in eight subscales from which a summary score called the Physical Component Score (PCS) can be derived. Calculation of the PCS is based on a formula developed for the US population (Ware et al 1994). The ALSWH PCS was calculated using a formula standardised to the relevant Australian population (Mishra & Schofield 1998). See Section 2.9.1 and Appendix 2.1 for more details about ALSWH PCS.

Deaths of study participants are recorded by linkage to the National Death Index where name, address and date of birth for each woman are matched annually (Powers et al 2000). The study office also logged all correspondence with participants and others notifying the withdrawal from the study or the death of a study participant. These additional death notifications were then compared with the National Death Index to ensure completeness of follow-up for mortality (Powers et al 2000).

4.3.3 ALSWH - Response, Retention Rates and Deaths

The focus of this chapter is the older cohort of women for whom Survey 1 was conducted in 1996 (aged 70-75 years, n=12432), Survey 2 in 1999 (aged 73-78 years) and Survey 3 in 2002 (aged 76-81 years). Between Survey 1 and Survey 2, 518 women died, 102 were withdrawn due to ‘frailty’, and 1111 did not complete Survey 2 but were not known to be deceased. Between Survey 2 and Survey 3 there were a further 564 deaths and 262 withdrawn due to frailty, and 1548 women did not complete Survey 3 but were not known to be deceased.
Table 4.1  Response, Retention Rates and Deaths of Older Cohort

<table>
<thead>
<tr>
<th></th>
<th>Survey 1</th>
<th>Survey 2</th>
<th>Survey 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age in Years</td>
<td>70-75</td>
<td>73-78</td>
<td>76-81</td>
</tr>
<tr>
<td>Eligible at current Survey</td>
<td>12432</td>
<td>11535</td>
<td></td>
</tr>
<tr>
<td>Ineligible</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>deceased between Surveys</td>
<td>529</td>
<td>569</td>
<td></td>
</tr>
<tr>
<td>frailty (e.g. dementia, stroke)</td>
<td>106</td>
<td>264</td>
<td></td>
</tr>
<tr>
<td>withdrawn before mail-out Survey date</td>
<td>262</td>
<td>515</td>
<td></td>
</tr>
<tr>
<td>Total ineligible</td>
<td>887</td>
<td>1348</td>
<td></td>
</tr>
<tr>
<td>Eligible at current Survey</td>
<td>11535</td>
<td>10187</td>
<td></td>
</tr>
<tr>
<td>Non-respondents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>withdrawn from the project</td>
<td>311</td>
<td>385</td>
<td></td>
</tr>
<tr>
<td>contacted but did not return</td>
<td>481</td>
<td>860</td>
<td></td>
</tr>
<tr>
<td>unable to contact participant</td>
<td>309</td>
<td>295</td>
<td></td>
</tr>
<tr>
<td>Total non-respondents</td>
<td>1101</td>
<td>1540</td>
<td></td>
</tr>
<tr>
<td>Respondents</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>completed Survey</td>
<td>12432</td>
<td>10434</td>
<td>8647</td>
</tr>
<tr>
<td>Retention rate as % eligible</td>
<td>*91%</td>
<td>*85%</td>
<td></td>
</tr>
</tbody>
</table>

*Table 4.1 results have been updated since provided by Lee et al (2005)*

Table 4.1 shows that 91% of women responded to Survey 2 in 1999 and 85% to Survey 3 in 2002. The main reason for non-response was the non-return of the questionnaire, this equated to 4% of eligible women at Survey 2 and 8% at Survey 3. It appears that these and other non-respondent women were more likely to report fair to poorer self-rated health at Survey 1 than the respondents (39% vs 25%, P-value <0.0001). There was also a statistically significant difference in the proportion of responders and non-responders at Survey 3 who reported fair/poor health at Survey 2 (35% vs 24%, P-value < 0.0001). Thus, it appears that the non-responders were in poorer health than respondents.

Figure 4.1 suggests that the group of women at Survey 3 who self-reported poor health had statistically significantly lower mean PCS (P-value <0.001) than other women in the study at Survey 1 and Survey 2. In fact, this group had a statistically significant lower mean PCS than the group of women who died between Survey 2 and Survey 3.
Figure 4.1 shows the differences in mean PCS of women who reported Excellent or Very Good or Good at Survey 3 are within 5 points (on PCS scale), which is considered clinically important (Kazis et al 2007). Likewise those women who were reported as Fair or Poor or Frail or Dead are also within 5 points of each other. This supports the approach by Diehr et al (2003), where they combined Excellent / Very Good / Good (1) and Fair / Poor / Dead (0).

According to the ALSWH Technical Report (2007) the demographic characteristics of the older cohort respondents from Survey 1 and Survey 3 were compared with those of women of the same age in the Australian population, using data from the 1996 and 2001 Censuses respectively. The demographic characteristics that were considered include: country of birth, marital status, education and lone person household. There was some under-representation of women from non-English speaking countries in the ALSWH sample at both Surveys (ALSWH 2007). The high level of missing data in the Census made comparisons difficult for marital status and educational qualifications but otherwise the comparisons showed few differences.
Figure 4.1  Mean PCS scores at Surveys 1 and 2 by General Health Classification at Survey 3
4.3.4 Measures

The following variables were generated in order to conduct the analysis. Firstly, a binary variable was generated to classify participants as healthy or not healthy. This categorical variable was then used to transform the PCS score to create a new variable called the “probability of being healthy in 3 years” (APCTD).

4.3.5 Definition of “Healthy”

The self-rated health question “In general, would you say your health is: Excellent, Very Good, Good, Fair or Poor?” (EVGGFP) is the first item in the SF-36. Responses to this question were used to define being ‘healthy’ at a particular time. People who answered Excellent, Very Good or Good were considered to be ‘healthy’ (score =1) while those in fair or poor health, those who had died, or had withdrawn due to frailty were considered to be ‘not healthy’ (score=0).

4.3.6 Derivation of the value of APCTD

Logistic regression was used to derive the equation to estimate the probability of being ‘healthy’ at the next survey as a function of the current PCS score. The transformation reported by Diehr et al (2003), where the follow-up measure of health was made one year later, was:

\[ \text{PCT} = \frac{e^{(-3.77+0.1089\times\text{PCS})}}{1+e^{(-3.77+0.1089\times\text{PCS})}} \times 100 \]  \hspace{1cm} (equation 4.1)

Initially, transformed scores were calculated for the ASLWH data using the Equation 4.1. The new variable was generated by using the PCS at Survey 1. The PCT variable was called PCT\(_1\) and represented the “probability of being healthy at Survey 2”. Next, another variable was created using the Equation 4.1 to represent the “probability of
being healthy at Survey 3". On this occasion the PCS at Survey 2 was used to create PCT₂.

In order to determine the impact of death, a further variable is generated called PCTD. This new variable included those participants who were known to have died between surveys, where the value of zero is given for the missing PCT. Since there are no deaths at Survey 1 then PCT₁=PCTD₁. For those who died between Surveys 1 and 2 a variable called PCTD₂ was created and a zero was included for those who were known to have died. Likewise the variable PCTD₃ was created to account for those women who died between Surveys 2 and 3.

Secondly, as the time interval between ALSWH Surveys was three years and the ALSWH uses an Australian version of the PCS, a new equation was derived. The ALSWH Survey 1 PCS was used to predict the percent ‘healthy’ at Survey 2 in a logistic regression model creating study-specific coefficients for transforming the PCS. The transformed variable was labelled APCT₁. For women who died, the APCT score is set to zero at all time points after their death, as there is no chance they will be healthy at any future follow-up. The set of transformed scores that include values for death are labelled APCTD. As mentioned earlier, there were no deaths at Survey 1, so APCTD₁ = APCT₁. When PCS are missing for reasons other than death, the PCS may be imputed using a range of methods (Engels & Diehr 2003; Twisk 2003; Diehr et al 2005). Imputation of missing data has been discussed in greater depth in Chapter 3 and later certain approaches to account for missing data will be applied in Chapter 5. The focus of this chapter has been the development and validation of the three-year transformation (APCTD).
4.3.7 Validation of APCTD

The following outlines the process of validation of the APCTD using ALSWH data. The ALSWH transformation was applied to PCS scores at Survey 2 to predict the probability of being healthy three years later. Values for the observed percent ‘healthy’ three years later (at Survey 3) were plotted against the predicted probability of being healthy, to assess the fit of the new equation to a dataset other than the one in which it was derived. In addition, values for APCT₁ were plotted against the observed percent ‘healthy’ at Survey 2 for each decile of women, defined by their APCT scores, to assess the fit of the transformation. Values for APCT₂ (predicted probability of being healthy three years after Survey 2) were plotted against the observed percent ‘healthy’ three years after Survey 2.

To compare the percent of predicted and observed ‘healthiness’ after 3 years, only those participants who had a PCS and were alive at Survey 2 were considered. To make comparisons, this group of women also needed to have answered the question: In general, would you say your health is...[excellent, very good, good, fair, poor] " in Survey 3. Within the group of women who were considered not healthy (0) there were those who had died since Survey 2 and others who had withdrawn due to illness (n=262, see Table 4.1).

4.3.8 Definition Diabetes Status at Survey 1 and Survey 2

To demonstrate the effect of deaths on the health related quality life scores for a case study of women with and without diabetes, the following definitions were established. In Survey 1 the following question was asked of participants: “Have you ever been told by your doctor that you have: …" The participant had the option of 16 health conditions to respond to, one of which was diabetes.
Existing Cases

An Existing case of diabetes was determined if the participant had answered

“Yes” at Survey 1: “Have you ever been told by a doctor that you have:… [diabetes]”

No Diabetes

A participant was considered not to have diabetes if she answered: “No" in Survey 1 and “No" in Survey 2 or “Missing” in Survey 1 and “No" in Survey 2.

Even though some participants had a “missing value ” at Survey 1 and had answered “no” to Survey 2, these women were still considered in the group of “no diabetes”.

Missing Classification

A participant was considered as “Missing” and hence excluded from analysis if she answered: “Missing” in Survey 1 and “Missing” in Survey 2; or “No” in Survey 1 and “Missing” in Survey 2.

The reason for excluding those who answered “No" then were “Missing” at Survey 2 is that the status was unable to be determined and at this stage it was decided to exclude them from the analysis. Note in this Chapter, for purpose of applying the transformations to the ALSWH data, only Existing Cases and No Diabetes will be considered in the analyses. In later chapters the definition of New Cases will be considered and discussed.
4.4 **Statistical Analysis**

The development and validation of the ALSWH equation involves the use of logistic regression and diagnostic tests such as the Hosmer-Lemeshow goodness-of-fit. Simple linear regression is used for the continuous variables by determining the relationship between the observed percent ‘healthy’ and predicted percent ‘healthy’. Graphs and lines of best fit are used to show these relationships. To determine whether the inclusion of values for death had any substantive effect on findings, the change in several measures of health between Survey 1 and Survey 2 were compared for women with and without self-reported diabetes. Independent t-tests were used to compare mean change in PCS, APCT and APCTD for the two groups. Paired t-tests were used to compare mean changes (Survey1–Survey2) in PCS, APCT and APCTD within the two groups. Statistical analyses were conducted using SAS 9.1 (2003) and Stata 9.1 (2006). A significance level of 5% was considered appropriate.

4.5 **Results**

4.5.1 **Application of Equation 4.1 using ALSWH data**

Equation 4.1 was used to generate the variables for PCT at each of Surveys 1, 2 and 3. The variables for PCTD were also generated for Surveys 1, 2 and 3. The descriptive statistics of these variables can be found in Table 4.2, while the descriptions of the shapes of the distributions are shown in Appendix 4.2.

4.5.2 **Development of the Transformation Equation APCT**

The results from the logistic regression which used the binary outcome variable (healthy=1, not healthy=0) and the predictor variable as PCS at Survey 1 were as follows:
The logit equation is represented by the log odds of being healthy:

\[
\text{logit} \left[ \pi(x_i) \right] = \alpha + \beta x_i = -4.98 + 0.12 \times \text{PCS}_i \quad \text{(equation 4.2)}
\]

where \( \pi \) = probability of being healthy in 3 years.

This can then be transformed into a logistic probability function and thus represented as a percentage (\times 100).

The transformation derived from the Survey 1 and Survey 2 ALSWH data was:

\[
\text{APCT} = \frac{e^{(-4.98+0.12\times\text{PCS})}}{1+e^{(-4.98+0.12\times\text{PCS})}} \times 100 \quad \text{(equation 4.3)}
\]

For example, using this transformation, a person reporting a PCS of 80 would be predicted to have a 99% probability of being healthy 3 years later, and the predicted probability of being healthy in 3 years is still 90% when the PCS is 60. For a person who has a low PCS (20), the predicted probability of being healthy in 3 years is 7%.

To determine whether the logit equation was a reasonable fit for the data the appropriate diagnostics and goodness of fit test were conducted. According to the Hosmer-Lemeshow goodness-of-fit statistic \( C = 9.08 \) which is compared with a chi-square distribution with 8 degrees of freedom. The corresponding P-value is 0.336 which indicates that the simple model is a reasonable fit. At a cut-off of 0.75, 70% of observations are correctly classified, where the sensitivity is 67%, and the specificity is 77%. Overall, based on these statistics the model will be considered a reasonable fit.

### 4.5.3 Comparisons of Distributions for Diehr and ALSWH equations

Transformed scores from both the Equation 4.1 and ALSWH Equation 4.3 were compared for all women. As expected, the Equation 4.1, which estimates the
probability of being healthy one year in the future, resulted in higher scores (PCT, mean 79.58, standard deviation 17.13) than the ALSWH equation which estimates the probability of being healthy in three years (APCT, mean 68.65, standard deviation 22.55) (see Table 4.2). These two means were statistically significantly different (P-value <0.0001, see Table 4.2). The shapes of the distributions of the transformed variables using both Diehr and ALSWH equations were very similar and both were highly skewed to the left (see Appendix 4.2). Although these data were skewed paired t-tests were used to compare PCTD and APCTD at all 3 surveys, the means are all statistically significantly different (P-value< 0.001).

Table 4.2  Descriptive Statistics – Comparisons of Equations at Surveys 1-3

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Survey 1</th>
<th></th>
<th>Survey 2</th>
<th></th>
<th>Survey 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (SD)</td>
<td>Minimum</td>
<td>Maximum</td>
<td>Mean (SD)</td>
<td>Minimum</td>
</tr>
<tr>
<td>PCS</td>
<td></td>
<td>49.96 (9.95)</td>
<td>17.86</td>
<td>75.45</td>
<td>49.58 (10.03)</td>
<td>13.45</td>
</tr>
<tr>
<td>Diehr equation</td>
<td></td>
<td>PCT 10851</td>
<td>79.58 (17.13)</td>
<td>13.89</td>
<td>98.84</td>
<td>78.94 (17.46)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCTD 10851</td>
<td>79.58 (17.13)</td>
<td>13.89</td>
<td>98.84</td>
<td>74.98 (24.22)</td>
</tr>
<tr>
<td>ALSWH equation</td>
<td></td>
<td>APCT 10851</td>
<td>68.65 (22.55)</td>
<td>5.47</td>
<td>98.25</td>
<td>67.80 (22.85)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>APCTD 10851</td>
<td>68.65 (22.55)</td>
<td>5.47</td>
<td>98.25</td>
<td>64.40 (26.74)</td>
</tr>
<tr>
<td>PCS</td>
<td></td>
<td>47.82 (10.18)</td>
<td>17.97</td>
<td>75.75</td>
<td>47.82 (10.18)</td>
<td>17.97</td>
</tr>
<tr>
<td>Diehr equation</td>
<td></td>
<td>PCT 7135</td>
<td>76.07 (18.48)</td>
<td>14.03</td>
<td>98.88</td>
<td>70.46 (26.68)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PCTD 7703</td>
<td>76.07 (18.48)</td>
<td>14.03</td>
<td>98.88</td>
<td>70.46 (26.68)</td>
</tr>
<tr>
<td>ALSWH equation</td>
<td></td>
<td>APCT 7135</td>
<td>63.92 (23.61)</td>
<td>5.54</td>
<td>98.31</td>
<td>63.92 (23.61)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>APCTD 7703</td>
<td>63.92 (23.61)</td>
<td>5.54</td>
<td>98.31</td>
<td>59.21 (28.20)</td>
</tr>
</tbody>
</table>
4.5.4 Example of Comparing Diehr and ALSWH Equations

To highlight the effects of the differences in the two equations the following graph has been developed. Figure 4.2 shows the PCT and APCT for various values of PCS. A person with a high PCS of 90 has a high probably of being “healthy” in one year (Diehr) and likewise similar result for the ALSWH equation (three years). A person in relatively poor physical health with a PCS of 10 has a low probability of being “healthy” in one year (Diehr) and in three years (ALSWH). However, the predictions for women whose physical component summary scores (PCS) are between 30-50, are more divergent.

![Graph](image)

Figure 4.2 A Comparison of Equation 4.1 and ALSWH Equation
4.5.5 Development of the ALSWH Equation using Survey 1 and Survey 2

To determine how well the ALSWH equation fitted the data from Survey 1 and 2, the observed percent ‘healthy’ after three years (at Survey 2) were plotted against the predicted percent ‘healthy’ (Survey 2). In order to create this plot, deciles were obtained by dividing APCT\textsubscript{2} into 10 equal groups. Each decile consisted of 968 women. The predicted percentage healthy column is the mean APCT\textsubscript{1} for each decile. The observed percentage healthy column of the table was created by determining the percentage of healthy subjects at Survey 2 in each of the APCT\textsubscript{1} deciles using the probability of being healthy at Survey 2 variable. The values of the points used in Figure 4.3 are shown in Table 4.3.

Table 4.3 Predicted percent Healthy vs observed percent Healthy at Survey 2

<table>
<thead>
<tr>
<th>APCT\textsubscript{1} Deciles N=9676</th>
<th>% Predicted Healthy at S2 (n=968)</th>
<th>% Observed Healthy at S2 (n=968)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 0 ≤ 33</td>
<td>23.4</td>
<td>23.9</td>
</tr>
<tr>
<td>&gt; 33 ≤ 47</td>
<td>40.4</td>
<td>40.4</td>
</tr>
<tr>
<td>&gt; 47 ≤ 60</td>
<td>53.5</td>
<td>52.5</td>
</tr>
<tr>
<td>&gt; 60 ≤ 69</td>
<td>64.9</td>
<td>65.5</td>
</tr>
<tr>
<td>&gt; 69 ≤ 77</td>
<td>73.5</td>
<td>71.2</td>
</tr>
<tr>
<td>&gt;77 ≤ 82</td>
<td>79.8</td>
<td>82.0</td>
</tr>
<tr>
<td>&gt; 82 ≤ 86</td>
<td>84.1</td>
<td>82.8</td>
</tr>
<tr>
<td>&gt; 86 ≤ 89</td>
<td>87.6</td>
<td>88.9</td>
</tr>
<tr>
<td>&gt; 89 ≤ 92</td>
<td>90.4</td>
<td>90.5</td>
</tr>
<tr>
<td>&gt; 92 ≤ 99</td>
<td>93.3</td>
<td>93.3</td>
</tr>
</tbody>
</table>
The darker line is a least-squares fit to those points in Table 4.2

The dotted line is the 45° line of perfect calibration

Figure 4.3  Predicted percent Healthy versus observed percent Healthy at Survey 2

Figure 4.3 shows the average predicted probability of being healthy at Survey 2 using the ALSWH equation (X axis) and the observed percent healthy (Y axis), using the 10 equal-sized groups of women defined by their APCT₁. Figure 4.3 suggests that those in the top decile (APCT scores of 92–99) have on average a 93% chance of being healthy in 3 years. Of that top decile, 93% were observed to be ‘healthy’. Women in the lowest decile (APCT scores ≤ 33) had on average a 23% predicted probability of being ‘healthy’ at Survey 2 and 24% were observed to be ‘healthy’. Figure 4.3 suggests a strong linear relationship for the predicted percent Healthy at Survey 2 versus the observed percent Healthy at Survey 2. A simple linear regression of these 10 data points provided the slope (β) coefficient as 0.994; 95%CI (0.95, 1.04) and an R-squared=99.7%.
4.5.6 Validation of ALSWH equation using Survey 2 and Survey 3

The ALSWH equation (Equation 4.3) has been developed using data from Survey 1 and Survey 2, but how appropriate is Equation 4.3 for further ALSWH Surveys? The decision to generate the ALSWH equation has been based upon evidence that Equation 4.1 is based on a very different population and only has a 12 months follow-up. To determine whether a new ALSWH Equation should be generated between ongoing surveys (that is from Survey 2 to 3 and Survey 3 to 4) then Equation 4.3 needed to be validated using data from Survey 2 and Survey 3.

To validate the ALSWH equation (4.3) a plot of the observed percent ‘healthy’ after three years (at Survey 3) versus the predicted percent ‘healthy’ (Survey 3) was created. Thus, in order to create Figure 4.4 deciles were obtained by dividing APCT$^2$ scores into 10 equal-sized groups of women. Each decile consisted of 832 women. The values are shown in Table 4.4. The predicted percentage healthy column is the mean APCT$^2$ for each decile. The observed percentage healthy column of the table was created by determining the percentage of healthy subjects at Survey 3 in each of the APCT$^2$ deciles using the “probability of being healthy at Survey 3” variable. The values shown in Table 4.4 have been represented in Figure 4.4 with a line of best fit drawn.

Figure 4.4 shows the average predicted probability of being healthy at Survey 3 using the ALSWH equation (X axis) and the observed percent healthy (Y axis), using 10 equal-sized groups of women defined by their APCT (n=8315). Figure 4.4 suggests that those in the top decile (APCT scores of 91–99) have on average a 93% chance of being healthy in 3 years. Of that top decile, 88% were observed to be ‘healthy’. Women in the lowest decile (APCT scores ≤ 31) had on average a 22% predicted probability of being ‘healthy’ at Survey 3 and 26% were observed to be ‘healthy’. Figure 4.4 suggests...
that the fit to the 45 degree line is reasonable. Using a similar validation approach to that of Diehr et al (2003), a simple linear regression of these 10 data points found that the slope (β) coefficient was 0.89; 95% CI (0.84, 0.93) and an R-squared=99.6%. Figure 4.4 shows that for women with lower predicted probability of being healthy, a higher percent were observed as being healthy three years later than predicted by the model. This may be partly explained by the finding that women who were not known to be dead at Survey 3 but did not complete the survey (and hence did not contribute data to the observed percent healthy) tended to be in poorer health at Survey 2 than women who did complete Survey 3.

Table 4.4  Observed percent ‘healthy’ after 3 years (at Survey 3) vs Predicted percent ‘health

<table>
<thead>
<tr>
<th>APCT2 Deciles N=8315</th>
<th>% Predicted Healthy at S3 (n=832)</th>
<th>% Observed Healthy at S3 (n=832)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 0 ≤ 31</td>
<td>22.2</td>
<td>26.5</td>
</tr>
<tr>
<td>&gt; 31 ≤ 45</td>
<td>38.2</td>
<td>42.6</td>
</tr>
<tr>
<td>&gt; 45 ≤ 58</td>
<td>51.8</td>
<td>56.2</td>
</tr>
<tr>
<td>&gt; 58 ≤ 67</td>
<td>62.7</td>
<td>65.0</td>
</tr>
<tr>
<td>&gt; 67 ≤ 75</td>
<td>71.3</td>
<td>72.0</td>
</tr>
<tr>
<td>&gt; 75 ≤ 81</td>
<td>78.2</td>
<td>77.9</td>
</tr>
<tr>
<td>&gt; 81 ≤ 85</td>
<td>83.3</td>
<td>82.6</td>
</tr>
<tr>
<td>&gt; 85 ≤ 89</td>
<td>87.2</td>
<td>86.5</td>
</tr>
<tr>
<td>&gt; 89 ≤ 91</td>
<td>90.1</td>
<td>89.4</td>
</tr>
<tr>
<td>&gt; 91 ≤ 99</td>
<td>93.0</td>
<td>88.3</td>
</tr>
</tbody>
</table>
The darker line is a least-squares fit to those points in Table 4.3
--- The dotted line is the 45° line of perfect calibration

Figure 4.4 Observed percent ‘healthy’ (at Survey 3) vs predicted percent ‘healthy’

4.5.7 Comparing the Changes in health of Women with and without Diabetes

The change in health of women with diabetes \((n=937)\) from Survey 1 to Survey 2 was not significantly different from the change in health of women without diabetes \((n=9811)\) when compared using the PCS \((P\text{-value}=0.89)\) and the transformed score APCT \((P\text{-value}=0.90)\) (see Table 4.5). Between Survey 1 and Survey 2, 8% of women with diabetes died \((n=88)\) and 4% of women without diabetes died \((n=405)\). When incorporating values for these deaths, APCTD showed a statistically significantly greater decline in health for women with diabetes (Table 4.5) \((P\text{-value}=0.02)\). The larger proportion of deaths in the diabetes group has impacted significantly on the
mean APCTD. That is, the introduction of a zero score for death has caused a greater reduction in the mean APCTD for this group.

### Table 4.5 Changes in health between Surveys 1 and 2 for Diabetes Groups

<table>
<thead>
<tr>
<th></th>
<th>SF-36 physical component summary score PCS</th>
<th>Transformed PCS APCT</th>
<th>Incorporating deaths APCTD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Diabetes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey 1 (n=937)</td>
<td>Mean (SD) 45.2 (9.7)</td>
<td>Mean (SD) 57.7 (23.2)</td>
<td>Mean (SD) 57.7 (23.2)</td>
</tr>
<tr>
<td>Survey 2</td>
<td>Mean (SD) 45.3 (10.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>No Diabetes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Survey 1 (n=9811)</td>
<td>Mean (SD) 50.5 (9.9)</td>
<td>Mean (SD) 69.8 (22.3)</td>
<td>Mean (SD) 69.8 (22.3)</td>
</tr>
<tr>
<td>Survey 2</td>
<td>Mean (SD) 49.9 (9.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diff in mean change</strong></td>
<td>-0.05 (0.3)</td>
<td>0.1 (0.7)</td>
<td>1.9 (0.8)</td>
</tr>
<tr>
<td><strong>P-value</strong></td>
<td>0.89</td>
<td>0.90</td>
<td>0.02</td>
</tr>
</tbody>
</table>

*Diff in mean change = (Diabetes Survey 1 - Diabetes Survey 2) - (No Diabetes Survey 1 - No Diabetes Survey 2). Note: N's vary at Diabetes - Survey 2 and No Diabetes - Survey 2 for all measures due to missing observations. N's shown are those women who were reported with a PCS at Survey 1.

### Table 4.6 Comparison of mean scores between Survey 1 and Survey 2

<table>
<thead>
<tr>
<th>Paired t-tests*</th>
<th>Diabetes</th>
<th>No Diabetes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical component summary score PCS</td>
<td>649</td>
<td>7842</td>
</tr>
<tr>
<td>Survey 1 -mean (SD)</td>
<td>46.3 (9.7)</td>
<td>51.0 (9.5)</td>
</tr>
<tr>
<td>Survey 2 -mean (SD)</td>
<td>45.4 (10.3)</td>
<td>49.9 (9.9)</td>
</tr>
<tr>
<td>P-value</td>
<td><strong>0.011</strong></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Transformed PCS (APCT)</td>
<td>649</td>
<td>7842</td>
</tr>
<tr>
<td>Survey 1 -mean (SD)</td>
<td>60.5 (23.0)</td>
<td>71.1 (21.3)</td>
</tr>
<tr>
<td>Survey 2 -mean (SD)</td>
<td>58.2 (24.4)</td>
<td>68.6 (22.5)</td>
</tr>
<tr>
<td>P-value</td>
<td><strong>0.005</strong></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Incorporating deaths (APCTD)</td>
<td>720</td>
<td>8184</td>
</tr>
<tr>
<td>Survey 1 -mean (SD)</td>
<td>58.9 (23.6)</td>
<td>70.4 (21.8)</td>
</tr>
<tr>
<td>Survey 2 -mean (SD)</td>
<td>52.5 (29.0)</td>
<td>65.8 (25.9)</td>
</tr>
<tr>
<td>P-value</td>
<td><strong>&lt;0.001</strong></td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*represents the paired t-test results of those women in the two groups who had both Survey measures for PCS, APCT and APCTD.
As shown in Table 4.6, not everyone had a PCS score at Survey 2. This may have been due to missing data for reasons other than dying between Survey 1 and 2. Of the 1119 women who were identified as having an Existing case of diabetes, 937 women had a recorded PCS value at Survey 1.

At Survey 2, only 649 of this group had a PCS value. Table 4.6 considers only women who have a reported PCS value for both Survey 1 and 2. Of these 649 women the change in PCS scores over time is statistically significantly different (difference = -0.9, P-value = 0.011). When the transformation is applied to PCS there is still a statistically significant difference in APCT for these 649 women (difference = -2.3, P-value = 0.005). Furthermore when deaths (APCTD) are accounted for there is an increase in numbers (n=720) thus improving the power of this paired t-test and statistical significance (difference = -6.4, P-value <0.001). A similar scenario is shown in Table 4.6 for the group of women identified as not having diabetes. For women without diabetes and who have PCS, APCT for both Survey 1 and Survey 2, there are statistically significant differences (both P-values <0.001). Likewise for the no diabetes group, the introduction of those who died (APCTD) between Survey 1 and Survey 2 has increased the statistical power (n=8184, difference = -4.6, P-value <0.001). The rationale for presenting both PCS and APCT is to demonstrate the relationship between the measures. The APCT is a function of PCS and thus it would be expected that results by groups (diabetes and no diabetes) are similar. In order to consider the impact of including deaths as zeros, comparisons were made between APCTD and APCT. For future reference to the probability of being health in three years, it will be assumed that deaths are included and thus only APCTD will be discussed in later chapters.
4.6 DISCUSSION

The main aim of this chapter was to determine whether the method to transform the PCS scores suggested by Diehr et al (2003) was valid when a different cohort and follow-up time was used. The results show that the methodology appears sound and the value of the method was illustrated by showing the impact on findings when deaths were included using the APCTD. Ignoring decedents resulted in a bias towards healthy survivors and no differences in changes over time between groups, while incorporating decedents in the analysis resulted in significant differences between groups in changes in health over time. This finding is consistent with results reported in prior publications by Diehr and colleagues (Diehr et al 2003; Diehr et al 2005).

In earlier work Diehr et al (1995) recommended replication of the method with other populations because the Ambulatory Care Quality Improvement Project (ACQUIP) cohort was sicker and older than the general population and included few women. The ALSWH study data was comprised of Australian PCS scores for older women, which allowed the method to be tested in a new setting. But once again, these results may not necessarily be generalisable to other populations.

The strengths and the limitations of this case study using the transformation equation will be discussed in Chapter 8. Also, in Chapter 8 recommendations about the possible and plausible usage of this method will be discussed as well as suggestions about future research which could be developed as a result of this method.
CHAPTER 5:

Analysing longitudinal changes in health related quality of life: A method to adjust for longitudinal missing data
List of Abbreviations used in Chapter 5

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSWH</td>
<td>Australian Longitudinal Study on Women’s Health</td>
</tr>
<tr>
<td>APCT</td>
<td>Probability of being healthy in three years (transformed PCS)</td>
</tr>
<tr>
<td>APCTD</td>
<td>Probability of being healthy in three years (includes deaths)</td>
</tr>
<tr>
<td>APCTDI</td>
<td>Probability of being healthy in three years (includes deaths and imputed data)</td>
</tr>
<tr>
<td>BP</td>
<td>SF-36 subscale domain - Bodily Pain</td>
</tr>
<tr>
<td>FIML</td>
<td>Full Information Maximum Likelihood</td>
</tr>
<tr>
<td>GEE</td>
<td>Generalised Estimating Equation</td>
</tr>
<tr>
<td>GH</td>
<td>SF-36 subscale domain - General Health</td>
</tr>
<tr>
<td>HR-QOL</td>
<td>Health Related Quality of Life</td>
</tr>
<tr>
<td>LISREL</td>
<td>Linear Structural Relations</td>
</tr>
<tr>
<td>MAR</td>
<td>Missing at Random</td>
</tr>
<tr>
<td>MCAR</td>
<td>Missing Completely at Random</td>
</tr>
<tr>
<td>MCS</td>
<td>Mental Component Summary score</td>
</tr>
<tr>
<td>MH</td>
<td>SF-36 subscale domain - Mental Health</td>
</tr>
<tr>
<td>MI</td>
<td>Multiple Imputation</td>
</tr>
<tr>
<td>MI-GEE</td>
<td>Multiple Imputation - Generalised Estimating Equations</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>MNAR</td>
<td>Missing Not at Random</td>
</tr>
<tr>
<td>MRE</td>
<td>Modified Regression Estimates</td>
</tr>
<tr>
<td>PCS</td>
<td>Physical Component Summary score</td>
</tr>
<tr>
<td>PCT</td>
<td>Diehr's Probability of being healthy in one year (transformed PCS)</td>
</tr>
<tr>
<td>PF</td>
<td>SF-36 subscale domain - Physical Functioning</td>
</tr>
<tr>
<td>RP</td>
<td>SF-36 subscale domain - Role Physical</td>
</tr>
<tr>
<td>RE</td>
<td>SF-36 subscale domain - Role Emotional</td>
</tr>
<tr>
<td>REML</td>
<td>Restricted Maximum Likelihood</td>
</tr>
<tr>
<td>SBP</td>
<td>Systolic Blood Pressure</td>
</tr>
<tr>
<td>SF</td>
<td>SF-36 subscale domain - Social Functioning</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short Form 36</td>
</tr>
<tr>
<td>VT</td>
<td>SF-36 subscale domain - Vitality</td>
</tr>
<tr>
<td>WinBUGS</td>
<td>Windows - Bayesian inference using Gibbs Sampling</td>
</tr>
<tr>
<td>WGEE</td>
<td>Weighted Generalised Estimating Equation</td>
</tr>
</tbody>
</table>
5.0 INTRODUCTION

This Chapter is an extension of earlier work from Chapter 4 which transformed the SF-36 Physical Component Summary score (PCS) to incorporate death into a new measure of health related quality of life. The transformed PCS score (APCT) is the estimated probability of being healthy in three years. For people who died the transformed score was set to zero after death (APCTD). This chapter will extend that work by developing a method for imputing longitudinal data that are missing for reasons other than death.

This chapter will consider the important concepts and terminology of Bayesian Inference and the missing data mechanism that were introduced in Sections 3.1.1.1 and 3.1.2.4 respectively. Specifically, this chapter will consider the developments of fully Bayesian techniques for imputing missing data and the application of the most appropriate techniques for longitudinal data where data missing may be missing not at random (MNAR).

5.1 BACKGROUND

Chapter 4 discussed the developments of the general approach to the Diehr et al (2003) PCS transformation. The Diehr transformation equation (see Chapter 4) has been applied to other settings by Fihn et al (2004); Ostbye & Taylor (2004); Strandberg et al (2004) and Diehr et al (2007). However, none of these applications of the Diehr transformation equation have adequately adjusted for missing values caused by reasons other than death. Diehr et al (2003) realized the importance of missing data, arguing for the “superiority of a method that not only accounted for death but also imputed values when the quality of life scores were missing for reasons other than death”. Imputing a score only for participants who had missing
data due to death (and not for other reasons) may bias the average quality of life scores too much. This chapter will attempt to address this potential problem by considering imputation methods for missing outcome (APCT) data (other than due to death).

Previously, in Chapter 3 many of the terms and methods that are the integral parts of the “field” of imputation were introduced and explained. Also, earlier in Section 3.1.1.1, the notion of Bayesian theory versus Frequentist theory was discussed to highlight the appeal of Bayesian theory, especially in the field of imputation. In order to fully understand how imputation techniques, using a Bayesian approach can be applied, the concepts of missing data mechanism (Rubin 1976) need to be understood. In Section 3.1.2.4 the definitions of MCAR, MAR and MNAR were outlined and presented using an example from Schafer and Graham (2002). There was also discussion in Chapter 3 about the current approaches for dealing with longitudinal missing data.

According to Carpenter et al (2002), there are four ways in which data from clinical trials can be analysed when drop-out is a factor:

1. “Discard data from all patients who did not complete the trial and analyse the remaining data.
2. Analyse only the observed data.
3. Use a single or multiple imputation to replace the missing observations with plausible values, then analyse the ‘completed’ dataset.
4. Build a longitudinal model for the data which includes a model for the drop-out process”.

Carpenter et al (2002) suggest that option 4 is the most complex computationally, but it is often the most useful. Carpenter et al (2002) reviewed the reasons for
patients dropping out of clinical trials which in many cases are similar to those of a longitudinal cohort study of elderly populations such as the Australian Longitudinal Study on Women’s Health. To attempt option 4 Carpenter et al (2002) discussed the implications of drop-out for the common methods of analysis and decided to use a similar class of models discussed by Diggle and Kenward (1994). Carpenter et al (2002) fitted models using the BUGS (Bayesian analysis using Gibbs Sampling) software using non-informative priors. They argued that BUGS is more flexible and can fit a variety of models without the effort required for the more familiar approaches using the EM algorithm such as Post et al (2010). The main purpose of their model based approach was to assess “how sensitive the conclusions of a longitudinal study are for missing data” (Carpenter et al 2002). For this Carpenter et al (2002) considered using the non-informative approach of the selection model for drop-out. Carpenter et al (2002) also suggested that their modelling approach can be estimated in either a frequentist or a Bayesian framework. However, they chose the Bayesian framework due mostly to the flexibility and recommend that a “model-based approach to missing data as the most appropriate way to assess the sensitivity of the conclusions to various drop-out mechanisms”. Although they used a longitudinal clinical trial to apply this model-based approach they recommend that “the modelling approach is appropriate for and readily adaptable, to data from a range of longitudinal studies” (Carpenter et al 2002). This chapter will adapt and apply Carpenter and Kenward’s recommended longitudinal missing data approach using a Fully Bayesian framework which can assume, not only missing at random (MAR) but missing not at random (MNAR) (Carpenter & Kenward 2005).
5.2 **AIMS AND OBJECTIVES**

The aim of this chapter is to demonstrate how to account for missing data in longitudinal studies, with a focus on comparing random effects models which impute missing outcome values assuming missing at random (MAR) or missing not at random (MNAR). To achieve this aim longitudinal data analyses are used to examine a case study that compares the longitudinal changes in health-related quality of life for older women with and without diabetes.

5.3 **METHODS**

5.3.1 **Subjects**

The focus of this chapter is the older cohort of ALSWH women for whom Survey 1 was conducted in 1996 (aged 70-75 years, n=12432), Survey 2 in 1999 (aged 73-78 years, 90% retention), Survey 3 in 2002 (aged 76-81 years, 85% retention) and Survey 4 in 2005 (aged 79-84, 84% retention). Between Survey 1 and Survey 2, 529 women died, 101 were withdrawn due to frailty, and 1368 did not complete Survey 2 but were not known to be deceased. Between Survey 2 and Survey 3 there were a further 568 deaths and 262 withdrawn due to frailty, and 2324 women did not complete Survey 3 but were not known to be deceased. Between Survey 3 and Survey 4 there were a further 766 deaths and 380 withdrawn due to frailty and 2667 women did not complete Survey 4.

Women who had missing data, but were not known to be dead, tended to be in poorer health at the previous survey than the longer-term respondents, suggesting data were not missing completely at random (MCAR). It is for this reason that MAR and MNAR will be considered in this chapter.
5.3.2 Measures

Deaths of study participants were identified by linkage to the National Death Index where name, address and date of birth for each woman are matched annually. The ALSWH study office also logged all correspondence with participants and others (for example, family members) notifying the withdrawal from the study or the death of a study participant. These additional death notifications were then compared with the National Death Index to ensure completeness of follow-up for mortality (Powers et al 2000).

5.3.3 Outcome

The transformed physical component summary score (APCT) was generated and validated in Chapter 4 and is shown below:

\[
APCT = \frac{e^{-4.98 + 0.12 \times PCS}}{1 + e^{-4.98 + 0.12 \times PCS}} \times 100
\]

(equation 4.3)

The benefit of using the transformed score is that death is equivalent to a value of zero whereas (as previously mention in Chapter 4) there is no score for death on the PCS scale. Hence for women who died, the APCTD score is set to zero at all time points after their death, as there is no chance they will be healthy at any future time. This new variable, known as APCTD, which will include those who have died, now supersedes the earlier APCT variable. In order to avoid confusion in this chapter, the term APCT will only be briefly mentioned in the results, whereas APCTD is the main focus as it assumes that deaths have been included after baseline (Survey 1) at each of the Surveys 2, 3 and 4.
5.3.4 Covariates

The categorical variables included in the longitudinal analysis and imputation models were: area of residence (0 = non-urban, 1 = urban), baseline age group (0 = 70-72 years, 1 = 73-75 years), smoking status (0 = no and 1 = yes), diabetes status (1 = No diabetes, 2 = New cases and 3 = Existing cases). A discrete variable to represent the number of co-morbidities was generated retrospectively after Survey 4. This variable represented the total number of (un-weighted) co-morbidities reported by participants in Surveys 1-4. The co-morbidities score ranges from 0-5, according to how many of five medical conditions (heart disease, stroke, osteoporosis, asthma and hypertension) had been reported at each survey period.

Classification of diabetes mellitus was determined retrospectively after Survey 4 according to the participant’s response at each survey to the question “Have you been diagnosed with or treated for diabetes in the past three years?” Three categories were defined: Existing Cases (participants who responded “yes” at Survey 1 in 1996); New cases (participants who answered “no” at Survey 1 but subsequently answered “yes” at either Survey 2, 3 or 4) and No Diabetes (participants who answered “no” at all 4 surveys). For the purpose of this case study, all of these covariates were considered time invariant.

5.3.5 Attrition

ALSWH attrition data is available on why participants have left the study. The reasons for attrition are coded as such: 1 = Responded at survey; 2 = Non-responder; 3 = Unable to be contacted; 4 = Too Frail to continue; 5 = Withdrawn from study and 6 = Deceased before survey.
5.4 Statistical Analysis

5.4.1 Multiple Imputation

As discussed previously in Section 3.1.7, an accepted approach to dealing with missing data is to use multiple imputation. The use of multiple imputation is appearing more commonly now in the literature (Horton & Kleinman 2007). There have also been a number of recent articles which discuss the fundamentals of using multiple imputation and how to use the current software (Horton & Kleinman 2007; Spratt et al 2010). Schafer and colleagues have developed a useful website for multiple imputation users and a guide to software implementations can be found at multiple-imputation.com (van Buuren et al 2006; Horton & Kleinman 2007). However, most of the software developments (Schafer 1997; van Buuren & Oudshoorn 1999; Raghunathan et al 2002; Royston 2005) with multiple imputation (see Chapter 3) “have difficulty in incorporating longitudinal information into the imputation methodology of these programs” (Carrigan et al 2007). Fully Bayesian techniques used in WinBUGS are “more suited to longitudinal imputation as they can incorporate hierarchical structure into the modelling process” (Carrigan et al 2007). WinBUGS can, like chained regressions, can deal more systematically with categorical data than some of the other programs available (Carrigan et al 2007).

In some situations, single imputation approaches, such as “mean substitution”, are used. In Chapter 3, issues surrounding single imputation were discussed. In this chapter, the simple longitudinal imputation approach called “last value carried forward” (LVCF) is partly used. This is because it only applies to those who have died and a value of zero is then carried forward as a pseudo last value for the participant. However, to account for missing outcome data (other than due to death) a form of longitudinal multiple imputation will be used which has been adapted from work by Carrigan et al (2007); Carpenter et al (2002) and Carpenter and Kenward
(2005). This chapter will discuss how to adapt and apply the current techniques to longitudinal data.

Carrigan et al (2007) suggest that "Maximum Likelihood approaches to imputation are often intractable in the mainstream software programs" and their implementation is "often violated in complex survey data" since they rely on the strict MAR assumption. There have also been developments in software packages to incorporate MI procedures, and some of these were discussed in Chapter 3. However MI procedures such as, SAS proc mi, rely on the assumption that data are multivariate normal and is known to have difficulty with categorical data (Carrigan et al 2007). Another approach developed for Stata (Royston 2005) known as ice which was discussed in Chapter 3, also assumes MAR for missing data. Stata's ice uses chained equations but can deal with categorical data. These MI procedures are mentioned here in Chapter 5 but are not considered any further at this stage. Stata's ice approach will be applied in later Chapters where the assumptions of missingness will be relaxed.

As previously mentioned, another statistical program which can deal with missing data is WinBUGS. This software can impute missing data using a Fully Bayesian approach. The end result of the technique does not produce multiple complete data sets such as in SAS and Stata but can be considered in a similar way in the final analysis. The WinBUGS Fully Bayesian approach to estimating parameters is achieved using Gibbs Sampling. The following sections will refer and describe to how the Fully Bayesian approach is conducted and how the Gibbs Sampling is achieved. The accounting for missing data will also be discussed and later the approach will be applied to Australian Longitudinal Study on Women’s Health data.
5.4.2 Gibbs Sampling and Missing Data

The most widely used sampling routine is Gibbs sampling because as Kynn (2006) suggests, it is flexible and reliable. For Gibbs sampling, instead of directly sampling from the full joint distribution, samples are drawn from each of the conditional distributions in the model. This means that all of the conditional distributions must be specified. WinBUGS 1.4.1 is the current windows based version of BUGS software described in Spiegelhalter et al (1996). Bayesian inference using Gibbs Sampling (BUGS) is a freely available software program and one of the more commonly used programs. It is considered an improvement on BUGS because of the more “user-friendly, point and click” environment and it now makes the Markov chain Monte Carlo techniques more accessible (Lunn et al 2000). Note that the default method in WinBUGS 1.4.1 is the Gibbs Sampling (for a more detailed description of Gibbs Sampling see Gelman (1995); Schafer (1997); Lunn et al (2000); Gelman et al (2003) and Carrigan et al (2007).

Using Gibbs Sampling, all unknown parameters are estimated conditional on all the other observed data and the other estimated parameters (Carrigan et al 2007). In this longitudinal analysis the missing APCTD would be sampled from a normal distribution with mean ($\mu_i$) and variance ($\sigma^2$). An iteration is complete when all the parameters and missing data APCTD have been estimated. The next iteration then uses these parameter estimates and the observed data to produce new estimates and new values for the originally missing APCTD. The process repeats for a set number of iterations - in the following case study 25000 iterations were used. Carrigan et al (2007) suggest the number of iterations should be greater than 1000 so that convergence to a solution is more likely. There is also burn-in stage (usually 5000 iterations) before the main process begins. The whole process of iteration is started using an initial set of values for each unknown parameter.
5.4.3 An iterative scheme

Initial starting values are chosen for each of the parameters. In this case these parameter estimates were derived using an earlier Maximum Likelihood model using the complete case data. Then, using the conditional distributions and the initial values the next value is drawn. Then, using the values at step 1 the next values are drawn and so on until convergence.

$$\tau^{(i)} \sim p(\tau \mid y, \alpha^{(i)}, \beta^{(i-1)})$$

$$\alpha^{(i)} \sim p(\alpha \mid y, \tau^{(i)}, \beta^{(i-1)})$$

$$\beta^{(i)} \sim p(\beta \mid y, \alpha^{(i)}, \tau^{(i)})$$

Where \( y \) represents the outcome, \( \alpha \) and \( \beta \) respresent parameters and \( \tau \) represents precision.

WinBUGS code for initial starting values:

```plaintext
list(a.area=0.01, a.smok=-0.09, a.diab=c(-0.37, -1.01), a.age=-0.13, a.com=-0.07, a.cons=c(2.3, 2.2, 1.5, 1.9), b0.mu=81.8, b0.tau=0.003, apctd.tau=0.003, b1.mu=-7.01, b1.tau=0.02, b.area=0.38, b.diab=c(-2.16, -14.03), b.age=-2.06, b.com=-1.80, b.smok=-4.0)
```

The estimated precision for APCTD (apctd.tau), random slopes (b1.tau) and random intercepts (b0.tau) are set. Likewise the beta estimates for the random slopes (b1.mu) and intercepts (b0.mu) are set. Since there is a longitudinal logistic regression model for missing there are two sets of estimates required. There are those estimates (b.age, b.smok, etc.) for the model of interest and those for the missing data model (a.age, a.smok, etc.).

\( a\.cons \) represents the estimated constants (\( \alpha \)) at each of the \( t = 4 \) time points.

\( b\.diab=c(-1.35, -6.34) \) represents the intial beta estimates for the categorical variable diabetes groups.
5.4.4 Handling Missing Data in WinBUGS

Kynn (2006) suggests that an advantage of WinBUGS is that it can naturally incorporate missing outcome data. The approach used in WinBUGS is similar to data augmentation. Van Dyk and Meng (2001) suggest that “the term data augmentation refers to methods for constructing iterative optimization or sampling algorithms via the introduction of unobserved data or latent variables”.

\[ Y_{mis} \sim p(Y_{mis} | Y_{obs}, \alpha, \beta, \tau) \]

The notation shown above was introduced in Chapter 3 where the complete data \( Y_{com} \) can be partitioned into two parts: the observed (\( Y_{obs} \)) and the missing (\( Y_{miss} \)) and is denoted as:

\[ Y_{com} = (Y_{obs}, Y_{miss}) \], while \( \alpha, \beta, \tau \) represent parameters.

Unfortunately, there is no easy way to handle missing values in the covariates (Kynn 2006). The sampling requires that the full conditional distributions (i.e. the right hand side of the equations) are specified. So any missing covariate \( x \) values would have to have a prior assigned. Kynn (2006) suggests that WinBUGS will treat the missing point as a parameter to be estimated and will choose it to minimize the error in the model, which can lead to very biased models.

5.4.5 Computation: Models

The following Sections will discuss three models which were considered. These models are:

- Model 1 - Complete-Case Model (ignorable missing)
- Model 2 - MAR Longitudinal imputation model
- Model 3 - MNAR Longitudinal imputation model (non-ignorable missing).
Model 1 Complete Case inference

Let $Y_{ij}$ be a continuous variable (APCTD): the probability of being healthy in three years.

Let $i = \text{the } i^{th} \text{ subject where } i = (1...i)$

$t = \text{the } t^{th} \text{ time where } t = (1...t)$

Model of Interest (and non-informative priors):

$Y_{it} = \beta_{0i} + \beta_1 \text{Time}_i + \beta_2 \text{Age}_i + \beta_3 \text{Comorb}_i + \beta_4 \text{Smoke}_i + \beta_5 \text{Diab}_{1i} + \beta_6 \text{Diab}_{2i} + \varepsilon_i$

where

$\beta_{6 \text{Diab}_{1i}} = \text{the effect size difference between New cases and Existing cases groups}$

$\beta_{6 \text{Diab}_{2i}} = \text{the effect size difference between No diabetes and Existing cases groups}$

$Y_{it} \sim N(\mu_{it}, \tau_{it}^{-1})$

$\mu_{it} \sim N(0, 10000)$

$\tau_{it} \sim \text{Gamma}(0.01, 0.01)$

random subject intercept - $\beta_{0i} \sim N(\mu_{0t}, \tau_{0t}^{-1})$

$\mu_{0t} \sim N(0, 10000)$

$\tau_{0t} \sim \text{Gamma}(0.01, 0.01)$

random time slope - $\beta_{1t} \sim N(\mu_{1t}, \tau_{1t}^{-1})$

$\mu_{1t} \sim N(0, 10000)$

$\tau_{1t} \sim \text{Gamma}(0.01, 0.01)$

5.4.6 Missing data patterns - Monotonic or Non-monotonic?

Imputing a zero has replaced the drop-outs due to death with a value. This has reduced the amount of missing data and has dealt with the monotonic missing pattern of those who have died. However, there are still monotonic missing data for
those who have dropped out due to frailty and possible dementia. There are also those participants whom the study has lost contact with. Another issue is intermittent missing outcomes (PCS). This may have occurred when participants missed completing a survey but returned to the study. Likewise a PCS score was unable to be calculated since too many items were missing from the survey. These types of missing data are known as non-monotonic. Hence the development of a longitudinal multiple imputation approach is useful because it can impute for both monotonic and non-monotonic missing data. For the purpose of the imputation analysis, the types of missing data patterns will be considered below. However, during the analysis the possible nature of why data were unobserved will not be differentiated. Reasons for doing so will be discussed later. The following considerations mentioned below (Twisk 2003) are more of an exploration at this stage.

5.4.7 Missing Data patterns

There will be an investigation to determine whether missing data are dependent on values of the outcome variable Y. Twisk (2003) suggests comparing the participants with data at t = t with participants who are missing data at t = t. The comparison is then carried out on APCT at t = t - 1. The difference between the groups can be compared using an independent sample t-test. The analysis to compare differences in APCT (t = t - 1) by the reasons for attrition with those who responded at the next survey (t = t) was conducted using oneway ANOVAs with Bonferroni multiple comparisons t-tests. The comparisons were also considered by diabetes classifications.
5.4.8 Inference under ignorable missing data

Model 2 - MAR longitudinal imputation model; ignorable missing data mechanism

The structure of the MAR longitudinal imputation model uses a random effects model to incorporate random subject intercept and random time slope (coefficient) (N=8368 women). The model has a random intercept in the imputation component of the model to incorporate within subject correlation in APCTD and hence take account the longitudinal study design (Carrigan et al 2007).

Note the complete case model (N=5451) and the MAR imputation model have the same model structure, the difference is that when APCTD data were missing in the MAR imputation model they were coded as “NA”. The use of “NA” in WinBUGS represents that a value is missing unlike other programs that use a dot “.”.

Let \( Y_{it} \) be a continuous variable (APCTD): the probability of being healthy in three years.

Let \( i = \) the \( i^{th} \) subject where \( i = (1...i) \)

\( t = \) the \( t^{th} \) time where \( t = (1...t) \)

Model of Interest (and non-informative priors):

\[
Y_{it} = \beta_{0i} + \beta_{1} Time_i + \beta_{2} Age_i + \beta_{3} Area_i + \beta_{4} Comorb_i + \beta_{5} Smoke_i + \beta_{6} Diab_{1i} + \beta_{7} Diab_{2i} + \epsilon_i
\]

where

\( \beta_{6} Diab_{1i} \) = the effect size difference between New cases and Existing cases groups

\( \beta_{7} Diab_{2i} \) = the effect size difference between No diabetes and Existing cases groups
\[ Y_{it} \sim N(\mu_{it}, \tau_{it}^{-1}) \]
\[ \mu_{it} \sim N(0, 10000) \]
\[ \tau_{it} \sim \text{Gamma}(0.01, 0.01) \]

random subject intercept - \( \beta_{0i} \sim N(\mu_{0t}, \tau_{0t}^{-1}) \)
\[ \mu_{0t} \sim N(0, 10000) \]
\[ \tau_{0t} \sim \text{Gamma}(0.01, 0.01) \]

random time slope - \( \beta_{1t} \sim N(\mu_{1t}, \tau_{1t}^{-1}) \)
\[ \mu_{1t} \sim N(0, 10000) \]
\[ \tau_{1t} \sim \text{Gamma}(0.01, 0.01) \]

### 5.4.9 Inference under non-ignorable missing data

**Model 3 - MNAR longitudinal imputation model; non-ignorable missing data mechanism.**

In model 2 above, MAR was assumed. In this model, MNAR was considered and the sensitivity of the results about the effect of diabetes to the assumption of MAR, were assessed. According to Carpenter and Kenward (2005), the MNAR assumption means that there needs to be a model which identifies the reason for missing values (in this case APCTD). This takes the form of a logistic regression where the dependent binary variable is 1 if APCTD is observed for participant \( i \) at time \( t \) and 0 if otherwise.

Model 3 was considered the same as the model of interest shown above with an additional MNAR component, a logistic regression model for the probability of missing. The outcome in this longitudinal logistic regression model is the probability that the APCTD value is observed based on the data from Surveys 1, 2, 3 and 4. The covariates used in the longitudinal logistic regression model were age, area, diabetes, smoking status and co-morbidities.
Model of Interest (and non-informative priors):

\[ Y_{it} = \beta_0 + \beta_1 \text{Time}_i + \beta_2 \text{Age}_i + \beta_3 \text{Area}_i + \beta_4 \text{Comorb}_i + \beta_5 \text{Smoke}_i + \beta_6 \text{Diab}_1i + \beta_7 \text{Diab}_2i + \epsilon_i \]

where

\[ \beta_6 \text{Diab}_2i = \text{the effect size difference between New cases and Existing cases groups} \]

\[ \beta_7 \text{Diab}_2i = \text{the effect size difference between No diabetes and Existing cases groups} \]

\[ Y_{it} \sim N(\mu_{it}, \tau_{it}^{-1}) \]

\[ \mu_{it} \sim N(0, 10000) \]

\[ \tau_{it} \sim \text{Gamma}(0.01, 0.01) \]

random subject intercept - \[ \beta_0 \sim N(\mu_{0t}, \tau_{0t}^{-1}) \]

\[ \mu_{0t} \sim N(0, 10000) \]

\[ \tau_{0t} \sim \text{Gamma}(0.01, 0.01) \]

random time slope - \[ \beta_{1t} \sim N(\mu_{1t}, \tau_{1t}^{-1}) \]

\[ \mu_{1t} \sim N(0, 10000) \]

\[ \tau_{1t} \sim \text{Gamma}(0.01, 0.01) \]

There are a number of features which have been adapted from Carrigan et al (2007) in order to produce the longitudinal imputation model. The main features were to describe the distributions of the random intercepts (\( b_{\text{int}}[i] \)) and random slopes (\( b_{\text{slo}}[i] \)) in the model of interest (shown below).

\[ i_{\text{reg}}[i] <- b_{\text{int}}[i] + b_{\text{age}} \times \text{age}[i] + (b_{\text{diab}}[1] \times \text{equals(diab}[i],2)) \]
\[ + (b_{\text{diab}}[2] \times \text{equals(diab}[i],1)) + b_{\text{comorb}}[i] + b_{\text{smok}} \times \text{smoke}[i] \]
\[ + b_{\text{area}} \times \text{area}[i]; \]

\[ b_{\text{int}}[i] \sim \text{dnorm}(b_{0\mu}, b_{0\tau}); \]
\[ b_{\text{slo}}[i] \sim \text{dnorm}(b_{1\mu}, b_{1\tau}); \]

The model shown above feeds into the following model which involves the random slopes.

\[ \text{apctd.m}[i,t] <- i_{\text{reg}}[i] + b_{\text{slo}}[i] \times (t-2.5); \]
5.4.9.1 Informative priors for the models

The following represent the Informative priors for parameters in model of interest. They assume a normal distribution with a mean of zero and a variance of 0.0001.

\[ \beta_{2i} \sim N(0.0, 0.0001), \beta_{3i} \sim N(0.0, 0.0001), \beta_{4i} \sim N(0.0, 0.0001), \beta_{5i} \sim N(0.0, 0.0001) \]
\[ \beta_{6i} \sim N(0.0, 0.0001), \beta_{7i} \sim N(0.0, 0.0001) \]

The following represent the Informative priors for logistic regression parameters in the missing data model. They assume a normal distribution with a mean of zero and a variance of 0.0001.

\[ \alpha_{1i} \sim N(0.0, 0.0001), \alpha_{2i} \sim N(0.0, 0.0001), \alpha_{3i} \sim N(0.0, 0.0001), \alpha_{4i} \sim N(0.0, 0.0001) \]
\[ \alpha_{5i} \sim N(0.0, 0.0001), \alpha_{6i} \sim N(0.0, 0.0001), \alpha_{7i} \sim N(0.0, 0.0001) \]

The following represent the informative priors for random intercept that assume a normal distribution with a mean of zero and a variance of 0.0001.

\[ \beta_{0i} \sim N(0.0, 0.0001), \beta_{0i} \sim \text{gamma}(0.01, 0.01) \]

The following represent the Informative priors for random slope that assume a normal distribution with a mean of zero and a variance of 0.0001.

\[ \beta_{3i} \sim N(0.0, 0.0001), \beta_{3i} \sim \text{gamma}(0.01, 0.01) \]

The following represent the Informative priors for logistic regression constant that assumes a normal distribution with a mean of zero and a variance of 0.0001.

\[ \alpha_{0it} \sim N(0.0, 0.0001) \]

5.4.9.2 Explanation of the MNAR model

The following modelling structure is a novel approach for a longitudinal observational study. It has been developed from the established work by Carrigan et al (2007) on longitudinal multiple imputation using WinBUGS and the Carpenter & Kenward (2005) example of assuming MNAR using WinBUGS. The final MNAR
produced here has adapted their work which incorporates the longitudinal imputation and the assumption of MNAR.

The MNAR model, like the MAR model, draws upon the data which includes missing data. The total of participants (i) and how many time points (t) needs to be established before the model of interest, can be dealt with. For loops in the code: for (i in 1:8368), are used to draw upon the earlier loaded dataset. In this case there are t=4 time points for i=8368 participants. All participants have 4 time points included in the dataset because a missing outcome value is recognized as “NA”. These missing values will be replaced with an imputed value as the modelling is conducted.

Model {
   for (i in 1:8368) for (t in 1:4) {

5.4.10 Determination of the coefficient $\gamma$ in the non-ignorable missing data mechanism

The APCTD variable was also included in the logistic regression model with a coefficient $\gamma$, as suggested by Carpenter and Kenward (2005). To determine an appropriate value for $\gamma$ to be used in the imputation component of the MNAR model, a range of values for $\gamma$ which included 0, 0.001 and 0.01 were considered. Results from the analysis used in the MNAR model suggested that the $\gamma$ coefficient = 0.001 was a reasonable approximation based on the models (results not shown). The statistical analyses for the MAR and MNAR modeling were conducted using WinBUGS1.4.1. The imputation approach applied in WinBUGS used a longitudinal multiple imputation. This is adapted from earlier work by Carpenter and Kenward (2005) and Carrigan et al (2007).
The logistic regression model and the $\gamma$ coefficient

$$\logit(\text{obs}_{i,t}) = \alpha_0 + \gamma \text{APCTD}_{i,t} + \alpha_1 \text{Age}_{i,t} + \alpha_2 \text{Area}_{i,t} + \alpha_3 \text{Comorb}_{i,t} + \alpha_4 \text{Smoke}_{i,t} + \alpha_5 \text{Diab1}_{i,t} + \alpha_6 \text{Diab2}_{i,t} + \epsilon_{i,t}$$

Constants in logistic regression model:

- $\alpha_0 \sim N(\mu_{0t}, \tau_{0t})$
- $\mu_{0t} \sim N(0, 0.0001)$
- $\tau_{0t} \sim \text{Gamma}(0.01, 0.01)$

In order to determine whether the missing outcome APCTD is missing not at random a binary variable was generated called observed. A Bernoulli distribution was used for this binary value. The WinBUGS code used was:

```winbugs
obs[i,t] ~ dbern(p.obs[i,t]);
logit(p.obs[i,t]) <- a.cons[t] + 0.0001*cut.apctd[i,t] + i.missing[i];
i.missing[i] <- (a.diab[1]*equals(diab[i],2)) + (a.diab[2]*equals(diab[i],1) + a.com*comorb[i] + a.smok*smoke[i] + a.age*age[i] + a.area*area[i];
```

$i.missing[i]$ is used to bring together the participants covariates and feeds them into the longitudinal logistic regression at each time point $t$. Since they are all baseline covariates and not missing the main part of this logistic model is the addition of the outcome which does have missing data. The $\gamma$ coefficient is required and is mentioned above.

$a.diab[1], a.age, a.area$ represent the parameter estimates that have are calculated at each iteration and were initially set at the beginning.

$\text{diab[i], smoke[i], comorb[i]}$ represent the observed data that the model draws upon. As in the case of the model of interest the parameter estimates for the
longitudinal logistic regression were established using Maximum Likelihood estimates from a logistic regression model and the observed data.

5.4.11 Using the cut(y) function in WinBUGS

Pattern–mixture models and selection models are two paths leading to the same joint distribution and have been used in clinical trials. Ekholm and Skinner (1998) used a pattern–mixture model and a selection interpretation for their sensitivity analysis. Similarly, the pattern–mixture model was related to a selection model to offer some reassurance that the assumed pattern–mixture models provide sensible implications about the selection mechanism (Kaciroti et al 2008). This can be derived in WinBUGS by adding a logistic regression that relates the missing data indicator to the full data (observed and imputed in step (b)). This part was implemented in WinBUGS by using the cut(y) function to act as a valve to control the flow of the information from the pattern–mixture model to the selection model, but not vice versa (Kaciroti et al 2008). This concept of controlling the flow of information will be adapted and used.

The WinBUGS code used was:

```winbugs
cut.apctd[i,t] <- cut(apctd[i,t]);
apctd[i,t] ~ dnorm (apctd.mu[i,t],apctd.tau);
apctd.mu[i,t] <- i.reg[i] + b.slo[i]*(t-2.5);
```

5.4.12 Using WinBUGS

To run an analysis in WinBUGS1.4.1 there are four basic steps:

i) specify a model;

ii) load the data;

iii) specify initial values; and

iv) run the Markov chain sampler (Carrigan et al 2007).
The analyses of the imputation models were performed in WinBUGS1.4.1. Each model was run for 25,000 iterations with an additional 5,000 iterations for burn-in (Carrigan et al. 2007). The data and the programs that were used to analyse them are shown in Appendix 5.1.

The uncertainty about the missing data imputations is automatically included in the Bayesian inference and so unlike MI methods in Stata or SAS there are no multiple datasets to combine. Examples of the graphs from the MAR modelling process are found in Appendix 5.2. The graphs shown in Appendix 5.2 are for the diabetes groups and include: iterations, dynamic traces, auto-correlations, running quantiles and density graphs. Results from the WinBUGS1.4.1 analysis were transferred to SAS 9.1 to estimate the mean quality-of-life scores at each survey according to diabetes status, after adjusting for covariates, using the least squares means (LSmeans) option with 95% confidence intervals (SAS 9.1 2003). These results were then displayed graphically. In order to compare adjusted LSmeans with and without imputed values for APCTD, the dataset derived from the WinBUGS analysis was converted to a SAS file before analyses. Independent t-tests were used to compare the size of the difference between groups of women with and without diabetes at each of the four surveys cross-sectionally (SAS 9.1 2003). Pearson’s Chi-squared tests were used to compare the proportion of missing data in each group at each survey and for comparing the proportion of deaths in each group at each survey (SAS 9.1 2003). Due to the large sample size a much smaller level of significance (i.e. P-value <0.001) is considered appropriate since very small differences can be statistically significant at the standard α level of 5%.
5.5 RESULTS

5.5.1 Overall results from ALSWH – Old Cohort

The percent of respondents missing PCS scores at Survey 1, 2, 3 and 4 were 13%, 8%, 17% and 13% respectively. Of the 12,432 participants in the original dataset, 9,181 women were able to be classified into one of the three diabetes categories. Participants with missing values for diabetes status (n=3,251) or who had inconsistent or missing responses on the covariates (n=813) were not included in the analysis. Thus the sample size used was 8,368 participants who had zero to four of the repeated outcome measures (PCS) and all covariate data.

Since APCT represents a transformed PCS then APCT=APCTD when a PCS value is not missing. If a PCS value was missing then so were APCT and APCTD. Those who died were imputed with a value of zero and this was carried forward thereafter. Table 5.1 shows the percent of respondents classified as an existing case (n=898) who had missing APCTD values at survey 1, 2, 3 and 4. Table 5.1 also shows the number of respondents classified as New cases (n=640) and the number of respondents without diabetes (n=6,830) with missing APCT values. As previously mentioned if the PCS is missing then so will the APCT be missing. Table 5.1 shows that over time the proportion of missing values increases in all diabetes groups. Those without diabetes had significantly less missing data than existing cases of diabetes at each survey (P-value <0.001). While those considered New cases have statistically significant differences in the proportion of missing data to the existing cases at surveys 2, 3 and 4 (P-value <0.001). Table 5.2a shows the proportion of women who withdrew between surveys is similar to the proportion of women who died between surveys. Table 5.2b shows, the proportion of deaths in women with Existing cases of diabetes was statistically significantly higher than those without diabetes at Survey 2, 3 and 4 (P-value <0.001). While the proportion of deaths in
women with Existing cases was also statistically significantly higher than the New cases at survey 3 and 4. No deaths had occurred in the New Cases group until between Survey 2 and Survey 3. Table 5.1 suggests that the proportion of women with missing APCT, for reasons other than death at Survey 3, increased in all diabetes groups. This may have been due to a greater number of items that could not be accounted for during the calculation of PCS. The length of the survey at Survey 3 may have also added to the missing number of questions in this round. The number of questions asked in Survey 4 was reduced. Table 5.1 shows that the missing PCS (APCT) data declines below levels at Survey 3, this decline occurs across all diabetes groups. Table 5.2b shows that by Survey 4 the Existing cases group have lost 31% to death compared to 17% for the No Diabetes group of women. This difference in proportions is statistically significant (P-value <0.001).

Table 5.1  Diabetes status defined retrospectively from 2005

| Diabetes Status | No Diabetes  
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(N= 6830)</td>
</tr>
<tr>
<td></td>
<td>Survey 1</td>
</tr>
<tr>
<td>1996</td>
<td>639</td>
</tr>
<tr>
<td>1999</td>
<td>9%</td>
</tr>
<tr>
<td>2002</td>
<td>76</td>
</tr>
<tr>
<td>2005</td>
<td>12%</td>
</tr>
<tr>
<td>(N= 898)</td>
<td>127</td>
</tr>
<tr>
<td></td>
<td>14%</td>
</tr>
</tbody>
</table>

Table 5.2a  Overall Mortality and Withdrawals

<table>
<thead>
<tr>
<th>Survey</th>
<th>Overall Mortality</th>
<th>%</th>
<th>Withdrawn</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey 1-2 (1996-1999)</td>
<td>529</td>
<td>4%</td>
<td>368</td>
<td>3%</td>
</tr>
<tr>
<td>Survey 2-3 (1999-2002)</td>
<td>569</td>
<td>5%</td>
<td>779</td>
<td>6%</td>
</tr>
<tr>
<td>Survey 3-4 (2002-2005)</td>
<td>769</td>
<td>6%</td>
<td>888</td>
<td>7%</td>
</tr>
</tbody>
</table>
Table 5.2b  Cumulative number (%) of deaths by diabetes status

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Diabetes (N=6830)</td>
<td>389</td>
<td>758</td>
<td>1181</td>
</tr>
<tr>
<td></td>
<td>6%</td>
<td>11%</td>
<td>17%</td>
</tr>
<tr>
<td>New Cases (N=640)</td>
<td>0</td>
<td>13</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>2%</td>
<td>8%</td>
</tr>
<tr>
<td>Existing Case (N=898)</td>
<td>91</td>
<td>172</td>
<td>274</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>19%</td>
<td>31%</td>
</tr>
<tr>
<td>Total Deaths</td>
<td>480</td>
<td>943</td>
<td>1503</td>
</tr>
</tbody>
</table>

5.5.2 Missing Data Patterns for those with Diabetes

Oneway ANOVAs with Bonferroni multiple comparison t-tests were used in Table 5.3 in order to determine whether missing data patterns may be ignorable. The purpose of using the Bonferroni multiple comparison t-tests was to compare the APCT between responders and non-responders (i.e. reason for attrition). These comparison tests were conducted for all diabetes groups. Tables 5.3, reports the mean APCT and standard deviation for the Existing Diabetes (sicker) group from surveys 1, 2 and 3 by the attrition status at survey 2, 3 and 4. Results from Bonferroni multiple comparison t-tests suggest that the missing APCT may not be ignorable. There are statistically significant differences in APCT between responders and those deceased at each survey. Those women who died had a lower mean APCT, at the previous survey compared to responders (P-value <0.0001, P-value <0.0001, P-value <0.0001). Likewise those who could not be contacted at Survey 4 have a statistically significant lower mean APCT at Survey 3 to those who died between Surveys 3 and 4 (P-value <0.0001). Notably those with diabetes have an increasingly greater proportion of attrition than those without
diabetes. Figure 5.1 highlights the differences in the attrition categories at survey 3 for all diabetes classifications between surveys 2 and 3.

Table 5.3  Probability of being healthy in 3 years by Attrition classifications for Existing Cases

<table>
<thead>
<tr>
<th>Attrition** Categories at next survey 2,3,4</th>
<th>APCT from Survey1</th>
<th>APCT from Survey2</th>
<th>APCT from Survey3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Responders</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>708</td>
<td>538</td>
<td>367</td>
</tr>
<tr>
<td>Mean</td>
<td>60.3*</td>
<td>59.3*</td>
<td>56.4*</td>
</tr>
<tr>
<td>SD</td>
<td>23.3</td>
<td>23.7</td>
<td>23.2</td>
</tr>
<tr>
<td>Did not do survey</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>59</td>
<td>60</td>
<td>22</td>
</tr>
<tr>
<td>Mean</td>
<td>54.6</td>
<td>66.0</td>
<td>47.5</td>
</tr>
<tr>
<td>SD</td>
<td>22.9</td>
<td>21.7</td>
<td>23.5</td>
</tr>
<tr>
<td>Unable to Contact</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>30</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Mean</td>
<td>51.2</td>
<td>57.8</td>
<td>56.5</td>
</tr>
<tr>
<td>SD</td>
<td>22.4</td>
<td>28.4</td>
<td>20.3</td>
</tr>
<tr>
<td>Frail</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>12</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>Mean</td>
<td>58.8</td>
<td>56.1</td>
<td>42.9</td>
</tr>
<tr>
<td>SD</td>
<td>21.4</td>
<td>27.3</td>
<td>20.7</td>
</tr>
<tr>
<td>Withdrawn</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>51</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>Mean</td>
<td>56.1</td>
<td>61.4</td>
<td>58.2</td>
</tr>
<tr>
<td>SD</td>
<td>24.7</td>
<td>22.7</td>
<td>26.4</td>
</tr>
<tr>
<td>Deceased</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>77</td>
<td>69</td>
<td>54</td>
</tr>
<tr>
<td>Mean</td>
<td>45.0*</td>
<td>45.4*</td>
<td>41.5*</td>
</tr>
<tr>
<td>SD</td>
<td>23.6</td>
<td>26.4</td>
<td>22.8</td>
</tr>
</tbody>
</table>

* P-value <0.001, Bonferroni multiple comparisons  
** Misusing Data patterns Suggested by Twisk (2003)

The missing data patterns for APCT suggested that the missingness was not monotone but mostly intermittent. Results from that analysis shown in Tables 5.3 suggest that women with an Existing case of diabetes and who died before survey 2 had on average a lower APCT score at survey 1 than those who responded at Survey 2 (P-value <0.001). Women with an existing case who could not be contacted (non-responders) at Survey 3 on average had a higher APCT score at Survey 2 than those who responded at Survey 3 (as shown by non-overlapping 95% confidence interval in Figure 5.1). Based upon this information about missingness it was decided to produce models using both MAR and MNAR.
* P-value < 0.001, Bonferroni multiple comparisons t-test

Figure 5.1  Mean APCT and Reason for Attrition between Survey 2 and 3 by Diabetes groups
Figure 5.2 represents the changes in the adjusted mean APCT (mean probability of being healthy in three years x 100) by diabetes status, prior to including zero scores for participants who died. Note that since the APCT is a direct transformation of PCS, the trends in the two sets of scores are similar for each diabetes group and differ only by the scale of the variables (so the PCS results are not shown). Figure 5.2 suggests that the trend in APCT decline over time appears to be similar across the three groups. The mean APCT for those with diabetes compared to those without diabetes is statistically significant at each survey (P-Value<0.001). The New cases come from the No diabetes over time and track closely to this no Diabetes group.
Figure 5.3 represents the changes in the adjusted mean APCTD (probability of being healthy in three years) after including scores for participants who died (i.e. APCTD= 0). In this analysis it can be seen that the adjusted mean APCTD decreases over all four surveys for women with diabetes. Figure 5.3 illustrates that the reinstating of deaths (zeros) has produced significant decreases in the mean probabilities of being healthy, particularly for the group of women with existing diabetes at survey 1. The mean difference between New cases and No Diabetes groups is not statistically significant over the course of the 4 Surveys, as represented by the overlapping 95% confidence intervals in Figure 5.3. There were statistically significant differences at each survey for Existing cases versus New cases and for Existing cases versus No Diabetes (P-value <0.001). These
differences are represented by Figure 5.3 where the 95% confidence intervals are not overlapping.

![Graph](image)

Figure 5.4 APCTDI – Mean Probability of being healthy in 3 years (Deaths and imputed values included), with 95% confidence intervals

Figure 5.4 represents the changes in the adjusted average APCTDI (probability of being healthy in three years) with the inclusion of scores for participants who died and imputed values for scores that were missing for other reasons. An important note here is that the imputations for APCTD (now known as APCTDI) were from the WinBUGS analysis mentioned earlier. Figure 5.4 shows a decline in mean APCTDI over the four surveys for women with diabetes. Cross-sectional comparisons of the mean APCTDI for women with and without diabetes is statistically significant for each of the four surveys (P-value<0.0001). Comparing Figures 5.3 and 5.4 suggest
that the existing cases group has higher mean probabilities of being healthy at surveys 2–4 when imputed values are included. This demonstrates the impact of replacing the larger number of missing values for the existing cases as participants were possibly dropping out of study over time due to frailty.

Results from Table 5.4, show adjusted coefficients from three random intercept models that were produced using WinBUGS1.4.1. Results are presented as mean coefficients after the 25000 iterations have been conducted. The main focus of these results is the comparisons of diabetes groups across the three models. The first random intercept model (a complete case model) where the outcome is APCT supports Figure 5.2 that surviving participants with and without diabetes have statistically significant different mean APCT. The second complete case model (where the outcome is APCTD) suggests that the inclusion of those who have died leads to an increase in the coefficients for the diabetes categories. While the results for the MAR and MNAR models show that the coefficients for diabetes are only marginally larger for the MAR model. However the 95% confidence limits for the MAR model were narrower. There were only slight differences between the MNAR and MAR models with the other covariates.
Table 5.4  Comparisons of regression coefficients for predicting APCTD

<table>
<thead>
<tr>
<th>Covariates</th>
<th>Complete Case APCT(^\ast) (n= 4352(^\ast))</th>
<th>Complete Case APCTD (n= 5451(^\ast))</th>
<th>Imputed model APCTDI ( N= 8368(^&amp;))</th>
<th>Imputed model APCTDI ( N= 8368(^&amp;))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (95% PI)**</td>
<td>Mean (95% PI)</td>
<td>MAR Mean (95% PI)</td>
<td>MNAR Mean (95% PI)</td>
</tr>
<tr>
<td>Intercept</td>
<td>80.19 (79.06, 81.32)</td>
<td>62.49 (61.22, 63.78)</td>
<td>63.27 (62.18, 64.31)</td>
<td>63.22 (62.07, 64.31)</td>
</tr>
<tr>
<td>Time (Survey)</td>
<td>-4.06 (-4.26, -3.86)</td>
<td>-7.15 (-7.41, -6.90)</td>
<td>-7.51 (-9.57, -5.42)</td>
<td>-7.04 (-7.26, -6.82)</td>
</tr>
<tr>
<td>No Diabetes New Cases</td>
<td>reference</td>
<td>reference</td>
<td>reference</td>
<td>reference</td>
</tr>
<tr>
<td>Existing Cases</td>
<td>-4.61 (-6.61, -2.66)</td>
<td>-1.79 (-4.35, 0.67)</td>
<td>-1.86 (-3.81, 0.09)</td>
<td>-1.83 (-3.80, 0.16)</td>
</tr>
<tr>
<td>Co-morbidities</td>
<td>-7.12 (-7.23, -6.74)</td>
<td>-0.63 (-0.56, 0.63)</td>
<td>-1.10 (-1.10, -0.14)</td>
<td>-1.10 (-1.10, -0.14)</td>
</tr>
<tr>
<td>age</td>
<td>-1.16 (-2.19, -0.14)</td>
<td>-3.73 (-4.95, -2.49)</td>
<td>-4.10 (-4.09, -1.96)</td>
<td>-3.01 (-4.09, -1.96)</td>
</tr>
<tr>
<td>smoke</td>
<td>-1.52 (-3.77, 0.78)</td>
<td>-8.74 (-11.22, -6.19)</td>
<td>-9.57 (-9.58, -5.55)</td>
<td>-7.54 (-9.58, -5.55)</td>
</tr>
<tr>
<td>area</td>
<td>0.45 (-0.59, 1.52)</td>
<td>0.89 (-0.37, 2.16)</td>
<td>0.67 (-0.38, 1.72)</td>
<td>0.71 (-0.39, 1.77)</td>
</tr>
</tbody>
</table>

\(^\ast\)APCT=APCTD since no deaths are included only surviving participants  
**2.5%, 97.5% Posterior Limits, PI=Posterior Interval  
\(^\ast\)N= 5451 represents the number of participants per survey with observed APCTD values (Random intercept and coefficient mixed model using WinBUGS).  
\(^\&\)N= 4352 represents the number of participants per survey with observed APCT values (Random intercept and coefficient mixed model using WinBUGS).  
\(^\&\)N= 8368 represents the number of participants per survey with (8368 x 4 =) 33472 observed APCTD values (Random intercept and coefficient mixed model using WinBUGS).
5.6 DISCUSSION

According to Figure 5.2, the result from Survey 4 suggests that for older women with existing cases of diabetes the mean probability of being healthy at survey 5 is 48% (95% CI: 45%, 51%). However the APCT transformation does not account for deaths and there is evidence from Table 5.2b that mortality rates are higher among women with diabetes than women without diabetes. Hence the mean probability of being healthy in 3 years for this group is more than likely overestimated.

![Figure 5.5](image-url)  
**Figure 5.5** APCTDI and APCTD overlaid to show differences
Finally, on comparing with and without imputed outcome (APCTD) data, Figure 5.5 is an illustration of when Figures 5.3 and 5.4 are overlaid (excluding the new cases). When missing values were imputed for both groups (known as APCTDI) the impact is greater on the Existing Cases, especially by Survey 3. The imputing of missing APCT at Survey 3 and Survey 4 for the Existing cases group increased the mean APCT (deaths included) and enlarged the group. This noticeable difference in the Existing cases' graphs for APCTD and APCTDI can be accounted for by the fact that in Table 5.1 the percentage of missing APCT was much greater for Existing cases than for the No diabetes group. Table 5.1 shows that for those women still alive in the Existing cases group for Surveys 3 and 4, they had 36% and 30% respectively, missing APCT (and PCS) compared to the No diabetes group who had 16% and 10% missing APCT.

Earlier work in Chapter 4 concluded that the use of a transformation of the PCS score to include death was justifiable. The retention rates indicate that to simply disregard those who have died and have not responded (withdrawn and/or missing) at the different time points may result in misleading conclusions about changes in HR-QOL. Other approaches to account for death and impute for missing values in longitudinal data are emerging in the literature (Diehr et al 2005; Diehr et al 2007; Kazis et al 2007; Shardell et al 2008) however these approaches have limitations as highlighted in the introduction. The method used in this chapter is a novel attempt to account for non-monotonic missing outcome data using Fully Bayesian analysis.

An advantage of including missing data and deaths is the improved statistical power for the study by increasing the sample size, which is very important when studying an older ageing cohort. It is also apparent from the models that low PCS scores at
one survey are associated with missing PCS scores at the following survey. Hence there is a strong case for assuming MNAR when imputing these outcome values.

The strengths and the limitations of using this longitudinal multiple imputation method will be discussed in Chapter 8. Also, in Chapter 8, recommendations about the possible and plausible usage of this method will be discussed. There will also be suggestions about future research which could be developed as a result of this method.
CHAPTER 6:

Methods for Accounting for Deaths –

Generating a PCS value for death using a health utility index (SF-6D)
List of Abbreviations used in Chapter 6

ADL Activities of Daily Living
ALSWH Australian Longitudinal Study on Women’s Health
ANOVA Analysis of variance
APCTD Probability of being healthy in three years (transformed PCS / includes deaths)
BP Bodily Pain dimension (derived from SF-36)
EVGGFP Excellent / Very Good/ Good / Fair / Poor
GH General Health dimension (derived from SF-36)
HR-QOL Health Related Quality of Life
HALex Health and Activity Limitation Index
HOS Health Outcomes Survey
HUI Health Utility Index
IADL Instructional Activities of Daily Living
MCS Mental Component Summary score
MH Mental Health dimension (derived from SF-36)
PCS Physical Component Summary score
PCSD PCS with death score included
RE Role Emotional dimension (derived from SF-36)
PF Physical Functioning dimension (derived from SF-36)
RP Role Physical dimension (derived from SF-36)
RTI Research Triangle Institute
SF Social Functioning dimension (derived from SF-36)
SF-6D Short Form - 6 Domains (Brazier)
SF-12 Short Form -12
SF-36 Short Form -36
SF-36 v1 Short Form -36 Version 1
SF-36 v2 Short Form -36 Version 2 1996
VT Vitality dimension (derived from SF-36)
6.0 INTRODUCTION

This Chapter will discuss another approach to accounting for deaths in longitudinal studies using Health Related-Quality of Life (HR-QOL) summary measures. Earlier in Chapter 2, a background into health utility indexes was discussed in preparation for the methodology presented in this chapter. This chapter will discuss the use of the Health and Activity Limitation Index (HALex), developed by Erickson (1998) and used by Trisolini et al (2005) to develop a PCS value for those cohort participants who die. This chapter will discuss and consider the possibility of an approach that uses another Health Utility Index, such as the SF-6D.

6.1 BACKGROUND

According to Ford et al (2008), a strong predictor of death in elderly people is self-rated health, especially in those who rate their health as poor. Conversely, studies have shown that the self-rated health measure (EVGGFP) is a strong predictor of survival. Participants who rate themselves in the “poor health” category often die sooner than those who rate their health as “excellent” (Ford et al 2008). It is for these reasons that assessing the impact of those who died needs to be considered.

The HALex is a generic measure of health that consists of two attributes: perceived health and activity limitation (Erickson 1998). The HALex as a health utility index was applied to a study by Trisolini et al (2002). Trisolini et al (2002) adapted the HALex and regressed it with health summary component scores to establish a value for death. The Trisolini et al approach has not been formally validated (Kazis et al 2007) and this is an opportunity to do so. The attractive feature of Trisolini et al’s approach is the simplicity of interpretation. The value given to death for PCS was 14 and MCS was 34.
The concept of the HALex is appealing but unfortunately, the ALSWH data does not contain sufficient information to generate the HALex index at each survey. The types of questions in the HALex include:

- “Are you limited in the kind or amount of work you can do because of any impairment or health problem?
- Does any impairment or health problem now keep you from doing any housework at all?
- Are you limited in the kind or amount of housework you can do because of any impairment or health problem?”

Other questions from the HALex can be found in Appendix 6.1.

Although some HALex questions are available in the ALSWH version of the SF-36v1 the possibility of using a more suitable health utility index, such as Brazier’s SF-6D mentioned earlier in Section 2.6.2.4, was worth pursuing. The ALSWH uses the earlier version of the SF-36 and so one component of this chapter will discuss and examine the conversion of the SF-36v1 to the SF-6D.

### 6.1.1 HALex Approach to account for death

Trisolini et al (2005) points out that many longitudinal studies using the SF-36 simply ignore deaths and analyse changes only for those alive at follow-up. Aware of the range of other approaches that have been undertaken, Trisolini et al (2005) still felt there were limitations to these existing approaches. Thus Trisolini and his colleagues explored the use of the HALex questionnaire to account for deaths in longitudinal data analysis by imputing PCS scores for decedents. They used a health utility index (HUI), which like other HUIs is defined on a scale ranging from 0 (death) to 1 (optimal health) (Ibrahim et al 1999). They used the data from baseline and follow-up collected by the Health Outcomes Survey (HOS). The HOS does not
provide utility assessments therefore they used questions from the HOS to estimate a utility index using the Health and Activity Limitation Index (HALex).

The HALex questionnaire has been used to evaluate Activities of Daily Living (ADL). Trisolini et al (2005) regressed PCS scores on the utility values and extrapolated the scores corresponding to a utility value of zero (death). Their analysis produced a score for death equivalent of 14 points for the PCS. Similar analysis using MCS produced a score for death of 34 points. Trisolini et al (2005) then imputed these scores for death in the Research Triangle Institute (RTI) analysis (Trisolini et al 2005) for those respondents who died before the two year follow-up. The HALex approach to death by Trisolini et al (2005) has not extended beyond a two year follow-up. Thus here is an opportunity to adapt and apply this HALex approach to the longer time frame of the Australian Longitudinal Study on Women’s Health.

6.1.2 Estimation of Health Utility Index scores

Erickson (1998) developed the matrix shown in Figure 6.1 that generated a HALex Utility score by activity limitation (Not Limited, Limited-other, Limited - Major, Unable – Major, Limited in IADLs and Limited in ADLs) and perceived health (EVGGFP). As shown in Figure 6.1, when activity limitation and perceived health are combined there are 30 health states. The table provides the calculated HALex Utility Index score for each health state. For example, perceived health is excellent and activity not limited would be scored as 1.00, while limited in ADLs and poor perceived health scores 0.10.

The Erickson matrix was adapted by Trisolini et al (2005) due to the limitations of the questions asked by the Health Outcomes Survey (HOS). The Trisolini et al (2002) matrix presented in Figure 6.2 was smaller in dimensions than the original matrix. The reason being, that the HOS did not have items related to Instrumental
Activities of Daily Living (IADLs) only for Activities of Daily Living (ADLs). Thus, Trisolini et al (2002) used HOS items to generate a 3 x 5 contingency table that classified three categories of activity limitations by five categories of perceived health. The activity limitation categories were:

\begin{itemize}
  \item[i)] No activity limitation - no difficulty in performing any of the six ADLs.
  \item[ii)] Some activity limitation - difficult with at least one ADL.
  \item[iii)] Unable to do at least one ADL (Trisolini et al 2002).
\end{itemize}

Figures 6.1 and 6.2 are presented to show comparisons of the Erikson HALex matrix and the modified HALex matrix. This modified HALex was presented at an international conference by Smith et al (2005) to demonstrate the Trisolini et al approach to calculating the utilities scores needed to determine their predicted PCS death score. The Trisolini et al (2005) matrix has 15 health states instead of the original 30 health states as shown in Figure 6.1. The values in Figure 6.2 range from 1.00 (Excellent health / No activity limitation) to 0.10 (Poor health / unable to do at least one ADL).
Table 2. Percent of population in and value assigned to health states defined by activity limitation and perceived health for persons 18 years of age or older, NHIS 1990

<table>
<thead>
<tr>
<th>Total</th>
<th>Excellent</th>
<th>Very good</th>
<th>Good</th>
<th>Fair</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>100.00</td>
<td>34.80</td>
<td>28.80</td>
<td>24.50</td>
<td>8.50</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>-</td>
<td>1.00</td>
<td>0.85</td>
<td>0.70</td>
<td>0.30</td>
</tr>
<tr>
<td>Not limited</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>83.10</td>
<td>33.40</td>
<td>26.10</td>
<td>19.30</td>
<td>4.00</td>
</tr>
<tr>
<td>Value</td>
<td>-</td>
<td>1.00</td>
<td>0.92</td>
<td>0.83</td>
<td>0.63</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>1.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Limited – other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>6.60</td>
<td>0.70</td>
<td>1.40</td>
<td>2.20</td>
<td>1.60</td>
</tr>
<tr>
<td>Value</td>
<td>-</td>
<td>0.87</td>
<td>0.70</td>
<td>0.72</td>
<td>0.52</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>0.75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Limited – major</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>3.60</td>
<td>0.40</td>
<td>0.70</td>
<td>1.30</td>
<td>0.90</td>
</tr>
<tr>
<td>Value</td>
<td>-</td>
<td>0.81</td>
<td>0.74</td>
<td>0.67</td>
<td>0.48</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>0.65</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Unable – Major</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>2.50</td>
<td>0.10</td>
<td>0.20</td>
<td>0.70</td>
<td>0.80</td>
</tr>
<tr>
<td>Value</td>
<td>-</td>
<td>0.68</td>
<td>0.62</td>
<td>0.55</td>
<td>0.38</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>0.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Limited in IADLs*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>2.80</td>
<td>0.10</td>
<td>0.30</td>
<td>0.70</td>
<td>0.90</td>
</tr>
<tr>
<td>Value</td>
<td>-</td>
<td>0.57</td>
<td>0.51</td>
<td>0.45</td>
<td>0.29</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>0.20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Limited in ADLs**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent</td>
<td>1.40</td>
<td>0.00</td>
<td>0.10</td>
<td>0.30</td>
<td>0.40</td>
</tr>
<tr>
<td>Value</td>
<td>-</td>
<td>0.47</td>
<td>0.41</td>
<td>0.36</td>
<td>0.23</td>
</tr>
<tr>
<td>Single attribute score</td>
<td>0.00</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

* IADL: Instrumental Activities of Daily Living
** ADL: Activities of Daily Living


Figure 6.1 Erickson (1998) original matrix for scoring the HALex


Figure 6.2 Trisolini et al matrix for scoring the HALex Utility Scores by Activity Limitation and Perceived Health Utility Scores
The graph in Figure 6.3 represents two relationships: one between HALex utility index scores and PCS values and the other for HALex utility index scores and MCS values. The predicted relationships (linear and curvilinear) are extended beyond 0.1 to zero on the HALex (X-axis) scale in order to extrapolate values for PCS and MCS on the Y-axis. According to Smith’s presentation the PCS death score is shown as 15, which is different to the Trisolini et al (2005) paper. The difference was due to rounding according to personal email correspondence with Dr Kevin Smith in January 2008.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS Score</td>
<td>38.4</td>
<td>12.4</td>
</tr>
<tr>
<td>MCS Score</td>
<td>51.0</td>
<td>10.8</td>
</tr>
<tr>
<td>HALex Utility</td>
<td>0.70</td>
<td>0.21</td>
</tr>
</tbody>
</table>

3. Results

- A total of 6,138 beneficiaries (63.8% response rate) completed all of the required survey items.
- Utility values were curvilinearly associated with component scores (Figure 1).
- Utilities were strongly correlated with PCS ($r=0.75$) and MCS ($r=0.45$) scores.
- Scores equivalent to death:
  - PCS = 15
  - MCS = 34


Figure 6.3 Regression results from Smith, et al (2005) conference presentation
6.1.3 Pros and Cons of the HALex Approach

Kazis et al (2007), point out that the RTI study (Trisolini et al 2002; Trisolini et al 2005) is the first to “use the utility based HALex approach to represent functional health status among dead”. In their white paper they discussed the pros and cons of the HALex approach as a recently developed health utility index and an application of the HALex by the RTI study. They suggest that the HALex as a HUI has gained recognition amongst health services researchers where it has been applied in several studies (Bradley et al 2000; Yabroff et al 2004) “however in relation to accounting for deaths, the technique of using the HALex is still new and there was limited evidence of validity” (Kazis et al 2007). Kazis et al (2007) recommended the HALex approach as a useful attempt to dealing with deaths, however they raised concerns that a 5-point change in scores may be quite different in meaning to the usual interpretation once deaths have been imputed with a score, especially in regards to the MCS death score of 34. According to Kazis et al (2007), an “assumption of such magnitude of death in the construct of mental health status is debatable”. Kazis et al (2007) raise a valid point about a 5-point change not meaning the same when deaths are included. However, a sensitivity analysis may be useful to compare results with and without and imputed value for deaths. It would appear that the actual values for death (14 and 34) may be higher than would possibly expect and this will be discussed further, later in this chapter.

Overall, the HALex approach does appear to be a very useful and simple way to acknowledge the impact of deaths in a longitudinal study of elderly participants. The idea of regressing PCS values against a HUI to obtain a PCS value for death is appealing. However, using the HALex and obtaining information about ADLs may not be possible to obtain from earlier HRQOL tools such as the SF-36v1. Therefore the concept of the HALex may be useful in another HUI setting.
6.1.4 Usage of Other Health Utility Indexes

Unfortunately, the SF-36 does not allow for a health utility index. Some researchers suggest that this limits the SF-36 instrument (Brazier et al 1998). Trisolini et al (2005) encountered this problem with the SF-36 and have used the HALex as a HUI. Such an approach to generating a HUI is adequate in the case of the HOS where additional questions in relation to ADLs are asked, however there is the problem of what to do if the additional questions are not asked. Smith et al (2005) suggested in his conference presentation that future research should assess the effect of using other measures of health utility. The focus of this chapter will be to follow up on this suggestion.

6.1.5 Using another Health Utility Index- Converting SF-36 to SF-6D

The SF-6D is a utility or preference based measure of health and is known as a health state utility. This type of measure puts the observable states of health on a preference continuum with a value of 0 represents death through to 1.0 for optimal health (Walters & Brazier 2005).

Brazier et al (1998) presented a study to estimate a preference-based single index from one of the larger generic profile measures of HR-QOL (SF-36). They suggested that the research presented a valid methodological key to converting the SF-36 to the SF-6D. The results can be applied to any SF-36 data set and hence considerably expand the available evidence base for conducting economic evaluation of health care interventions (Brazier et al 1998).

All responders to the original SF-36 questionnaire can be assigned an SF-6D score, provided the 11 items used in the six dimensions of the SF-6D have been
completed. The SF-6D preference-based measure can be regarded as a continuous outcome scored on a scale of 0.301–1.00, with 1.00 indicating ‘full health’. Attempts by researchers have been made to improve the assignment of SF-6D scores by targeting missing items (discussed briefly in Section 3.1.4.6). More details about the SF-6D can be found in Chapter 2 and at the following website: (http://www.qualitymetric.com/default.aspx).

6.1.6 Further Developments of the SF-6D health utility index

In Chapter 2, the SF-6D was mentioned as a very useful health utility index. Brazier et al (1998) developed the SF-6D and since then this health utility has been gaining popularity, especially in recent years (Rasanen et al 2006; Dann et al 2008). Fryback (2003) described the SF-6D as one of the more useful health utility indexes and argued it to be “state of the art”. Fryback (2003) liked the SF-6D because it appeared to extend and/or improve on what he and other researchers had attempted. That is, “to collapse the SF-36 into a single summary score equivalent to the preference-based health indexes” (Fryback 2003). The SF-6D can be used as the original measurement instrument, or it can be calculated from the SF-12 or SF-36. It is this useful conversion from the SF-36v1 that is of interest here, because to generate a PCS value for death one initially requires a health utility index score (Trisolini et al 2005).

The SF-6D health utility score is estimated using a set of parametric preference weights obtained from a sample of the general population using the recognised valuation technique of standard gamble 2 (Brazier et al 2007). See Section 2.6.1 for a brief description of the standard gamble technique. In addition a new excel programme is now available from the University of Sheffield to convert SF-36 data into the SF-6D utility score estimated using a set of nonparametric Bayesian preference weights (Kharroubi et al 2007). These nonparametric preference
weights are an improvement on the parametric preference weights as the nonparametric model has advantages over the conventional parametric random effects model that is reflected in improvements in the predictive ability of the model (Brazier et al 2007). Two potential advantages of the nonparametric model as suggested by Kharroubi et al (2007) are: i) very flexible method for exploring the impact of covariates; ii) a Bayesian approach offers the use informative priors to help guide and inform studies in other countries. However, Kharroubi et al (2007) suggest there can be a downside to the Bayesian approach. The “computational complexity, and specialist software that is needed”; and it is more difficult to summarise in a table of coefficients like Brazier et al (2004), due to the much greater number of parameters generated (Kharroubi et al 2007). The Health Economics and Decision Science website recommends that researchers use the algorithm in future work, but could still use the original algorithm for comparability purposes.

6.1.7 Death as a changing Quality of Life value

A PCS value for death, derived from the ALSWH data, will be compared with the Trisolini et al (2005) value of 14. However, a question that was considered was should the death score remain constant in a longitudinal study? If the HR-QOL of the surviving cohort can be shown to decline over time should those who have died also be considered as declining? If the intention is to consider accounting for deaths, is it appropriate for a death score to remain constant over time? It may be possible that a decedent’s PCS value could be higher than those still alive but in poor physical health. This is not to say that a low PCS is a strong predictor of death but more the point that if a death score is imputed there is the risk of inflating the mean PCS, especially in the “sicker” (say existing diabetes) group.
It is proposed then that the death score, like the PCS, may possibly decline in value over time. That is, the way women respond to these Surveys is dependent on their age and culture etc., therefore the estimated value for death may be considered a function of the study population. In this chapter the possibility of considering death as a health state will be explored and different values for death at the different time points will be extrapolated. The extrapolation will reflect the overall physical health of the surviving cohort at that point in time. Using the ALSWH data, if a participant died before survey 2 then they may receive a death PCS score of (say the Trisolini et al value) 14 and then at Survey 3 based on the surviving cohort the value for death may decline to 12 and so on over time. Thus participants who died before Survey 2 would receive a higher death value than those who died between Survey 2 and 3.

Furthermore, in an email discussion with Dr Kevin Smith (February 2010) about using a uniformly assigned PCS value for death over time, he raised the issue of trying to explain, conceptually, any differences in the death values. These questions will be considered and discussed later in the chapter.
6.2 AIMS AND OBJECTIVES

Although Trisolini et al (2002) determined a PCS score of 14 for death this was generated for a specific sample. Therefore, it is necessary to determine whether this value is valid in other populations. Furthermore, could another health utility index be used to generate a similar value for death, hence providing another approach when the HALex (ADLs) is not available?

In this chapter the SF-36v1 will be converted to the SF-6D using programming code provided by Brazier and Hanmer at the website:

http://www.openhealthmeasures.org/repository/

Then the new SF-6D indexes will be used to determine a PCS value for death using the ALSWH dataset.

The aims of this chapter are to:

Using the ASLWH data:

1. Convert the SF-36v1 to SF-6D at surveys 1 to 4;
2. Conduct regression analyses involving Linear and Polynomial regression to determine an appropriate PCS value(s) for death at each survey period;
3. Compare the Trisolini et al (2005) PCS death value of 14 to the ALSWH PCS death value;
4. Use longitudinal multiple imputation in WinBUGS to impute for missing PCS values for surveys 1 to 4;
5. Assess the effects of the adapted SF-6D method to account for deaths in a longitudinal study using random effects models.
6.3 METHODS

6.3.1 Subjects

The focus of this chapter is the older cohort of women for whom Survey 1 was conducted in 1996 through to Survey 4 in 2005. Deaths and retention rates for each survey have been reported earlier in Section 5.5.

6.3.2 Measures

6.3.2.1 Converting from SF-36v1 to SF-6D

The following section will explain the conversion of the SF-36v1 to the SF-6D using the appropriate programming code. The Stata 8.2 and SAS 8.2 programming code has been provided by Janel Hanmer on her website:

http://www.openhealthmeasures.org/repository/.

There are two programming code versions since the SF-36 has two versions. In the earlier SF-36 version a number of the questionnaire items used a Likert scale with values 1-6, these items were altered in version 2 to a Likert scale with values 1-5. Since the ASLWH has used the SF-36v1 of the items with the Likert scale with 1-6 a conversion to 1-5 scale was required. How this was done is explained in Section 6.3.2.4.

Figure 6.4 shows the items from the SF-36 questionnaire used to generate the 8 subscales and 2 health component summary scores. The items used in the SF-6D are directly related to 11 items in the SF-36. Details of the items used by Brazier et al (2004) for the SF-6D have been shown in the Appendix 6.2. In short, (from the SAS and Stata code of Hanmer) they are considered as the following items in Section 6.3.2.2. In addition to the SF-36v1, the ALSWH made localised changes to the SF-36 items. These localised changes are explained in Section 6.3.2.3.
6.3.2.2 The Coding of SF-36 items

This code presumes the fields are numbered and coded as follows:

- SF3 = vigorous activities where 1="limited a lot" to 3="not limited at all"
- SF4 = moderate activities where 1="limited a lot" to 3="not limited at all"
- SF12 = bathing and dressing where 1="limited a lot" to 3="not limited at all"
- SF15 = physical limited work where 1="yes" and 2="no"
- SF18 = emotional accomplish less where 1="yes" and 2="no"
- SF21 = bodily pain where 1="none" to 6="very severe"
SF22 = pain interferes with work where 1="not at all" to 5="extremely"
SF24 = nervous where 1="all the time" to 6="none of the time"
SF27 = energy where 1="all the time" to 6="none of the time"
SF28 = downhearted and blue where 1="all the time" to 6="none of the time"
SF32 = social activities where 1="all the time" to 5="none of the time"

Initially the SF-36 items were coded by Janel Hanmer in July 2005. In 2007, Hanmer, in discussion with Brazier and Beuchner, incorporated some more code. Noted here also is that the updated SAS and Stata programming code now use the consistent model from Brazier & Roberts (2004) (Column 2, Table 4) (see Appendix 6.2). In 2009, Hamner published a method for “predicting an SF-6D preference-based score using MCS and PCS Scores from the SF-12 or SF-36" (Hanmer 2009). In February 2009, through a personal email correspondence with Hanmer, an error with the Stata code was brought to her attention.
6.3.2.3 Localised changes to SF-36v1 items for ALSWH old survey

In order to generate the SF-6D index score the following variables are generated from the 11 items, that represent the 6 domains, hence, the title Short Form 6D.

The SF-6D domain variables are:

- (SFPhysical) - consists of 3 items (SF3, SF4, SF12)
- (SFRole) - consists of 2 items (SF15 and SF18)
- (SFSocial) - consists of 1 item (SF32)
- (SFPain) - consists of 2 items (SF21 and SF22)
- (SFMental) - consists of 2 items (SF24 and SF28)
- (SFVital) - consists of 1 item (SF27)

6.3.2.4 Generation of a random number to convert items

Stage 1

Since the items SF24, SF27 and SF28 have a Likert scale of 1-6 they needed to be reduced to scale of 1-5. The classification of 3 in the old Likert scale (1-6) needed to be reassigned because it did not exist on the newer 1-5 scale. In order to "preserve" the data’s integrity the collapsing of old category 3 to either new category 2 or 3 needed to be randomly assigned. The following coding was carried out was on SF24:

If the SF24 value =3 and corresponding random number was < 0.5 then the value of the new item became 2. If the SF24 value = 3 and the random number was ≥ 0.5 then the item was replaced with 3. A direct map from a value of 1 was made, while the following replacements were made for other values in the scale:

- if SF24=2 then SF24r=2
- if SF24=4 then SF24r=3
- if SF24=5 then SF24r=4
- if SF24=6 then SF24r=5
Note: the same procedure occurred for SF27 and SF28 (See Appendix 6.2). Also, note a seed is set so that same sequence can be repeated if needed.

Stage 2

The next stage in the conversion was to create a variable that assigned a value of 1 if the dimension is in "worse state". A zero is assigned if not the case.

\[
\text{if SFPhysical} \geq 4 \text{ or SFRole} \geq 3 \text{ or SFSocial} \geq 4 \text{ or SFPain} \geq 5 \text{ or SFMental} \geq 4 \text{ SFVital} \geq 4 \text{ then most}=1 \text{ , else most}=0
\]

Stage 3

The next stage was to assign a weighted value from Table 4 of Brazier et al (2004) paper (see Appendix 6.2) for each level of the dimension. For example, for the physical functioning domain:

\[
\begin{align*}
\text{if SFPhysical} &= 1 \text{ then pf1}=0 \\
\text{if SFPhysical} &= 2 \text{ then pf1}=-0.035 \\
\text{if SFPhysical} &= 3 \text{ then pf1}=-0.035 \\
\text{if SFPhysical} &= 4 \text{ then pf1}=-0.044 \\
\text{if SFPhysical} &= 5 \text{ then pf1}=-0.056 \\
\text{if SFPhysical} &= 6 \text{ then pf1}=-0.117 \\
\text{else SFPhysical} &= .
\end{align*}
\]

The variable “most” was also re-coded as 0.61, if originally the value was 1.

Stage 4

This new variable (most1) was then added to the other 6 dimension variables to obtain the SF-6D index.

\[
\text{SF-6D} = 1 + pf1 + rl1 + sc1 + pn1 + mh1 + v1 + most1
\]  
(equation 6.1)
The SF-6D indexes were generated in the same manner for surveys 1-4. The Stata code used for Survey 1 is shown in Appendix 6.2.

**Need to set seed**

Note that since the random numbers are assigning values to questions SF24, SF27 and SF28 then the total values for SF-6D can vary. That is, if the programming code is run multiple times the resulting SF-6D values will vary each time, hence the requirement to set a seed (starting point), for the random number generator.

### 6.3.3 Development of the Outcome Variable PCSD

In Chapters 4 and 5, a distinction was made in regards to PCS and the ALSWH PCS since there were comparisons to the Diehr et al (2003) transformation equation (PCTD). However in this chapter, when the acronym PCS is used it is in reference to the ALSWH PCS. The outcome variable in this chapter will be PCS at surveys 1, 2, 3 and 4. A new variable (and acronym) was generated, called **PCSD** which represents the PCS variable with the value for death included at surveys 2, 3 and 4.

#### 6.3.3.1 Determination of PCS value for Death

Trisolini et al (2005) regressed PCS scores on the utility values and extrapolated the scores corresponding to a utility value of zero (death). This approach was adapted and applied to the ALSWH data, however the health utility index (HUI) used was the SF-6D. Note that the focus of this research is only the PCS score, however, a similar approach could be considered for the MCS, but will not be attempted in this thesis.
6.3.3.2 Simple Regression models

Trisolini et al (2005) regressed PCS scores on the HALex utilities and found logarithmic models appeared to fit the data better than simple linear regression models (Trisolini et al 2005). Figure 6.3 displays these curvilinear relationships. As such, logarithm, quadratic and cubic equations will be considered to extrapolate a PCS value using the ALSWH dataset.

The dependent variable used in the simple linear regression will be PCS at survey 1 while the independent variable was the SF-6D index at survey 1. For the two polynomial regressions, the independent variables will include additional squared and a cubic terms based on the SF-6D at survey 1. The logarithm regression involves the natural log transformation of the dependent variable (PCS at survey 1). Further simple linear, logarithm and polynomial regressions were also used for survey 2, survey 3 and survey 4. The linear and polynomial regression equations at survey 1 are of the following general form:

Linear Equation 6.1  
\[ \hat{Y}_t = \alpha + \beta_1 \times SF\_6D \]

Quadratic Equation 6.2  
\[ \hat{Y}_t = \alpha + \beta_1 \times SF\_6D + \beta_2 \times (SF\_6D - C_2)^2 \]

Cubic Equation 6.3  
\[ \hat{Y}_t = \alpha + \beta_1 \times SF\_6D + \beta_2 \times (SF\_6D - C_3)^2 - \beta_3 \times (SF\_6D - C_3)^3 \]

Log Equation 6.4  
\[ \log \hat{Y}_t = \alpha + \beta_1 \times SF\_6D \]

Where, t = 1 to 4 , Y = PCS, \( \alpha \) = Y-intercept, \( \beta \) = slope , \( C_2 \) and \( C_3 \) = constants

Later a final decision will be made as to which equation should be used to extrapolate a value for death. A decision will also be made as to whether or not a changing value for death over time should be used.
6.3.4 Covariates

Once again a case study of diabetes will be used to apply the newly developed outcome PCSD. The random effects longitudinal analyses includes the same variables as in the models presented in Chapter 5: area of residence (0=non-urban, 1=urban), age cohort (0=70-72 years, 1=73-75 years), smoking status (0=no and 1=yes), diabetes status (1=No diabetes, 2=New cases and 3=Existing cases). A integer variable to represent the number of co-morbidities was also generated. This was retrospectively generated as the cumulative total number of co-morbidities reported by participants from Survey 1 through to Survey 4. The co-morbidities score ranges from 0-5, according to how many of five medical conditions (heart disease, stroke, osteoporosis, asthma and hypertension) had been reported at each survey period.

6.4 Statistical Analysis

6.4.1 Developments of SF-6D and PCS death value

Stata10.1 was used to generate the SF-6D indexes for each participant at each survey. The linear and polynomial regression equations and graphs using SF-6D and PCS (at each survey) were generated using JMP8.0.

6.4.2 Cross-sectional Analysis

To determine whether the inclusion of values for death (PCSD) had any substantive effect on findings, cross-sectional comparisons were made between the Existing cases, New cases and No diabetes groups using oneway ANOVA with multiple comparison t-tests (Bonferroni tests) at each survey. Paired t-tests were used to test for differences between surveys for survivors in each diabetes group. Statistical
analyses and graphs were conducted and created using SAS 9.1, Stata10.1 and JMP8.0.

6.4.3 Longitudinal Data Analysis

In order to conduct similar analyses to Chapter 5, a PCS value for death was imputed for those participants who had died after survey 1. Random effects models were conducted to examine the effects of using the adapted SF-6D method to account for deaths in a longitudinal study using the SF-36v1. Consideration was given to the “missingness” of the data and so MAR and MNAR random effects mixed models were compared. Other comparisons such as complete case for PCS and PCSD were also considered for the diabetes case study. For the outcome PCSD, a model was considered when a PCS value was added for death and carried forward for that participant, thus giving them complete outcome data (PCSD). The longitudinal models used, including an imputation component, are shown below in section 6.4.4.1. These are all random effects models that employed random slopes and intercepts, similar to the modeling in Chapter 5. The statistical modelling was conducted using WinBUGS1.4.1 (2003).

6.4.4.1 Using WinBUGs

To run an analysis in WinBUGS1.4.1a similar approach was taken to that discussed in Chapter 5. Stata10.1 (2008) was used to generate the random effects Restricted Maximum Likelihood (REML) regression models, shown in Tables 6.9 and 6.10.
6.4.4.2 Longitudinal Model using a Fully Bayesian Approach to imputing missing (assumption of MNAR)

The following MNAR longitudinal model was created in Chapter 5, the main difference is the outcome variable (PCSD). The MAR model was similar to the model used in Chapter 5. The MNAR model like the MAR model draws upon the data which imputes for missing data. The total of participants (i) and how many time points (t) needs to be established before the model of interest, can be dealt with. For loops in the code: for (i in 1:8368), are used to draw upon the earlier loaded dataset. In this case there are t=4 time points for i=8368 participants. All participants have 4 time points included in the dataset because a missing outcome value is recognized as “NA”. These missing values will be replaced with an imputed value as the modelling is conducted.

Model {
  for (i in 1:8368)
    for (t in 1:4) {

In this Chapter, let $Y_{it}$ be a continuous variable (PCSD) denoting the Physical Component Summary score with deaths included.

Model of Interest (and non-informative priors):

$Y_{it} = \beta_0 + \beta_1 Time_i + \beta_2 Age_i + \beta_3 Area_i + \beta_4 Comorb_i + \beta_5 Smoke_i + \beta_6 Diab_{1i} + \beta_7 Diab_{2i} + \epsilon_i$

where

$\beta_6 Diab_{1i}$ = the effect size difference between New cases and Existing cases groups
$\beta_7 Diab_{2i}$ = the effect size difference between No diabetes and Existing cases groups

$Y_{it} \sim N(\mu_{it},\tau_{it}^{-1})$

$\mu_{it} \sim N(0, 10000)$

$\tau_{it} \sim \text{Gamma}(0.01, 0.01)$
random subject intercept - $\beta_{0i} \sim N(\mu_{0t}, \tau_{0t}^{-1})$

$\mu_{0t} \sim N(0, 10000)$
$\tau_{0t} \sim \text{Gamma}(0.01, 0.01)$

random time slope - $\beta_{1t} \sim N((\mu_{1t}, \tau_{1t}^{-1})$)

$\mu_{1t} \sim N(0, 10000)$
$\tau_{1t} \sim \text{Gamma}(0.01, 0.01)$

WinBUGS code used:

```
i.reg[i] <- b.int[i] + b.age*age[i] + (b.diab[1]*equals(diab[i], 2))
    + (b.diab[2]*equals(diab[i], 1)) + b.com*comorb[i]
    + b.smok*smoke[i] + b.area*area[i];

b.int[i] ~ dnorm(b0.mu, b0.tau);
b.slo[i] ~ dnorm(b1.mu, b1.tau);
```

The model shown above, feeds into the following model which involves the random slopes.

```
pcsd.mu[i, t] <- i.reg[i] + b.slo[i]*(t - 2.5);
```

The following represent the informative priors for parameters in model of interest. They assume a normal distribution with a mean of zero and a variance of 0.0001.

$\beta_{2i} \sim N(0.0, 0.0001), \beta_{3i} \sim N(0.0, 0.0001), \beta_{4i} \sim N(0.0, 0.0001), \beta_{5i} \sim N(0.0, 0.0001)$
$\beta_{6i} \sim N(0.0, 0.0001), \beta_{7i} \sim N(0.0, 0.0001)$

The following represent the informative priors for logistic regression parameters in the missing data model. They assume a normal distribution with a mean of zero and a variance of 0.0001.

$\alpha_{3i} \sim N(0.0, 0.0001), \alpha_{4i} \sim N(0.0, 0.0001), \alpha_{5i} \sim N(0.0, 0.0001), \alpha_{6i} \sim N(0.0, 0.0001)$
$\alpha_{7i} \sim N(0.0, 0.0001), \alpha_{8i} \sim N(0.0, 0.0001), \alpha_{9i} \sim N(0.0, 0.0001)$
The following represent the informative priors for random intercept that assume a normal distribution with a mean of zero and a variance of 0.0001.

\[ \beta_{0i} \sim N(0.0, 0.0001), \quad \beta_{0i} \sim \text{gamma}(0.01, 0.01) \]

The following represent the informative priors for random slope that assume a normal distribution with a mean of zero and a variance of 0.0001.

\[ \beta_{1i} \sim N(0.0, 0.0001), \quad \beta_{1i} \sim \text{gamma}(0.01, 0.01) \]

The following represent the informative priors for logistic regression constant that assumes a normal distribution with a mean of zero and a variance of 0.0001.

\[ \alpha_{0t} \sim N(0.0, 0.0001) \]

### 6.4.4.3 The logistic regression model and the \( \gamma \) coefficient

\[
\logit(\text{obs}_{it}) = \alpha_{0t} + \gamma_{\text{PCSD}_{it}} + \alpha_1 \text{Age}_{it} + \alpha_2 \text{Area}_{it} + \alpha_3 \text{Comorb}_{it} + \alpha_4 \text{Smoke}_{it} + \alpha_5 \text{Diab}_{1t} + \alpha_6 \text{Diab}_{2t} + \epsilon_{it}
\]

Constants in logistic regression model -

\[ \alpha_{0t} \sim N(\mu_{0t}, \tau_{0t}) \]

\[ \mu_{0t} \sim N(0, 0.0001) \]

\[ \tau_{0t} \sim \text{Gamma}(0.01, 0.01) \]

In order to determine whether the missing outcome PCSD is missing not at random a binary variable was generated called observed. A Bernoulli distribution was used for this binary value. The WinBUGS code used was:

```plaintext
obs[i,t] ~ dbern(p.obs[i,t]);

logit(p.obs[i,t])<- a.cons[t]+0.0001*cut.pcsd[i,t]+i.missing[i];

i.missing[i]<-(a.diab[1]*equals(diab[i],2))
+ (a.diab[2]*equals(diab[i],1)
+ a.com*comorb[i] + a.smok*smoke[i] +
 a.age*age[i] + a.area*area[i]);
```

`i.missing[i]` is used to bring together the participants covariates and feeds them into the longitudinal logistic regression at each time point \( t \). Since they are all
baseline covariates and not missing the main part of this logistic model is the addition of the outcome which does have missing data. The $\gamma$-coefficient is required and is mentioned in Section 5.4.7.

$a$.diab[1], a.age, a.area represent the parameter estimates that have are calculated at each iteration and were initially set at the beginning.

$\text{diab}[i]$, $\text{smoke}[i]$, $\text{comorb}[i]$ represent the observed data that the model draws upon. Like in the case of the model of interest the parameter estimates for the longitudinal logistic regression were established using Maximum Likelihood estimates from a logistic regression model and the observed data.

For the MNAR model the $\gamma$ coefficient = 0.001 was used. This was same value for the cut function for $\gamma$ used in section 5.4.7.

6.4.4.4 Multiple Imputation for missing covariates using Stata

In order to determine the impact of the missing baseline covariate data, an approach that assumes MAR was also applied using Stata. A technique in Stata known as ice (Imputation using Chained Equations) can deal with all missing data (continuous and categorical variables), however, the assumption of MNAR is not applicable. In order to determine the impact of missing covariates and outcomes the ice approach was used as well as the WinBUGS approach. The resulting five complete data sets from the multiple imputation, are combined using a Stata command called mim and then analysed using random effects models with random slopes and intercepts.
6.4.4.5 The Stata Code used for Multiple Imputation

Royston and colleagues have developed programming code for Stata that can be downloaded from the Stata website and installed in Stata. The structure for the ice command is generally quite simple to construct. In this attempt at using the ice command in Stata10.1, baseline binary covariates were considered along the PCSD variable and the co-morbidities (categorical) variable as previously mentioned in the WinBUGS methods.

In this Chapter the ice command has been used as a naïve approach to missing data and is not the focus at this stage. Later in Chapter 7, this ice command developed by Royston (2005) will be given greater importance when approaches to missing outcomes and time dependent covariates are accounted for. Here in Chapter 6 the ice command was considered very simply to impute missing baseline covariates and missing outcomes, as a way of making comparisons, with the WinBUGS method models. Although no statistical comparisons of the models were possible.

Firstly, in constructing the necessary command structure, all the variables that are to be considered in the M=5 datasets are listed. Using the cmd command all variables that will be imputed can be assigned a regression process linear regression (regress) for a continuous variable or logistic regression (logit) for a binary variable or ordered logistic regression (ologit) for ordinal categorical. If a variable which is to be imputed is not assigned a regression process, by default Stata will try and choose an appropriate generalised linear regression model. Using regressions such as mlogit or ologit in this ice procedure can be problematic with Stata 10.1, and thus setting these at the cmd is important to avoid an error.
Next step is to use the eq command. This eq command allows a subset of the variables to be chosen to use as covariates in the regressions. If not used selected, Stata will by default select all variables as covariates for each regression. If variables are included and do not have missing data then the original variable remains unchanged in the m=5 datasets. The original dataset remains in the newly structured multi-levelled dataset. New variables are automatically generated, one to identify individual participants (_mi) and one to identify the new dataset (_mj). The multiple datasets including the original (_mj=0) are then saved automatically using saving and m(5) instructs Stata to conduct the process “m” times - in this case 5 times. Note that diabetes status (diab_retro: (1) Existing cases, (2) New cases and (3) No diabetes) was considered to have no missing data since unidentified participants were removed before multiple imputation.

6.4.4.6 Combining MI datasets to be used in a Longitudinal Analysis

As mentioned, the Stata10.1 command known as ice produces multiple complete datasets. In order to summaries these datasets and conduct a longitudinal analysis of the “m” complete datasets Royston (2005) developed the following method mim: (manipulating multiply imputed datasets). The mim: command combines the imputed data sets and allows for longitudinal modelling with “averaged regression estimates”. The following code is an example using the ALSWH data, the variable IDalias identifies the individual participants’ intercepts:

```
ice IDalias pcsd age smoke area diab_retro comorb, ///
   cmd(pcsd: regress, age: logit, smoke: logit, ///
   area: logit, comorb: ologit) ///
   eq(pcsd: age smoke area diab_retro comorb, ///
   comorb: age smoke area diab_retro pcs, ///
   age: comorb smoke area diab_retro, ///
```
smoke: comorb age area diab_retro, ///
area: smoke comorb age area diab_retro) ///
seed(1285964) m(5) (saving imp, replace)
use imp, clear

The longitudinal model used in Stata 10.1 is shown below. The covariance matrix was considered “unstructured”. The model used the `mim` command to combine all 5 fully imputed datasets. There are actually 6 datasets stored because the original dataset remains unchanged and is labelled as dataset 0. The datasets are stored in one large dataset called `'imp'` and the original dataset is not used in the final analysis of `mim`.

The following Stata code was used for the random slope and intercept model:

```
xi: mim: xtmixed pcsd survey i.diab_retro age comorb area
smoke || IDalias: survey, cov(unstructured)
```

**Cov** = covariance -variance matrix structure

**xtmixed** = random effects command

**survey** = time points

The variable `IDalias` was used to determine the random intercept of each participant. The variable **survey** represents time (1-4) in the longitudinal model and is used to determine the random slopes.
6.5 RESULTS

6.5.1 Descriptive Statistics – SF-6D at Surveys 1-4

Table 6.1 shows that the minimum and maximum values of the SF-6D in the ALSWH are 0.301 and 1.0. The value of 0.301 is actually a floor value (that is impossible to go below this value) based on these data and the conversion from the SF-36v1. At Survey 1-4 the mean SF-6D utility scores are 0.72, 0.73, 0.71 and 0.69 respectively, suggesting a possible decline in scores over time (apart from survey 2). Table 6.1 also reported the proportion of missing values, that is those participants who were not dead but had no score calculated due to missing SF-36v1 items. The proportions of missing SF-6D (but not dead) are similar to those proportions reported earlier for APCT (PCS) in Section 5.5.1. The proportion of missing values for survey 1, 3 and 4 of 13%, 20% and 16%, respectively, are considered high. The cause of the majority of the missing data can be contributed to the fact that (like the PCS) if any items are missing then the final score cannot be calculated. Imputation of missing SF-6D utility scores or SF-36 items will not be considered in this Chapter. The sample size at baseline (N=10759) for the SF-6D is still considered quite large and hence will be used in the exploratory analysis shown later.

Table 6.1 Descriptive Statistics – SF -6D at Surveys 1, 2, 3 and 4

<table>
<thead>
<tr>
<th>SF-6D</th>
<th>n</th>
<th>Mean (SD)</th>
<th>Minimum</th>
<th>Maximum</th>
<th>% Missing SF-6D (not dead)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey 1</td>
<td>10759</td>
<td>0.722 (0.13)</td>
<td>0.301</td>
<td>1.00</td>
<td>N= 1243213.4%</td>
</tr>
<tr>
<td>Survey 2</td>
<td>9593</td>
<td>0.729 (0.13)</td>
<td>0.301</td>
<td>1.00</td>
<td>N=104348.1%</td>
</tr>
<tr>
<td>Survey 3</td>
<td>6929</td>
<td>0.710 (0.13)</td>
<td>0.301</td>
<td>1.00</td>
<td>N=864619.9%</td>
</tr>
<tr>
<td>Survey 4</td>
<td>6012</td>
<td>0.692 (0.12)</td>
<td>0.301</td>
<td>1.00</td>
<td>N=715316.0%</td>
</tr>
</tbody>
</table>
6.5.2 Correlations between PCS and SF-6D at Survey 1-4

Initially, a linear relationship between PCS and SF-6D was anticipated, so as an introductory step towards choosing a possible regression equation Pearson's correlations were conducted at each Survey. Results in Tables 6.2 suggest that there are strong correlations between PCS and SF-6D at each of the surveys. The Pearson’s correlations are all statistically significant (P-value<0.0001) at surveys 1, 2, 3 and 4 where the correlation coefficients (ρ) are 0.72, 0.72, 0.74 and 0.74 respectively. Therefore, at this point a linear relationship appears plausible with the observed data at each survey.

Table 6.2 Correlations for PCS and SF-6D at surveys 1, 2, 3 and 4

<table>
<thead>
<tr>
<th>Measure</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS Score at S1</td>
<td>10851</td>
<td>49.96</td>
<td>9.95</td>
<td>17.86</td>
<td>75.45</td>
</tr>
<tr>
<td>SF-6D score at S1</td>
<td>10759</td>
<td>0.722</td>
<td>0.128</td>
<td>0.301</td>
<td>1.0</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.7177*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCS Score at S2</td>
<td>9560</td>
<td>49.58</td>
<td>10.03</td>
<td>13.45</td>
<td>75.69</td>
</tr>
<tr>
<td>SF-6D score at S2</td>
<td>9593</td>
<td>0.729</td>
<td>0.126</td>
<td>0.301</td>
<td>1.0</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.7241*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCS Score at S3</td>
<td>7135</td>
<td>47.82</td>
<td>10.18</td>
<td>17.97</td>
<td>75.75</td>
</tr>
<tr>
<td>SF-6D score at S3</td>
<td>6929</td>
<td>0.710</td>
<td>0.125</td>
<td>0.301</td>
<td>1.0</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.7379*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCS Score at S4</td>
<td>6206</td>
<td>46.25</td>
<td>10.14</td>
<td>15.43</td>
<td>70.26</td>
</tr>
<tr>
<td>SF-6D score at S4</td>
<td>6012</td>
<td>0.692</td>
<td>0.125</td>
<td>0.301</td>
<td>1.0</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td>0.7354*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* P-value<0.0001
6.5.3 Comparing Results from R-squared ($R^2$) and Y-intercepts ($\alpha$) and beta ($\beta$) estimates

Figure 6.5 shows the R-squared ($R^2$) values for each polynomial and linear regression model at survey 1. Comparing all four regressions, the $R^2$ values are all similar. The $R^2$ results for the linear, quadratic, cubic and logarithm regressions were 0.5150, 0.5151, 0.5153 and 0.5055, respectively. These $R^2$ values differ at the fourth decimal place (for Surveys 1 to 3) and suggest, approximately 51.5% of the variation in Y (PCS) can be explained by the SF-6D. Thus it appears that for all regression equations the relationships are all behaving in a linear way as can be seen in Figures 6.5 - 6.8. So it is not surprising to see similar $R^2$ values in Figures 6.6 - 6.8. What is very obvious from Figures 6.5 - 6.8 is the divergence from the linear once the curves reach the lower regions of the data points. Tables 6.3 shows summaries of the $\beta$ estimates and Y-intercepts ($\alpha$) for each regression model at Survey 1, generated from the results of regressing the PCS and SF-6D by different functions. The logarithm model has exponentiated the Y-intercept value (19.99) so comparisons to the other Y-intercepts can be made and matches what is shown in Figure 6.5. Noticeably, in Figure 6.5, when comparing the regression “curves”, the Y intercept for the linear regression is lower than the logarithm and the cubic, while the quadratic Y-intercept is slightly lower than the linear. In Figures 6.6- 6.8 the linear has a lower Y-intercept than the cubic, logarithm and quadratic equations.
### Table 6.3: Comparisons of Linear, Logarithmic and Polynomial Regression estimates for Survey 1

<table>
<thead>
<tr>
<th>Co-variates</th>
<th>Linear Equation</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Co-variates</td>
<td>( \beta ) Estimates</td>
<td>Std Error</td>
<td>t Ratio</td>
<td>P-value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>9.15</td>
<td>0.40</td>
<td>23.14</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-6D1</td>
<td>56.01</td>
<td>0.54</td>
<td>104.11</td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Transformed log(PCS) *\**

<table>
<thead>
<tr>
<th>Co-variates</th>
<th>Transformed log(PCS) <em>*</em> (anti-log)</th>
<th>Std Error</th>
<th>t Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>19.99</td>
<td>0.01*</td>
<td>340.03*</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>SF-6D1</td>
<td>3.40</td>
<td>0.01*</td>
<td>102.30*</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

**Quadratic Equation**

<table>
<thead>
<tr>
<th>Co-variates</th>
<th>Quadratic Equation</th>
<th>B Estimates</th>
<th>Std Error</th>
<th>t Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.34</td>
<td>0.43</td>
<td>21.90</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>SF-6D1</td>
<td>55.85</td>
<td>0.56</td>
<td>100.58</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>(SF-6D1-0.731)^2</td>
<td>-4.38</td>
<td>3.70</td>
<td>-1.18</td>
<td>0.2371</td>
<td></td>
</tr>
</tbody>
</table>

**Cubic Equation**

<table>
<thead>
<tr>
<th>Co-variates</th>
<th>Cubic Equation</th>
<th>( \beta ) Estimates</th>
<th>Std Error</th>
<th>t Ratio</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>8.48</td>
<td>0.63</td>
<td>13.41</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>SF-6D1</td>
<td>57.09</td>
<td>0.88</td>
<td>65.17</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>(SF-6D1-0.724)^2</td>
<td>-8.36</td>
<td>4.29</td>
<td>-1.95</td>
<td>0.0515</td>
<td></td>
</tr>
<tr>
<td>(SF-6D1-0.724)^3</td>
<td>-36.00</td>
<td>19.64</td>
<td>-1.83</td>
<td>0.0668</td>
<td></td>
</tr>
</tbody>
</table>
*Noted the fitted curve for the transformed log(PCS) is shown on the same PCS scale (Y-axis) as the other fitted lines.

Figure 6.5 Regression results of PCS on SF-6D at survey 1 using linear, logarithmic and polynomial regression models
*Noted the fitted curve for the transformed log(PCS) is shown on the same PCS scale (Y-axis) as the other fitted lines.

Figure 6.6 Regression results of PCS on SF-6D at survey 2 using linear, logarithm and polynomial regression models
*Noted the fitted curve for the transformed log(PCS) is shown on the same PCS scale (Y-axis) as the other fitted lines.

- Transformed Fit Log: $\text{Transfomed Log } R^2 = 0.5348$
- Polynomial Fit Degree: Quadratic regression $R^2 = 0.5449$
- Polynomial Fit Degree: Cubic regression $R^2 = 0.5450$
- Linear Fit: Linear regression $R^2 = 0.5446$

Figure 6.7 Regression results of PCS on SF-6D at survey 3 using linear, logarithm and polynomial regression models
*Noted the fitted curve for the transformed log(PCS) is shown on the same PCS scale (Y-axis) as the other fitted lines.

Figure 6.8 Regression results of PCS on SF-6D at survey 4 using linear, logarithm and polynomial regression models

```
To estimate an extrapolated value for death from ALSWH, baseline results in Table 6.3 and Figure 6.5 were considered and are in line with Trisolini et al's (2002) original approach. From Figure 6.5 the data appears to produce similar fitting “lines” for cubic, quadratic logarithmic and linear regressions until the end point for SF-6D utility values are reached. The predicted values from the polynomial and logarithm equations tend to then move away from the linear, as would be expected. In Figures 6.6-6.8, similar graphs have occurred, whereas in Figure 6.8 (Survey 4) the
```
The curvature of the “fitted line” is more pronounced for the cubic equation than in the earlier cubic graphs.

Figures 6.5 – 6.8 show the fitting of the four functions: linear, logarithm, quadratic and cubic. The fit of the equations based on the observed plotted values remain very close in all functions. However, once the values of SF-6D reach the lower limit (floor value) of 0.301 the functions begin to move away from each other as the line/curves approach the Y-axis. Due to the strong correlation of the observed values a linear relationship appears reasonable and more appropriate with these data. There is obvious divergence away for the linear regression line as the fitted equations project beyond the real values of 0.301. It is important to remember the exploratory nature of this simple modeling and the main point of these fitted equations is to determine whether an extrapolation beyond 0.301 is in fact viable. Judging by Figures 6.5 – 6.8 there is quite a difference to where the functions cross the Y-axis, especially, for the logarithm and cubic equations.

Focusing on Survey 1 and results in Figure 6.5, the relationship would indicate a strong positive linear relationship, that is, a high SF-6D index relates to a high PCS value and a low SF-6D index relates to a low PCS value. Thus from all Figures 6.5-6.8, the cubic result suggests that there appears to be a turning point at approximately 0.1 for SF-6D. It would appear the upward nature of the cubic curve becomes more prominent as the sample size decreases in surveys 3 and 4 where there are less observed values. The fact that this turning point occurs beyond observed values for the SF-6D index suggest it would be difficult to justify the choice of the cubic model. A small cluster of points appears to be impacting on the shape of the polynomial functions.
In Figure 6.5 and 6.6 the linear and quadratic relationships appear very similar, such that the plotted points (observed data) of both equations appear linear. Due to the nature of the quadratic equation the points eventually begin to move away from the linear regression and possibly due to the small cluster of points highlighted in Figure 6.7. The curvilinear shape becomes more obvious in survey 3 and 4 (Figures 6.6 and 6.7) again when the sample sizes are reduced due to deaths and or missing PCS and SF-6D scores.

Trisolini et al (2002) chose to use the logarithmic regression which they felt best explained the data. They also point out that reliability adjustments to correct slopes for measurement error were not made. The reason being is that the HALex utility values are deterministic. By this they mean that utility values were calculated from an algorithm and not reported directly by respondents (Trisolini et al 2002). Thus the variability amongst respondents is not so problematic.

Finally, since the distance between zero and the lower end of the SF-6D index (0.301) is wider than for the HALex index (0.1), the use of the linear regression here may be a more conservative approach to extrapolating a PCS value for death. It is also worth noting that all the curves are approximately linear in the observed data range.
6.5.4 Calculation of the PCS Death scores using a Linear Equation at each survey

6.5.4.1 Linear Regression Equations

Earlier in this chapter, it was shown that the variables PCS and SF-6D are highly correlated so using a simple linear regression model where $Y = PCS$ and $X = SF-6D$, the following linear regression models were created at surveys 1, 2, 3 and 4. To extrapolate a value for PCS, the simple regression lines could be extended as generated in Figures 6.5-6.8 or similarly the regression equations could be used and zero inserted for the SF-6D utility score ($X$ value) to find a value for PCS ($Y$ value) at each survey.

Note $Y_t = PCS_t$ and $X_t = SF-6D_t$ where $t = 1, 2, 3, 4$.

\[
\hat{Y}_{pcs1} = \alpha + \beta \times SF_6D = 9.15 + 56.01 \times 0 = 9.15
\]
\[
\hat{Y}_{pcs2} = \alpha + \beta \times SF_6D = 7.34 + 57.67 \times 0 = 7.34
\]
\[
\hat{Y}_{pcs3} = \alpha + \beta \times SF_6D = 5.37 + 59.41 \times 0 = 5.37
\]
\[
\hat{Y}_{pcs4} = \alpha + \beta \times SF_6D = 5.09 + 58.88 \times 0 = 5.09
\]

Inserting $X=0$ for each equation results in values of 9.2, 7.3, 5.4 and 5.1 (rounded to 1 decimal place) for $\hat{Y}$. Later it will be discussed whether the PCS death value should vary over time or should a constant value for death be considered at each survey.

6.5.5 Considerations of the PCS value for death

6.5.5.1 Calculation of Floor and Ceiling PCS values using ALSWH data

An obvious question arises in all this and gets back to the simple point of why not just simply impute a PCS value of zero for death? Firstly, when considering the ceiling and floor effects of the PCS it can be shown through the ALSWH data that it is not possible to score a value of 0 and for that matter 100. This is due to the weighting applied to the z-scores used to calculate the PCS. Also, it needs to be remembered that zero on the subscales (PF, RP, BP, GH, VT, SF, RE and MH) and
for the PCS does not equal death. The ALSWH researchers used the following codes to calculate the raw z-scores for each subscale:

\[
\begin{align*}
PF_{ZB} &= (PF - 63.35)/25.94 \\
RP_{ZB} &= (RP - 57.39)/43.24 \\
BP_{ZB} &= (BP - 65.09)/26.68 \\
GH_{ZB} &= (GH - 65.36)/22.04 \\
VT_{ZB} &= (VT - 60.02)/20.90 \\
SF_{ZB} &= (SF - 81.09)/25.58 \\
RE_{ZB} &= (RE - 75.80)/37.87 \\
MH_{ZB} &= (MH - 76.52)/17.21
\end{align*}
\]

Firstly, to achieve a value from 0-100 for the subscales (PF, RP, BP, GH, VT, SF, RE and MH) missing items and recoding of item categories needed attention before deriving the subscales. For the purpose of explaining the ceiling (highest possible value) and floor (lowest possible value) effects on the PCS by the use of examples, these issues (missing items and recoding) will be over looked for now.

The means and standard deviations used to calculate the subscales were from observed data of the ALSWH factor analytic (aged 70-74 years) sample. Using weighting from the factor analysis, the raw PCS scores were calculated by adding the weighted z-scores for each subscale. It is interesting to note; that the number of items per sub-scale are:

- PF = 10 items;
- RP = 4 items;
- BP = 2 items;
- GH = 5 items;
- VT = 4 items;
- SF = 2 items;
- RE = 3 items;
- and MH = 5 items.

Thus, 35 items of the SF-36 make up the PCS and MCS whereas the health transition item (compared to one year ago how would you rate your general health now?) is not included in the two summary component scores. When considering the weightings of each of the raw z-scores, negative weights are given to the role emotional (RE) and mental health (MH) subscales.
The ALSWH weightings are as follows:

\[
\text{prawb} = (\text{PF}_ZB \times 0.39753) + (\text{RP}_ZB \times 0.26911) + (\text{BP}_ZB \times 0.37008) + (\text{SF}_ZB \times 0.02453) + (\text{MH}_ZB \times -0.27163) + (\text{RE}_ZB \times -0.22946) + (\text{VT}_ZB \times 0.13676) + (\text{GH}_ZB \times 0.24270)
\]

Where \text{prawb} is the raw PCS value when weighted by subscales and \text{PF}_ZB represents the raw z-score for the Physical Functioning subscale and so on with the labeling of \_ZB.

So what does this mean? Consider the following scenarios:

**Scenario 1 - All subscales are equal to zero**

\[
\begin{align*}
\text{PF}_ZB &= (0 - 63.35)/25.94 = -2.44 \\
\text{RP}_ZB &= (0 - 57.39)/43.24 = -1.33 \\
\text{BP}_ZB &= (0 - 65.09)/26.68 = -2.44 \\
\text{GH}_ZB &= (0 - 65.36)/22.04 = -2.97 \\
\text{VT}_ZB &= (0 - 60.02)/20.90 = -2.87 \\
\text{SF}_ZB &= (0 - 81.09)/25.58 = -3.17 \\
\text{RE}_ZB &= (100 - 75.80)/37.87 = 0.64 \\
\text{MH}_ZB &= (100 - 76.52)/17.21 = 1.36
\end{align*}
\]

\[
\text{prawb} = (-2.44 \times 0.39753) + (-1.33 \times 0.26911) + (-2.44 \times 0.37008) + (-2.87 \times 0.24270) + (-2.87 \times 0.13676) + (-3.17 \times 0.0245) + (-2.00 \times -0.22946) + (-4.45 \times -0.27163) = -1.754
\]

The final PCS is obtained by standardisation of the raw z-scores using a mean of 50 and standard deviation of 10, which is the standard practice for the PCS.

\[
\text{PCS} = (-1.754 \times 10 + 50) = 32.46
\]

Thus the first scenario would equate to an ALSWH PCS score of 32.46

**Scenario 2 - PF, RP, BP, GH, VT, SF are all zero while RE and MH are 100**

\[
\begin{align*}
\text{PF}_ZB &= (0 - 63.35)/25.94 = -2.44 \\
\text{RP}_ZB &= (0 - 57.39)/43.24 = -1.33 \\
\text{BP}_ZB &= (0 - 65.09)/26.68 = -2.44 \\
\text{GH}_ZB &= (0 - 65.36)/22.04 = -2.97 \\
\text{VT}_ZB &= (0 - 60.02)/20.90 = -2.87 \\
\text{SF}_ZB &= (0 - 81.09)/25.58 = -3.17 \\
\text{RE}_ZB &= (100 - 75.80)/37.87 = 0.64 \\
\text{MH}_ZB &= (100 - 76.52)/17.21 = 1.36
\end{align*}
\]
\[ \text{prawb} = (-2.44 \times 0.39753) + (-1.33 \times 0.26911) + (-2.44 \times 0.37008) + (-2.87 \times 0.24270) + (-2.87 \times 0.13676) + (-3.17 \times 0.0245) + (0.64 \times -0.22946) + (1.36 \times -0.27163) = -3.94 \]

\[ \text{PCS} = (-3.938 \times 10) + 50 = 10.62 \]

Thus the second scenario would equate to an ALSWH PCS score of 10.62. It is important to note that if a participant has high MH and RE scores this equates to a decrease in the PCS. Therefore, the floor value on the ALSWH data basically means that no participant will be able to score below 10.62. Likewise calculating the ceiling value using the following scenario:

\[ \text{PF} = 100; \text{RP} = 100; \text{BP} = 100; \text{GH} = 100; \text{VT} = 100; \text{SF} = 100; \text{RE} = 0; \text{MH} = 0, \]

suggests that no participant in the ALSWH data can score above 86.3.

Finally, what do these scenarios mean and can they possibly occur? As previously mentioned in Table 6.2 the minimum PCS value recorded for an ALSWH participant at Survey 2 was 13.45, this score also happened to be for a participant who died between survey 2 and survey 3. This participant scored the following for each of the subscales:

\[ \text{PF} = 0; \text{RP} = 0; \text{BP} = 0; \text{GH} = 20; \text{VT} = 0; \text{SF} = 0; \text{RE} = 100; \text{MH} = 96. \]

While at baseline a participant was scored as (0, 0, 0, 0, 0, 0, 0, 0). These scenarios are possible whereas it would be less likely that (100, 100, 100, 100, 100, 0, 0) would occur as the survey continues and the participants become elderly. The decline in RE and MH is possible but maintaining very high scores for the other subscales may be less achievable at the same time point. The last scenario to consider is (100, 100, 100, 100, 100, 100), this results in a score of 64.55 (N=12 at baseline), again for the majority of participants maintaining extremely high subscales is unlikely over time.
6.5.6 Comparisons using general Australian populations aged 65 years older.

Out of interest, comparisons using another Australian general elderly population of 65 years and older were considered. Using the ALSWH data PCS scores were re-calculated based upon the factor analytic weightings and raw z-scores from the Australian study by Stevenson (1996) of an elderly population of 65 years and older. Stevenson (1996) was the first to standardise the Rand SF-36 with an Australian elderly population. The weightings, means and standard deviations are shown below and were applied to the baseline subscales for ALSWH data.

\[
\begin{align*}
PF_{ZA} &= (PF - 57.3)/28.8 \\
RP_{ZA} &= (RP - 56.0)/42.8 \\
BP_{ZA} &= (BP - 65.4)/28.6 \\
GH_{ZA} &= (GH - 61.1)/22.4 \\
VT_{ZA} &= (VT - 57.4)/21.4 \\
SF_{ZA} &= (SF - 77.3)/27.7 \\
RE_{ZA} &= (RE - 72.1)/37.0 \\
MH_{ZA} &= (MH - 75.3)/17.3 \\
\end{align*}
\]

\[
\text{prawa} = (PF_{ZA} \times 0.44) + (RP_{ZA} \times 0.41) + (BP_{ZA} \times 0.32) \\
+ (GH_{ZA} \times 0.14) + (VT_{ZA} \times -0.08) \\
+ (SF_{ZA} \times 0.02) + (RE_{ZA} \times 0.004) + (MH_{ZA} \times -0.31)
\]

Where prawa is the raw PCS value when weighted by subscales and PF_{ZA} represents the raw z-score for the Physical Functioning subscale and so on.

The means, standard deviations and weightings are different to the ALSWH and likewise the samples are different, with Stevenson’s (1996) sample having included men. Interestingly, the two negatively weighted subscales from Stevenson (1996) were VT and MH. Applying the same two scenarios from above, equates to PCS scores of 39.75 and 21.94. These results are considerably higher than the two earlier scenarios.
Figure 6.9  Regression results of PCS on SF-6D at survey 1 using the Stevenson (1996) subscale weightings for linear, logarithm and polynomial regression models.

Inserting values for zero, the results for the Y-intercept for the linear equation is as follows:

\[ \hat{Y}_{\text{PCS}} = \alpha + \beta_1 \times \text{SF}_6 \text{D} = 14.07 + 50.62 \times \text{SF}_6 \text{D} \]

\[ = 14.07 \text{ when } \text{SF}_6 \text{D} = 0 \]  \hspace{1cm} (Linear)

Interestingly, in Figure 6.9 the shape of cubic equation is clearly not appropriate, while the linear and quadratic equations (using the baseline data) are very similar. The logarithmic regression intercept value is above both the linear and quadratic equations. Also, interesting to note here the Stevenson (1996) Australian population PCS value of death (14.07) is similar to that suggested by Trisolini et al (2002), more of a coincidence than for any real reason.
6.5.7 What PCS value then to use for death?

Based upon the discussion in the previous section and what was shown earlier by the simple linear regression modeling of PCS and SF-6D at Survey 1, a rounded PCS value of 9 is reasonable for the ALSWH. The lowest possible value would be 10.62, so imputing a value of 9 seems a reasonable approximation for those who died. In the case of considering the use Trisolini et al’s PCS value of 14, it would appear that it would be possible that participants whom have died would be imputed with a higher PCS (i.e. 5 points) than those still alive but scoring poorly (low) on the PCS scale.

Another reason for using a PCS value of 9 rather than 14 is because the value of 14 is based on the use of the HALex, whereas here the SF-6D was used. A point to note is that the lowest observed value for the SF-6D is 0.301 where the HALex may produce values as low as 0.1. This meant that the use of a quadratic or cubic equation had further to extrapolate with the ALSWH data, which is another reason not to use the quadratic and cubic equations, due to the greater uncertainty. Also in support of the value of 9 and not 14, is that the PCS is weighted and standardized to the ALSWH cohort which as was shown earlier produced a much different floor value to that of the Stevenson (1996) sample. However, for comparative purposes the results from using cross-sectional analysis will also include a value of 14.

Finally, along with comparisons of a constant value of 9 and 14 for deaths, a decreasing value of 7, 5 and 5 at surveys 2, 3 and 4 was also considered. This analysis was compared with using a constant value of 9 to determine whether a declining death score should be used since the cohort’s PCS values are declining over time. The following values shown in Table 6.4 are the rounded values from either of the ALSWH models. According to approach 2 in Table 6.4, if a participant
died before survey 2 a PCS value of 7 would be imputed, whilst at survey 3 that participant would be scored as 5 and likewise for survey 4 a value of 5. Note in Table 6.4, the results are rounded down to nearest whole number.

Table 6.4 Approaches for imputing PCS values for death

<table>
<thead>
<tr>
<th>Survey</th>
<th>Approach 1</th>
<th>Approach 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>No deaths</td>
<td>No deaths</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

In Table 6.4, Approach 2 imputes a value of 7 at Survey 2 instead of 9, since this was the result from the regression of PCS and SF6D at survey 2. The decreasing values of Approach 2 takes into consideration the observed data at each survey, whereas the constant value of 9 was the result obtained from the linear regression of PCS and SF6D at survey 1. Approach 1 is an extension of Trisolini et al (2002) since the death PCS value was obtained from baseline data then used when deaths occur at any follow-up survey.

Table 6.5 has been shown to compare mean PCS values when deaths are excluded and included (i.e. PCSD) at surveys 2, 3 and 4. In Table 6.5 the use of PCS death values: 14, 9, 7 and 5 were included at survey 2, 3 and 4 respectively. These are the results from the simple linear regression models mentioned earlier and the Trisolini et al value. Table 6.5 also shows the minimum and maximum values for PCS and PCSD at each survey. Cross-sectional comparisons of PCS and PCSD by diabetes groups will be considered later. Interestingly, the minimum PCS values at survey 2 and 4 are 13.45 and 15.43 respectively. The constant value of 14 is shown merely as a comparison with a constant value of 9. Results (not shown) of
comparison tests suggest that similar results would be present if 14 or 9 were applied. As the value of 9 has been generated from the ALSWH data this value will be used throughout and reported, likewise comparisons to a decreasing value for death (7, 5, 5) will also be reported.

Table 6.5 Descriptive Statistics of Approach 1 – Comparisons of PCS and PCSD at Surveys 1-4

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean (SD)</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS</td>
<td>10851</td>
<td>50.0 (9.95)</td>
<td>17.9</td>
<td>75.5</td>
</tr>
<tr>
<td>PCSD</td>
<td>10851</td>
<td>50.0 (9.95)</td>
<td>17.9</td>
<td>75.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survey 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS</td>
</tr>
<tr>
<td>PCSD9</td>
</tr>
<tr>
<td>PCSD14</td>
</tr>
<tr>
<td>PCSD7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survey 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS</td>
</tr>
<tr>
<td>PCSD9</td>
</tr>
<tr>
<td>PCSD14</td>
</tr>
<tr>
<td>PCSD5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Survey 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCS</td>
</tr>
<tr>
<td>PCSD9</td>
</tr>
<tr>
<td>PCSD14</td>
</tr>
<tr>
<td>PCSD5</td>
</tr>
</tbody>
</table>
6.5.8 Cross-Sectional Data Analysis

6.5.8.1 Cross-sectional comparisons of PCSD at Survey 1-4

Oneway ANOVAs with Bonferroni multiple comparisons for post hoc pair-wise comparisons were used to examine the mean differences between diabetes groups. The comparisons are cross-sectional and Table 6.7 and Table 6.8 show the results when considering a different PCS value for death. Table 6.6 shows the results of the mean comparisons at each survey to highlight what occurs when deaths are ignored.

6.5.8.2 Results from Comparing Survivor groups

In Table 6.6 the results of the multiple mean comparisons of the diabetes groups focuses on the survivors. The results suggest that from the comparisons of Existing cases and No diabetes the differences are statistically significant at each survey (P-value<0.001). The mean PCS values for the remaining survivors for Existing cases decrease by (45.3 - 42.8) 2.5 points from survey 1 to 4. The decrease in mean PCS for the No diabetes group is (50.9 - 46.7) 4.2 points over the 4 surveys. The multiple comparison Bonferroni tests suggest that the mean differences between the survivors in the two groups at each survey are statistically significant. All comparisons are statistically significant except for survey 4 for New cases verses Existing cases, where the difference is 0.5 of a PCS point. It appears that the survivors at each survey are maintaining a similar level of physical health. The results from paired t-tests (not all results shown) for Existing cases of diabetes survivors between surveys 1 and 2 (n=649 pairs, mean difference= -0.86, P-value =0.011) and between surveys 2 and 3 (n=438 pairs, mean difference = -2.26, P-value <0.001) and finally surveys 3 and 4 (n=319 pairs, mean difference = -2.09, P-value <0.001) are statistically significant different. There appears to be declines of approximately 2 points between surveys for surviving existing Diabetes cases with
complete data. It must be remembered that the sample sizes for each group are quite large so small results are generally statistically significant, even with the significance set at 0.001. Whether these differences are clinically important is another matter. Comparisons for the No diabetes survivors are statistically significant (again due to large sample sizes) and also appear to be declining at approximately 2 points (results not shown). Thus the comparisons in Table 6.6 suggest that the difference in means between the two groups of survivors is approximately 5 points at each survey (ranging from 4.0 - 5.6).

Table 6.6  Cross-sectional comparisons of PCS at Survey 1, 2, 3 and 4

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean (SD)</th>
<th>**Bonferroni multiple comparison test P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>937</td>
<td>45.3 (9.9)</td>
<td>Exist v New 3.7 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>647</td>
<td>49.0 (9.6)</td>
<td>No D v New 2.0 (&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6577</td>
<td>50.9 (9.7)</td>
<td>No D v Exist 5.6 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>748</td>
<td>45.6 (10.2)</td>
<td>Exist v New 2.1 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>671</td>
<td>47.7 (9.3)</td>
<td>No D v New 2.5 (&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6232</td>
<td>50.2 (9.8)</td>
<td>No D v Exist 4.6 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>496</td>
<td>43.4 (9.8)</td>
<td>Exist v New 2.3 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>537</td>
<td>45.7 (9.6)</td>
<td>No D v New 2.8 (&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5283</td>
<td>48.5 (10.1)</td>
<td>No D v Exist 5.1 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>391</td>
<td>42.8 (9.5)</td>
<td>Exist v New 0.5 (1.00)</td>
</tr>
<tr>
<td>New Cases</td>
<td>504</td>
<td>43.2 (9.4)</td>
<td>No D v New 3.5 (&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5252</td>
<td>46.7 (10.2)</td>
<td>No D v Exist 4.0 (&lt;0.001)*</td>
</tr>
</tbody>
</table>

** New = New cases, No D= No diabetes, Exist=Existing cases
*Statistically Significant at alpha level of 0.001
Figure 6.10 Cross-sectional comparisons of PCS at Survey 1, 2, 3 and 4

Figure 6.10 represents the mean PCS values of Existing cases and No diabetes groups at surveys 1 to 4. Figure 6.10 highlights the results shown in Table 6.6. This figure suggests that over time, as participants have died (in both groups), or have no PCS value (for other reasons) the mean difference between the two groups narrows over time.

### 6.5.8.3 Results from Comparing groups with Decedents ‘reinstituted’

In Table 6.7, the mean group difference of Existing cases and No diabetes is statistically significant (P-value <0.001) at all four surveys. At survey 2, 3 and 4 a PCS value of 9 was included for those who died. For each comparison of those with existing diabetes and those without diabetes, the mean difference widens at each survey. The greatest mean difference is at survey 4 where the proportion of deaths for either group is the greatest, and more subjects are ‘reinstituted’. At survey 1 the
mean difference between new cases and no diabetes is statistically significant, whereas, when the deaths are ‘reinstated’ the mean differences at surveys 2-4 are non-significant. Results are similar to those in Chapter 4 but the outcome there was comparing the probabilities of future health.

Table 6.7 Cross-sectional comparisons of PCSD9 (death = 9) and Surveys 1-4

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean (SD)</th>
<th>**Bonferroni multiple comparison test P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>937</td>
<td>45.3 (9.9)</td>
<td>Exist v New 3.7 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>647</td>
<td>49.0 (9.6)</td>
<td>No D v New 2.0 (&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6577</td>
<td>50.9 (9.7)</td>
<td>No D v Exist 5.6 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>842</td>
<td>41.5 (15.0)</td>
<td>Exist v New 6.2 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>671</td>
<td>47.7 (9.3)</td>
<td>No D v New -0.1 (1.00)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6647</td>
<td>47.6 (13.8)</td>
<td>No D v Exist 6.1 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>685</td>
<td>33.9 (17.5)</td>
<td>Exist v New 10.9 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>550</td>
<td>44.8 (11.0)</td>
<td>No D v New -1.5 (0.096)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6087</td>
<td>43.3 (16.3)</td>
<td>No D v Exist 9.3 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>687</td>
<td>28.2 (18.2)</td>
<td>Exist v New 11.9 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Cases</td>
<td>554</td>
<td>40.1 (13.3)</td>
<td>No D v New -0.7 (1.00)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6509</td>
<td>39.5 (17.5)</td>
<td>No D v Exist 11.2 (&lt;0.001)*</td>
</tr>
</tbody>
</table>

** New = New cases, No D = No diabetes, Exist = Existing cases
*Statistically Significant at α level of 0.001
There does appear to be some impact on the results at each survey when comparing Tables 6.6 and 6.7. At survey 1 for both PCS and PCSD the results are the same since no deaths have been imputed. However, as a PCS value of 9 was imputed for deaths and carried forward, results in Table 6.7 compared to Table 6.6 appear to show the mean difference, between existing diabetes and no diabetes groups, widening.

Figure 6.11  Cross-sectional comparisons of PCSD9 at Survey 1, 2, 3 and 4.

Figure 6.11 highlights the widening in mean PCS when death values of 9 were included at Surveys 2, 3 and 4. In Figure 6.11 the mean difference at Survey 4 was approximately 3 times as wide as in Figure 6.10, due to the inclusion of deaths=9. Similar comparisons were made using a death value 14 (Trisolini et al's value). Results are not shown but the same noticeable differences were observed. The larger value for death (14) resulted in a slightly narrower mean difference in existing cases and no diabetes groups at each survey compared to using a value of 9 in
Tables 6.6. However, the mean differences in all comparisons, at all surveys, are much larger in Table 6.8 when compared to Table 6.7 when no deaths are considered. Interestingly, by Survey 4 the mean difference between Existing cases (30.3) and No diabetes (40.4) was 10.1 which is not as wide as when death is equal to 9.

Table 6.8  Cross-sectional comparisons of PCSD7, 5, 5 at Surveys 1-4.

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean (SD)</th>
<th>**Bonferroni multiple comparison test P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>937</td>
<td>45.3 (9.9)</td>
<td>Exist v New 3.7 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>647</td>
<td>49.0 (9.6)</td>
<td>No D v New 2.0 (&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6577</td>
<td>50.9 (9.7)</td>
<td>No D v Exist 5.6 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>842</td>
<td>41.3 (15.5)</td>
<td>Exist v New 6.4 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>671</td>
<td>47.7 (9.3)</td>
<td>No D v New -0.2 (1.00)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6647</td>
<td>47.5 (14.1)</td>
<td>No D v Exist 6.2 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>685</td>
<td>32.8 (19.1)</td>
<td>Exist v New 11.9 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>550</td>
<td>44.7 (11.3)</td>
<td>No D v New -1.9 (0.03)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6087</td>
<td>42.8 (17.5)</td>
<td>No D v Exist 10.0 (&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>687</td>
<td>26.5 (20.0)</td>
<td>Exist v New 13.3 (&lt;0.001)*</td>
</tr>
<tr>
<td>New Cases</td>
<td>554</td>
<td>39.8 (14.1)</td>
<td>No D v New -1.1 (0.552)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>6509</td>
<td>38.7 (18.8)</td>
<td>No D v Exist 12.2 (&lt;0.001)*</td>
</tr>
</tbody>
</table>

** New = New cases, No D= No diabetes, Exist=Existing cases
*Statistically Significant at α level of 0.001

In Tables 6.8, the PCS value for death was allowed to vary at survey 2, 3 and 4. The values used were from the earlier linear regressions. The values for the Y-intercepts were rounded to whole numbers 7, 5, 5. There are statistically significant differences in the Existing cases versus No diabetes (P-value <0.001) and likewise for the New cases versus Existing cases (P-value <0.001). The mean difference at
survey 4 when death=5 was 12.1 (P-value<0.001). This mean difference of 12.1 is larger than when the death values of 14 or 9 are used. In Figure 6.12 the decline in both Existing cases and No diabetes groups is more noticeable than in Figure 6.10 when there no deaths. The decreasing death values are having an impact on mean results, but Figures 6.11 and 6.12 look very similar and both differ greatly from the graph shown in Figure 6.10.

**Figure 6.12**  Cross-sectional comparisons of PCSD7,5,5 at Survey 1, 2, 3 and 4.

### 6.5.9 Longitudinal Data Analysis

The results shown in this section have used the imputed PCS death values of 9 and then 7, 5, 5 at surveys 2, 3 and 4 respectively. The cross-sectional results above were not adjusted for covariates, whereas in this section possible baseline covariates will be considered. The main purposes of the longitudinal analysis are to evaluate the impact of imputing values for the missing outcomes and baseline
covariates and compare the modelling approaches and not to determine the “best fitting” model.

Tables 6.9 and 6.10 show the results of the range of covariate adjusted using Fully Bayesian random effects models. The purpose of these Tables is to show comparisons of the diabetes groups using a similar imputation approach to that used in Chapter 5. This time however, the outcome is PCSD and not the probability of being health in 3 years (APCTD). Table 6.9 and 6.10 show the results from a number of different models that are adjusted comparisons of diabetes groups. Each table shows a complete case model for PCS (without deaths) which has been generated using WinBUGS. This complete case model is repeated in both tables because all other models involve the imputation of possible PCS death values and other missing outcome or baseline covariate data. This model (model 1) could be considered similar to an approach where researchers may have used the repeated measures MANOVA. As has been previously mentioned this complete case analysis approach is no longer considered acceptable in longitudinal analysis especially when there is a large proportion of missing data. Three of the next four models (models 2-4) in each table show results from WinBUGS models again using a similar approach to that in Chapter 5. While the last model (model 5), has been included in each Table to show results where all missing data (PCS and baseline covariates) are imputed (except for diabetes status). In order to remain consistent when comparing the diabetes groups, missing diabetes status was not imputed.

6.5.9.1 Comparisons of diabetes status - WinBUGS – PCSD9

In Table 6.9 the complete case model (n=4352), Model 1 considers only those who have complete data for all covariates and a PCS value at each survey. Using WinBUGS, the covariate adjusted mean difference -3.26; 95% CI (-4.24, -2.30) between Existing diabetes and No diabetes groups is statistically significant. In
model 2 where a PCS death value of 9 has been included, the mean difference in PCS is -6.70; 95% CI (-7.72, -5.70) and has actually doubled the mean difference between the two groups. This result is also statistically significant. The inclusion of those who died increased the sample size by 25.2% (n=5451). Model 2 has been included as a “pseudo complete case” model since reinstating those who died means that the decedents will have a recorded PCS value before and after death. Thus, many decedents then had pseudo complete case data for all 4 surveys. There were however, those women who were still not included in model 2 because they had missing PCS values before death.

The importance of considering the impact of missing data has been a regular theme throughout the thesis and is again considered in Tables 6.9 and 6.10. Models 3 and 4 in Table 6.9 use the WinBUGS multiple imputation approaches for MAR and MNAR. Model 3 assumes MAR where the adjusted mean difference between Existing diabetes and No diabetes was -8.07; 95% CI (-8.96, -7.20). In model 4, the mean difference between the two groups is -8.10; 95% CI (-8.94, -7.31). The imputation of missing PCS values has widened the adjusted mean differences in both models. However, results are very similar for both the MAR and MNAR models.
Figure 6.13 Longitudinal Analysis – Mean PCS when death=9 but without other imputation of PCS

Figure 6.13 represents the mean PCS over time for Existing and No diabetes groups when a death value of 9 is included. The adjusted means clearly show a widening gap between the two groups. The results shown in Figure 6.13 were generated from a random effects model which included a time X group interaction and allowed for random intercepts (Results from the model are not shown).

6.5.9.2 Comparisons of diabetes status- WinBUGS – PCSD 7, 5, 5

In Table 6.10 the complete case model 1 was again included. Whereas in model 2 the death score was allowed to vary (7, 5, 5) at surveys 2, 3, 4. In model 2 the adjusted mean difference in Existing and No diabetes is -6.56; 95% CI (-7.59, -5.55). This result is also statistically significant. The effect size for the varying death values is similar to when 9 is imputed in model 2 (Table 6.9) -6.70; 95% CI: (-7.72, -5.70).
In Table 6.10 model 3 assumes MAR and the difference in Existing diabetes and No diabetes was -8.50; 95% CI (-9.44, -7.58) while model 4 (MNAR) the mean difference between the two groups is -8.51; 95% CI ( -9.46, -7.59). These results for the MAR and MNAR models like in Table 6.9 are similar. Models 3 and 4 show that mean difference was greater than for model 2, where missing PCS values are not considered only deaths. Whereas the MAR and MNAR models in Table 6.10 compared to Table 6.9 show the effect size to be slightly wider. The greater value for death, used in Table 6.19, appears to decrease the mean difference in mean PCS when all models are compared across the two tables.

6.5.9.3 Comparisons of diabetes status using Stata (MAR)

Tables 6.9 and 6.10 also show imputed models (model 5), where the assumption of MAR was assumed. These models are presented as a naïve approach to “missingness”. Stata's `ice` command allows for multiple imputation of dependent and independent variables with the assumption of MAR (Royston 2005). Tables 6.9 and 6.10 report the result of a random effects model (model 5), where missing baseline covariates, missing PCS (for reasons other than death) and PCS death values have been imputed. The result was 5 complete imputed dataset using the imputed (differing) death values, then using the `mim` command to generate an adjusted difference in diabetes and no-diabetes groups.

In Table 6.9 the mean difference in Existing cases and No diabetes groups was -6.10; 95% CI: (-6.74, -5.47) (P-value <0.001). This model was shown here because it allows for missing baseline covariates to be imputed unlike the WinBUGs models shown. In Table 6.10, where the death score was allowed to vary from 7 to 5, the results of model 5 were -4.39; 95%CI (-4.99, -3.79) (P-value <0.001). The results for both models (5) in Table 6.9 and 6.10 are statistically significant and similar.
Interestingly, comparing model 5 to the “pseudo complete case” model 2 the results from the WinBUGs and Stata models in Table 6.9 appear to be similar (not as close in Table 6.10). These results will be discussed later.
Table 6.9 Comparisons of regression coefficients using WinBUGs and Stata (PCSD9)

<table>
<thead>
<tr>
<th>PCSD9</th>
<th>Complete Case Model 1 PCS* (n= 4352*)</th>
<th>Pseudo Complete Case Model 2 PCSD (n= 5451*)</th>
<th>Imputed model Model 3 PCSDI (n=8368)</th>
<th>Imputed model Model 4 PCSDI ( N= 8368)</th>
<th>Imputed missing baseline covariates and PCSD Model 5 PCSDI ( N=9181)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25000 iterations 5000 burn ins</td>
<td>25000 iterations 5000 burn ins</td>
<td>25000 iterations 5000 burn ins</td>
<td>25000 iterations 5000 burn ins</td>
<td>Stata MICE 5 Imputations</td>
</tr>
<tr>
<td><strong>Mean (95% PI)</strong></td>
<td><strong>Mean (95% PI)</strong></td>
<td><strong>Mean (95% PI)</strong></td>
<td><strong>Mean (95% PI)</strong></td>
<td><strong>Mean (95% PI)</strong></td>
<td><strong>Mean (95% PI)</strong></td>
</tr>
<tr>
<td>Intercept</td>
<td>59.18 (58.67, 59.72)</td>
<td>56.34 (55.77, 56.94)</td>
<td>45.71 (45.15,46.27)</td>
<td>45.72 (45.16, 46.26)</td>
<td>55.28 (54.81, 55.74)</td>
</tr>
<tr>
<td>Time (Survey)</td>
<td>-1.79 (1.87, -1.71)</td>
<td>-3.89 (-4.03, -3.75)</td>
<td>-3.80 (-3.92, -3.68)</td>
<td>-3.80 (-3.92, -3.69)</td>
<td>-3.47 (-3.59, -3.36)</td>
</tr>
<tr>
<td>No Diabetes New Cases</td>
<td>-2.03 (-2.86, -1.21)</td>
<td>-1.08 (-2.16, -0.003)</td>
<td>-0.47 (-1.39, 0.39)</td>
<td>-0.46 (-1.52, 0.47)</td>
<td>-1.07 (-1.82, -0.31)</td>
</tr>
<tr>
<td>Existing Cases</td>
<td>-3.26 (-4.24, -2.30)</td>
<td>-6.70 (-7.72, -5.70)</td>
<td>-8.07 (-8.96, -7.20)</td>
<td>-8.10 (-8.94, -7.31)</td>
<td>-6.10 (-6.74, -5.47)</td>
</tr>
<tr>
<td>Co-morbidities</td>
<td>-3.09 (-3.30, -2.87)</td>
<td>-0.74 (-0.98, -0.50)</td>
<td>0.40 (0.16, 0.64)</td>
<td>0.39 (0.16, 0.63)</td>
<td>-0.50 (-0.67, -0.32)</td>
</tr>
<tr>
<td>age</td>
<td>-0.31 (-0.75, 0.13)</td>
<td>-0.98 (-1.52, -0.44)</td>
<td>-1.51 (-2.03, -0.98)</td>
<td>-1.52 (-2.04, -0.99)</td>
<td>-1.11 (-1.52, -0.70)</td>
</tr>
<tr>
<td>smoke</td>
<td>-0.35 (-1.37, 0.68)</td>
<td>-1.14 (-2.22, -0.03)</td>
<td>-4.08 (-5.12, -3.03)</td>
<td>-4.09 (-5.12, -3.08)</td>
<td>-2.09 (-2.84, -1.34)</td>
</tr>
<tr>
<td>area</td>
<td>0.23 (-0.23, 0.69)</td>
<td>0.38 (-0.18, 0.96)</td>
<td>0.39 (-0.17, 0.93)</td>
<td>0.27 (-0.16, 0.92)</td>
<td>0.49 (-0.08, 0.91)</td>
</tr>
</tbody>
</table>

*PCS=PCSD since no deaths are included only surviving participants ** 2.5%, 97.5% Posterior Limits, PI=Posterior Interval

* N= 5451 represents the number of participants per survey with observed PCSD values (PCSD=9) (Random intercept and coefficient mixed model using WinBUGS).  
* N= 4352 represents the number of participants per survey with observed PCSD values (PCSD=9) (Random intercept and coefficient mixed model using WinBUGS).  
* N= 8368 represents the number of participants per survey with (8368 x 4 =) 33472 observed PCSD values (PCSD=9) (Random intercept and coefficient mixed model using WinBUGS).  
* N= 9181 the number of participants per survey with (9181 x 4 =) 36 724 observed PCSD values (PCSD=9) and all covariates (Random intercept and coefficient mixed model Stata).
### Table 6.10 Comparisons of regression coefficients using WinBUGs and Stata (PCSD7-5-5)

<table>
<thead>
<tr>
<th>PCSD14</th>
<th>Complete Case Model 1 (PCS* (n= 4352*) 25000 iterations 5000 burn ins)</th>
<th>Pseudo Complete Case Model 2 (PCSD (n= 5451*) 25000 iterations 5000 burn ins)</th>
<th>Imputed model Model 3 (PCSDI (n=8368*) 25000 iterations 5000 burn ins)</th>
<th>Imputed model Model 4 (PCSDI (N= 8368) 25000 iterations 5000 burn ins)</th>
<th>Imputed missing baseline covariates and PCSD Model 5 (PCSDI (N=9181) Stata MICE 5 Imputations)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Covariates</strong></td>
<td><strong>β</strong> (95% PI)**</td>
<td><strong>β</strong> (95% PI)</td>
<td><strong>β</strong> (95% PI)</td>
<td><strong>β</strong> (95% CI)</td>
</tr>
<tr>
<td></td>
<td>Intercept</td>
<td>59.18 (58.67, 59.72)</td>
<td>56.73 (56.16, 57.34)</td>
<td>45.1 (44.53, 45.69)</td>
<td>45.14 (44.57, 45.72)</td>
</tr>
<tr>
<td></td>
<td>Time (Survey)</td>
<td>-1.79 (1.87, -1.71)</td>
<td>-4.17 (-4.32, -4.02)</td>
<td>-4.08 (-4.21, -3.95)</td>
<td>-4.08 (-4.21, -3.95)</td>
</tr>
<tr>
<td></td>
<td>No Diabetes New Cases</td>
<td>-2.03 (-2.86, -1.21)</td>
<td>-1.09 (-2.19, -0.01)</td>
<td>-0.33 (-1.39, 0.70)</td>
<td>-0.35 (-1.39, 0.70)</td>
</tr>
<tr>
<td></td>
<td>Existing Cases</td>
<td>-3.26 (-4.24, -2.30)</td>
<td>-6.56 (-7.59, -5.55)</td>
<td>-8.50 (-9.44, -7.58)</td>
<td>-8.51 (-9.46, -7.59)</td>
</tr>
<tr>
<td></td>
<td>Co-morbidities</td>
<td>-3.09 (-3.30, -2.87)</td>
<td>-0.82 (-1.07, -0.57)</td>
<td>0.63 (0.38, 0.88)</td>
<td>0.61 (0.36, 0.86)</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>-0.31 (-0.75, 0.13)</td>
<td>-0.92 (-1.46, 0.37)</td>
<td>-1.59 (-2.13, -1.04)</td>
<td>-1.61 (-2.15, -1.06)</td>
</tr>
<tr>
<td></td>
<td>smoke</td>
<td>-0.35 (-1.37, 0.68)</td>
<td>-0.92 (-2.01, 0.20)</td>
<td>-4.36 (-5.43, -3.27)</td>
<td>-4.37 (-5.40, -3.30)</td>
</tr>
<tr>
<td></td>
<td>area</td>
<td>0.23 (-0.23, 0.69)</td>
<td>0.38 (-0.18, 0.96)</td>
<td>0.42 (-0.16, 0.99)</td>
<td>0.38 (-0.20, 0.94)</td>
</tr>
</tbody>
</table>

*PCS=PCSD since no deaths are included only surviving participants ** 2.5% , 97.5% Posterior Limits, PI=Posterior Interval

* N= 5451 represents the number of participants per survey with observed PCSD values (PCSD=7, 5, 5) (Random intercept and coefficient mixed model using WinBUGS).

* N= 4352 represents the number of participants per survey with observed PCSD values (PCSD=7, 5, 5) (Random intercept and coefficient mixed model using WinBUGS).

* N= 8368 represents the number of participants per survey with (8368 x 4 =) 33472 observed PCSD values (PCSD=7, 5, 5) (Random intercept and coefficient mixed model using WinBUGS).

* N= 9181 represents the number of participants per survey with (9181 x 4 =) 36724 observed PCSD values (PCSD=7, 5, 5) and all covariates (Random intercept and coefficient mixed model Stata).
6.6 DISCUSSION

6.6.1 Cross-Sectional comparisons of death values

The results shown in Table 6.7 where a PCS value of 9 was included for deaths indicate that the mean differences at each survey for Existing cases and No diabetes are highly significant. In Tables 6.7 and 6.8 it was quite noticeable by survey 4 that the mean difference in PCS values for the two groups was 11.2 points and 12.2 points respectively and both results may be considered clinically important. In Tables 6.7 and 6.8 the mean differences again highlight the importance of considering those who died when making comparisons, especially when considering a sicker group. Those who are surviving with ‘long term’ diabetes may have a similar perception of health at their stage in life as those without diabetes and are possibly maintaining a level of physical health in relation to the burden of the disease.

6.6.2 Comparisons to Trisolini et al’s PCS value for death of 14

Earlier results and discussion lead to the use of ALSWH data generated value for death. This value was created using baseline data when all observed data were available. Using a constant of 14 points for death was considered to be too high for these data and would “overvalue” the death score. Assuming that the expected lowest PCS score within the surviving participants to be higher than the PCS score for death is reasonable. Thus 14 points for death over time may be higher than the lowest score for a participant. It was previously mentioned that the floor value for PCS was 10.62 and by Survey 3 and 4 the lowest PCS score was 13.45. Results from the longitudinal analysis suggest that using a value of 9 for death and imputing for missing PCS that the difference in existing diabetes and no diabetes groups maybe slightly wider. Over
the course of the ALSWH this gap may widen and may have greater clinical importance.

6.6.3 The Use of Decreasing PCS Values for Death (7, 5, 5)

It appears that as the observed data from the study decreases it could be hypothesized by Survey 5 that the PCS death value would be lower than 5. These decreasing PCS values for death may be reflecting the ageing of the population over time. Another question then arises if a decreasing death value is considered, that is, could the PCS death value (Y-intercept) become zero or negative (due to the linear extrapolation method) over time? As noted above the ceiling PCS values and SF6D values suggest that this may not be possible as health continues to decline. However, an extrapolated PCS value close to zero will not be totally ruled out at this stage.

6.6.4 The Impact of Using of Multiple Imputation

In Table 6.9 and 6.10, it is clear no matter which value for death is considered that multiple imputation impacted on the results. As appears to be the case in Chapter 5, if deaths are reinstated then there also needs to be some compensation for this. The impact of reinstating deaths in the Existing cases group has a much greater impact on the mean PCS since it was originally a smaller groups and lost a larger proportion of women over time. Thus, allowing for the imputation of missing PCS values appears to soften the impact of reinstating deaths. The results from the MAR and MNAR appear to represent a fairer difference in groups over time. Noticeably the results from the MAR and MNAR models using WinBUGS are similar. When baseline covariates and PCS values were imputed using Stata’s ice procedure the results (in Tables 6.9 and 6.10) are still statistically significant for diabetes groups and again may be a fairer indication of what is occurring over time. These models have been adjusted for baseline
covariates so as imputation of covariates could be considered. It is not possible to compare across the all models in Table 6.9 and 6.10 but at least comparisons between MNAR and MAR are useful. The initial starting values for the Beta estimates and variances (precision-tau) used in the WinBUGS models were originally generated from Stata mixed effect models. There were noticeable differences in the final coefficients for covariates and time in both Tables 6.9 and 6.10, but again the use of the different statistical approaches may account for this and any comparisons between Stata and WinBUGS models are not appropriate.

It was mentioned in the results that models 2 and 5 in Table 6.9 were similar and although straight comparisons are not appropriate they do raise interest. It is important to remember that the “pseudo complete case” was restricted, so that any participants with missing covariates were excluded. So the only participants reinstated in this model 2 were those who had died and had complete baseline covariates. The deaths were reinstated with a PCS value for death (9) so that they now appear to have complete data at every time point. In model 5 all missing PCS and missing baseline covariates are imputed. Model 2 was merely shown to demonstrate the impact of the deaths if only they are reinstated. Model 2 would not be a recommended model for reporting. If MAR was considered then model 5 might be a more appropriate model to report than model 1 or model 2.

The strengths and the limitations of this case study using a PCS value for death will be discussed in the Summary Chapter 8. Also, in this final Chapter recommendations about the possible and the plausible usage of this method will be discussed. There will also be suggestions about future research which could be developed as a result of this method.
CHAPTER 7:

Accounting for deaths in longitudinal studies –

Using the adapted HOS Methodology

(Same or Better Vs Worse)
### List of Abbreviations used in Chapter 7

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSWH</td>
<td>Australian Longitudinal Study on Women’s Health</td>
</tr>
<tr>
<td>APASB</td>
<td>Australian Probability of being Alive and Same or Better health in 3 years</td>
</tr>
<tr>
<td>APASBD</td>
<td>Australian Probability of being Alive and Same or Better health in 3 years (includes deaths)</td>
</tr>
<tr>
<td>APCT</td>
<td>(Australian) Probability of being healthy in three years</td>
</tr>
<tr>
<td>APCTD</td>
<td>Probability of being healthy in three years (includes deaths)</td>
</tr>
<tr>
<td>ANOVA</td>
<td>ANalysis Of VAriance</td>
</tr>
<tr>
<td>CMS</td>
<td>Centers for Medicare and Medicard Services</td>
</tr>
<tr>
<td>HOS</td>
<td>(Medicare) Health Outcome Survey</td>
</tr>
<tr>
<td>HAL</td>
<td>Health Assessment Laboratory group</td>
</tr>
<tr>
<td>ICE</td>
<td>Imputation using Chained Equations</td>
</tr>
<tr>
<td>MCS</td>
<td>Mental Component Summary score</td>
</tr>
<tr>
<td>MICE</td>
<td>Multiple Imputation using Chained Equations</td>
</tr>
<tr>
<td>MI</td>
<td>Multiple Imputation</td>
</tr>
<tr>
<td>MIM</td>
<td>manipulating multiply imputed datasets</td>
</tr>
<tr>
<td>PASB</td>
<td>Probability of being Alive and Same or Better health in 3 years</td>
</tr>
<tr>
<td>PASBD</td>
<td>Probability of being Alive and Same or Better health in 3 years (includes deaths)</td>
</tr>
<tr>
<td>PCS</td>
<td>Physical Component Summary score</td>
</tr>
<tr>
<td>PCSD</td>
<td>PCS with a score for death</td>
</tr>
<tr>
<td>PCTD</td>
<td>Diehr's Probability of being healthy in three years (includes deaths)</td>
</tr>
<tr>
<td>SAS</td>
<td>Statistical Analysis Software</td>
</tr>
<tr>
<td>SEM</td>
<td>Standard Error of Measurement</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short Form 36</td>
</tr>
<tr>
<td>SF-36 v1</td>
<td>Short Form 36 Version 1</td>
</tr>
<tr>
<td>SF-36 v2</td>
<td>Short Form 36 Version 2 1996</td>
</tr>
<tr>
<td>UVIS</td>
<td>univariate imputation sampling</td>
</tr>
<tr>
<td>VR-12</td>
<td>Veteran’s Rand (Version) Short Form 12</td>
</tr>
<tr>
<td>WinBUGS</td>
<td>Windows for Bayesian Using Gibbs Sampling</td>
</tr>
</tbody>
</table>
7.0 INTRODUCTION

This chapter will discuss another approach to accounting for deaths in longitudinal studies using HR-QOL summary measures. Mentioned in earlier chapters were a number of researchers engaged in examining ways to account for deaths in longitudinal studies (Trisolini et al 2005; Diehr et al 2007; Kazis et al 2007; Selim et al 2007). This chapter will consider an approach to account for deaths currently being used by the Medicare Health Outcome Survey in USA (Rogers et al 2004; Selim et al 2007).

The HOS approach considers the probability of being alive and in same or better health in 2 years. The approach has defined same or better health using a cut-off value for a change in Physical Component Summary score over two time points. This approach uses a conditional probability by combining the probability of being Alive in two years with the probability of being in Better or same health in two years.

This approach will be adapted and applied to the ALSWH data as a third approach to account for deaths in longitudinal studies. The focus again will be changes in health over time between those women with Existing cases of diabetes compared with those women who are considered as New cases and No diabetes.

7.1 BACKGROUND

The Medicare Health Outcome Survey (HOS) is an annual evaluation of the physical and mental health of Medicare beneficiaries enrolled in managed care plans (Selim et al 2007). During the 1990’s a need for better management care plans was undertaken by the Centers for Medicare and Medicard Services (CMS) (Selim et al 2007). According to the CMS website: [http://www.hosonline.org/surveys/hos/partners.aspx](http://www.hosonline.org/surveys/hos/partners.aspx).
CMS has commissioned a number of partners to assist in developing and evaluating HOS. Selim et al (2007) suggest that in order for reports of the performance of the health care plans to be effective, profiling is required using the best statistical methods to control for variation in the case-mix characteristics. The HOS program uses the outcome of change in health status after adjustment with a multi-modelling case mix methodology. This complex case-mix methodology was developed by Rogers et al (2004) (Selim et al 2007). They define case-mix as “those patient specific characteristics that are outside the control of the plan and that could contribute to better or worse outcomes, but over which the plan has little or no influence” (Selim et al 2007). In a technical report commissioned by CMS, the RTI International group developed statistical methods to “adjust the predicted two-year outcomes for the case-mix characteristics of the groups under study” (Selim et al 2007). In their new adjusted case-mix method they used a limited number of variables such as: age, baseline physical and mental health status, congestive heart failure and angina/coronary artery disease in order to predict physical health at follow-up in 2 years (Selim et al 2007). The new case-mix method instead of using the multi-modelling approach to death used a single model and a correction factor to account for missing case-mix information (Selim et al 2007). This appeared to be a very useful approach, however, there was a limitation to the case-mix adjustments in “identifying differences in among health care systems that have clinical or policy relevance” (Selim et al 2007). A limitation is the possible bias of results due to the exclusion of participants with incomplete or missing data. (Selim et al 2007).

The following sections will outline in more detail the two case-mix methods and discuss the pros and cons of both methods. In an attempt to overcome the limitations and complex nature of the case-mix methods, later there will be discussion of a new approach that will use multiple imputation and extend the method beyond two time points (baseline and a two year follow-up).
7.1.1 Original and Current HOS Case-Mix methodology

Ware and his colleagues at the Health Assessment Laboratory (HAL) modified an earlier approach to handling deaths for PCS scores. The new method is known as the HOS Case-Mix methodology (Trisolini et al 2002; Rogers et al 2004; Selim et al 2007) but sometimes is referred to as the HAL Case-Mix approach (Trisolini et al 2005).

The HOS performance measurement results are computed using the Case-Mix methodology (Selim et al 2007). The original HOS Case-Mix approach uses information on baseline health status from HOS beneficiaries such as: PCS, MCS, chronic conditions and socio-demographic data (Selim et al 2007). A series of eight multiple logistic regression death models are created and three different models for the outcomes of the same or better physical health or mental health over 2 years (Rogers et al 2004). The models range from a more complex to a basic structure, which controls for such factors as: socio-demographic characteristics, specific chronic conditions, specific health status domains, age, gender, ethnicity and Medicaid status (Kazis et al 2007). For example, a very basic structure might involve a model that includes just the PCS score and age as covariates. For beneficiaries with data for just these two covariates, then they (according to their profile) would have the expected probability from this model imputed. A complex structure may involve all the suggested covariates and thus those beneficiaries are imputed with an expected value that is similar to their profile (similar age, gender, PCS etc). Expected values were calculated for each participant (beneficiary) using the most complex model as possible based on what covariate data were available for the beneficiary (Selim et al 2007).
7.1.2 Selim et al’s Alternative Case-Mix methodology

The HOS Case-Mix approach uses models which are theory-driven that could be a downside for end-users. The employment of an approximate case-mix for an individual can be difficult and has not been used beyond two a two year follow-up. Selim et al (2007) saw a need to make the “Case-Mix methodology more straightforward and understandable to both a technical and lay audience” (Selim et al 2007). They put forward an alternative Case-Mix methodology and compared their alternative to the current HOS Case-Mix methodology.

There were two important differences in the methods used to calculate the expected values. The alternative Case-Mix methodology suggested by Selim et al (2007) used a single multivariate logistic regression model to calculate the expected probability values for death. In contrast, the Rogers et al (2004) HOS Case-Mix method uses a series of eight different death models.

The second difference was that the Selim et al (2007) alternative Case-Mix method used a correction factor to account for those beneficiaries with missing Case-Mix information. According to Selim et al (2007) “the correction factor was calculated as the difference between observed value of those beneficiaries with case-mix data and those with missing data and multiplied by the proportion of beneficiaries with missing case-mix data”.

Overall, after conducting a comparison of methods, Selim et al (2007) reported that based on their findings the results were similar and that continued use of the current HOS Case-Mix methodology was recommended.
7.1.3 Pros and Cons of the HOS Case-Mix methodology

An attractive aspect of the HOS Case-Mix methodology is the fact that it not only allows for deaths to be incorporated in health status, but it also provides a separate model for death. Thus, allowing the user of the method to adjust for covariates associated with death separately to those associated with changes in health status (Kazis et al 2007).

The main concern Kazis et al (2007) has with this methodology is that a negative focus on mental outcomes may not be fairly judged if the negative MCS outcome entails death. Thus users should study MCS and PCS outcomes with and without death included to get a greater understanding of what is occurring (Kazis et al 2007).

7.1.4 HOS Case-Mix methodology replicated by RTI group

The HOS Case-Mix methodology was replicated by a group known as the Research Triangle Institute (RTI). Trisolini et al (2005) replicated the method using their Medicare Fee-For-Service (FFS) beneficiaries data (random sample of N=6547) by applying a SAS software program that the HAL group provided (Trisolini et al 2005). The HAL Case-Mix method used Medicare and Choice Beneficiaries data and their findings are described in several reports (Ware et al 2001; Rogers et al 2004) and an earlier version of the HAL methodology was applied in the MOS (Ware et al 1996b; Trisolini et al 2005).

According to Trisolini et al (2005), the HAL Case-Mix approach assigns respondents to one of 3 categories, depending on whether they were “better,” “the same,” or “worse” at follow-up compared with baseline. Respondents were assigned to categories based on whether their change in PCS (or MCS) fell within ±2 times the standard error of measurement for a single observation. Trisolini et al (2005) and Selim et al (2007) used the same cut-off as Rogers et al (2004) and Ware et al (1996b) for defining a change in
PCS as better. The cut-off value was 5.66 points for PCS and 6.72 points for change in MCS.

7.1.5 Adaptation of the alternative HOS Case-Mix methodology

Although Selim et al (2007) recommended the continued use of the current HOS Case-Mix methodology, their method has provided an opportunity to study the accountability of deaths in longitudinal data analysis. This chapter will explore an adaptation of the alternative approach by Selim et. al (2007) by extending this approach. Selim et al (2007) made mention of the use of one single logistic regression model for deaths and the use of correction factor for missing covariates instead of the Rogers et al (2004) original cascading series of eight logistic regression models.

In this chapter the Selim et al (2007) alternative approach will be adapted where similar covariates will be considered. However, rather than using their idea of a correction factor, multiple imputation will be used at different stages throughout the adapted methodology. Also, as in Chapter 4, the inclusion of zeros for deaths will be applied at each of the follow-up surveys.
7.2 **AIM AND OBJECTIVES**

The aim of this Chapter is to assess the effects of using the adapted HOS methodology (also known as the HAL Approach) to account for deaths in a longitudinal study. There are five objectives for this Chapter:

1. Determine an effective ALSWH cut-off point for Same or Better Vs Worse using logistic regression models;
2. Determine whether other cut-off points for Better Vs Same Vs Worse would be more effective using ordered logistic regression models;
3. Compare the Rogers et al cut-off (5.66) to a ALSWH cut-off;
4. Conduct a longitudinal analysis using the probability of being alive and Same or Better health in three years and account for deaths (imputing a value of zero);
5. Use longitudinal multiple imputation to impute missing covariates and outcomes at each Survey.
7.3 METHODS

7.3.1 Study sample

The study sample is the older cohort of the Australian Longitudinal Study on Women. The data analysed included Surveys 1-4. See Chapter 5 for a description of the sample.

7.3.2 Definitions of Variables of Interest

The following sections will discuss the variables used in the statistical analysis. There will also be descriptions of the development of new variables of interest.

7.3.2.1 Outcome Measures

The primary outcome was developed using the conditional probability of two secondary outcomes. The secondary outcomes are probabilities and are discussed in the following sections. The primary outcome known as the probability of being Alive and in the Same or Better health is based on earlier work by the HAL group (Rogers et al 2004) and then used by Trisolini et al (2005) and Selim et al (2007). The following equation represents the conditional probability multiplied by 100.

\[
\text{Probability of being Alive and in Same or Better Health in 3 years} = [ (Pr_{\text{Alive}}) \times Pr_{\text{same\_better}} ] \times 100 \tag{equation 7.1}
\]

Where,

\( Pr_{\text{same\_better}} \) = probability of being in Same or Better health in 3 years (secondary outcome)

\( Pr_{\text{Alive}} \) = probability of being Alive in 3 years (secondary outcome)

Previously, Rogers et al (2004), Trisolini et al (2005) and Selim et al (2007) considered a change over two years. A combination of the two approaches, and an expansion to a change in health over three years was used here.
Secondary Outcomes

7.3.2.2 Binary variable for Death required for transformation equation
In order to calculate the predicted probability of death in 3 years a binary variable for
death was generated. Participants who had died between Survey 1 and 2 were coded
as (1) for dead, whilst those who were known to be ongoing participants in the
longitudinal study were coded as zero. This binary outcome was then used in a multiple
logistic regression. Further details of the application of this process follow in later
sections.

7.3.2.3 Classification of Same or Better or Worse
Using the original definition outlined by Rogers et al (2004), ALSWH participants were
categorised into the three groups: Better, Same or Worse, based on their change in
PCS from Survey 1 to Survey 2. Note that the focus in this Chapter is the
consideration of PCS only and not MCS. In order to generate a transformation equation
to use for predicting the probability of being “healthy” in 3 years a binary variable was
constructed which combined the Same and Better categories. Other combinations of
the three categories that were tried are discussed in Section 7.3.9.

a cut-off value for the ALSWH data was established. Selim et al (2007) considered the
cut-off points for the operational definition of “the same or better” as two standard
errors of measurement (SEM) for a single score or 1.414 standard errors of change.
They then defined “the Same or Better” group mean change in PCS to be –5.66 points
or higher (Selim et al 2007; Selim et al 2007b). In the adapted HAL approach Trisolini
et al (2005) also used the same cut-off value (5.66) definition established by Rogers et
In order to replicate this approach a cut-off value for the ASLWH data was determined and compared to the Rogers et al cut-off. The cut-off value for change in PCS in the ALSWH data was calculated using the 2 standard errors of measurement (SEM) method.

### 7.3.2.4 Definition of Standard Error of Measurement

To establish a cut-off value (change in PCS) using ALSWH data a definition of Standard Error Measurement (SEM) was required.

The Standard Error of Measurement can be calculated using the equation from Rejas et al (2008),

\[
SEM = \sigma \times \sqrt{1-r}
\]

(equation 7.2)

Where, \( \sigma \) = Standard deviation of the measurement

\( r \) = reliability coefficient

Consideration was given as to whether 2 SEMs or 1 SEM should be used in establishing the ALSWH cut-off. In a personal email correspondence in July 2009 with Dr Bill Rogers, he suggested

“if the cut-off is lowered to 1 SEM of change then the signal and error are both raised, but the efficiency stays about the same. Because of long tails, the efficiency of measuring changes in points is also not much better than using a cut-off, but it's hard to argue that big declines in health are unimportant. So I agree that 1 SEM would be just fine, but why change it?” (Rogers 2009)

Therefore, from this correspondence with Dr Rogers the use of 2 SEMs was used for the ALSWH study.

The 2 SEM was estimated for the ALSWH PCS by multiplying the baseline standard deviation of the ALSWH PCS by the square root of one minus its reliability coefficient (Streiner & Norman 1998; Selim et al 2007; Rejas et al 2008). This result was then
doubled to achieve 2 SEM. The scale reliability was estimated using Cronbach’s alpha coefficient as suggested by Rejas et al (2008).

7.3.3 Rogers et al’s Cascading HOS Approach to Modelling Deaths

Rogers et al (2004) used a series of eight logistic regression models to estimate the probability of death for a beneficiary in two years. These estimated probabilities were assigned based on a case-mix profile. That is, depending on the number of missing covariates for an individual then a beneficiary was given a probability associated with that logistic regression model which assigned the same probabilities as a similar profile. This estimated probability for each individual would then be used later to calculate the probability of being alive and in the Same or Better health in 2 years. This “cascading” case-mix approach by Rogers et al (2004) may be considered similar to a form of mean substitution.

7.3.3.1 Selim et al’s Demographic covariates

According to Selim et al (2007), they identified from prior studies by Wilson and Cleary (1995) and Hornbrook and Goodman (1996) three domains of case-mix characteristics: socio-demographics, co-morbidities and baseline health status. For the socio-demographic variables, Selim et al (2007) chose, age, gender, race/ethnicity, marital status, education, income and Medicaid eligibility, since they were previously theorised and shown to be using the VR-12 (Williams 1996; Selim et al 2007). Age was considered as a continuous variable, while race/ethnicity was grouped as Whites, African-Americans, Hispanics and others. Whereas marital status, education and income were all dichotomized, such as married or not, at least 12 years of education vs. less than 12 years and less than $20000 vs. equal or more than $20000 (Selim et al 2007).
7.3.3.2 Co-morbidities and Cancer covariates

Co-morbidities considered by Selim et al (2007) included the following self-reported diagnoses: acute myocardial infarction, coronary artery disease, congestive heart failure, other heart conditions, stroke, diabetes, chronic obstructive pulmonary disease (COPD), asthma, cancer (other than skin cancer), gastrointestinal disorder, arthritis (hip and hand) and sciatica.


7.3.4 Extension of the alternative HOS Case-Mix Approach to a longitudinal study

The HOS Case-Mix approach used the change in PCS (Same or Better) and has only been applied for two time points: baseline and a two year follow-up. At this stage the approach has not been extended beyond this time period. The ALSWH data provides an opportunity to build on the work of Rogers et al (2004); Trisolini et al (2005) and Selim et al (2007) to develop an approach that can be applied to longitudinal studies when more time points are being analysed and when the time between Surveys is greater than two years.
7.3.5 Generating the Outcome as a conditional probability which accounts for deaths

The Approach is a four stage process:

**Stage 1** – Calculate the probability of being Alive in 3 years

**Stage 2** – Calculate the probability of being in Same or Better health in 3 years

**Stage 3** – Calculate the conditional probability of being Alive and in the Same or Better health in 3 years, given that the participant has had a probability of being Alive and a probability of being in the Same or Better health calculated earlier.

**Stage 4** – Calculate the probability of being Alive and in the Same or Better health in 3 years which includes deaths.

**Stage One:**

7.3.5.1 Modelling the Probability of Death in three years

Initially, a binary variable known as death was generated to indicate who had died between Survey 1 and Survey 2. A participant was considered alive if they had responded Survey 2. This was identified in the data by a categorical variable called attrition. Respondents were coded as 1. The reasons for attrition were coded as such:

1 = Responder; 2 = Non-responder; 3 = Unable to be contacted; 4 = Too Frail to continue; 5 = Withdrawn from study and 6 = Deceased before survey.

The variable for death is the basis for the development of a death transformation equation. The transformation equation was used to determine the probability of death in 3 years at the remaining Surveys.
7.3.5.2 ALSWH Covariates used for the logistic regression Death model

The modelling of death by the next survey was how the original approach was conducted. Eventually, the resulting probability of death, for each participant, becomes a probability of being Alive at the next survey. This will be discussed later. Based upon the suggestion from Selim et al (2007) certain covariates were used in the initial logistic regression to determine the regression coefficient estimates. The regression estimates were then used in a transformation equation to generate the predicted probability of death based on the substitution of the individual participants’ covariates at each Survey. Some of the covariates used were age, PCS and MCS, area of residence and education status. Other covariates used are described below in the following section.

7.3.5.3 Co-morbidities

A limitation of the ALSWH data was that some questions changed over the course of the study. For the purpose of the analysis in this chapter and to remain true to the HOS approach the prevalence the following chronic diseases were used: diabetes, heart disease, hypertension, osteoporosis, stroke and asthma.

7.3.5.4 Cancers

In terms of the prevalence of cancers the only consistent cancer listed in the ALSWH data was “skin cancer” and “other cancer” from Survey 2 through to Survey 4. Thus to determine an approximate prevalence of cancer, a new variable (cancer) was generated at each Survey.

At Survey 1 a participant was identified as “Yes, has cancer” if she answered “Yes” to being diagnosed as having breast, cervical, lung or bowel cancer.

At Survey 2 a participant was identified as having at least one cancer if she answered “Yes” to “other cancer” or had been identified at baseline as “Yes”.
At Survey 3 cancer prevalence was calculated by answering “Yes” to “other cancer” at Survey 3 or being identified at Survey 1 or Survey 2 as “Yes”. “Skin cancer” was not included since the type of skin cancer was not known. Likewise, at Survey 4 a similar classification was conducted.

7.3.5.5 ALSWH Death Model using Survey 1 and Survey 2 data

The complex HOS Case-Mix methodology (Rogers et al 2004) has been reduced to one logistic regression for death similar to the alternative Case-Mix method (Selim et al 2007). However, a further adaptation and an extension of the alternative Case-Mix methodology is the use of longitudinal multiple imputation of missing covariates.

As mentioned earlier, in the HOS Case-Mix approach the expected probability of death was calculated for each participant using a case-mix approach based on the number of missing baseline covariates. There were eight Case-Mix logistic regression models considered and each participant received an expected probability based on a particular logistic regression model. Rather than apply this approach, multiple imputation was conducted to impute for missing baseline covariates. The resulting estimates from the combined multiple complete datasets were then used to generate the probability of death in three years. This MI approach will allow for participants with missing covariates at baseline to be included, in much the same way as the cascading arrangement of logistic regression models for deaths in the HOS case-mix approach. In order to produce probabilities for as many participants as possible, longitudinal multiple imputation was applied to impute missing follow-up covariates and a more detailed explanation is discussed in the results section.

Using the ALSWH data to predict the probability of death in three years, multiple logistic regression was applied where death was the outcome variable. The covariates used were: PCS and MCS at Survey1, baseline age, area of residence (urban / non-
urban), level of education (tertiary: yes/no), living alone (yes/no), managing on income (yes/no), at least one chronic disease (heart disease, stroke, diabetes, hypertension, asthma or osteoporosis) and at least one cancer (other than skin cancer).

The logistic regression model used for the ALSWH Approach to predict death:

\[
\text{logit (death)} = \alpha + \beta_1 \text{PCS}_i + \beta_2 \text{MCS}_i + \beta_3 \text{age}_i + \beta_4 \text{area}_i + \beta_5 \text{educ}_i + \beta_6 \text{alone}_i \\
+ \beta_7 \text{manage}_i + \beta_8 \text{cmor}_i + \beta_9 \text{cancer}_i
\]

\text{(equation 7.3)}

The \(\alpha\) intercept and \(\beta\) coefficients are used in a similar way to equation 4.3 in Section 4.6.2 to generate the predicted probability of death in 3 years for each follow-up Survey. The estimates from the initial logistic regression are held constant and the individual participant’s data are applied to calculate the individual’s predicted probability of death at the next Survey in 3 years. Comparisons were made between the \(\beta\) coefficients when multiple imputation was not considered and are reported in Section 7.5.1. This transformation approach is similar to that used in Chapter 5 to produce the APCT scores and was an extension of the earlier work using PCT by Diehr et al (2003).
7.3.5.6 The Probability of Death in three years

An attempt to replicate the HOS Case-Mix modelling approach was considered (see Appendix 7.1), however the complex nature of the Case-Mix modelling for death is a deterrent for further exploration beyond one follow-up. That is, the application at future follow-up Surveys could be difficult for an end user. Below is the transformation equation (7.5) that was considered.

\[
\text{Probability of Death in three years } = \frac{e^{\alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{MCS}_1 + \beta_3 \text{age}_1 + \beta_4 \text{area}_1 + \beta_5 \text{educ}_1 + \beta_6 \text{alone}_1 + \beta_7 \text{manage}_1 + \beta_8 \text{cmorb}_1 + \beta_9 \text{cancer}_1}}{1 + e^{\alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{MCS}_1 + \beta_3 \text{age}_1 + \beta_4 \text{area}_1 + \beta_5 \text{educ}_1 + \beta_6 \text{alone}_1 + \beta_7 \text{manage}_1 + \beta_8 \text{cmorb}_1 + \beta_9 \text{cancer}_1}}
\]

(equation 7.4)

Taking the Selim et al (2007) alternative Case-Mix approach one step further, multiple imputation was used to generate the “averaged” regression coefficients. An explanation of the “averaged” regression coefficients will follow in a later section. Earlier work by the HAL and RTI groups modelled death using only baseline and two year follow-up data from different populations to the ALSWH population. This Chapter now extends the Selim et al’s alternative Case-Mix approach by using the ALSWH Surveys 1-4 data which have 3 year time intervals.
Stage Two

7.3.6  Modelling the Probability of Same or Better Health in three years

Two cut-off values to model the predicted probability of Same or Better versus Worse health in three years were used. The recommended change in PCS cut-off used by Rogers et al (2004) and Selim et al (2007) was considered along with a cut-off based on the ALSWH data. These cut-off results were compared but the comparisons have not been reported.

7.3.6.1 Creating a cut-off for Same or Better using ALSWH data

As previously mentioned the ALSWH researchers developed their own Australian population-based calculations for the PCS and MCS. Since this was the case, a cut-off for change in PCS from baseline to follow-up was derived using the ALSWH PCS. To determine a change in PCS cut-off the use of the Standard Error of Measurement (SEM) method was deemed appropriate. Using 2 SEMs was considered to be reasonable to determine the cut-off value. To calculate 2 SEM the following analysis was conducted using equation 7.6 from Rejas et al (2008).

The equation requires the calculation of the reliability coefficient ($r$). This can be considered equivalent to the scale reliability coefficient, also known as Cronbach’s Alpha. Using the ALSWH $PCS_2$ and $PCS_1$ the reliability coefficient was determined as 0.8109 (Stata 10.1).
The SEM equation also required the standard deviation of baseline PCS. The standard deviation was found to be approximately 9.95.

Hence,

\[
2 \text{ SEM} = 2 \times \sigma \times \sqrt{1 - r} \\
= 2 \times 9.95137 \times \sqrt{1 - 0.8109} \\
= 8.6548 \\
\text{(equation 7.5)}
\]

where \( \sigma \) = standard deviation of baseline PCS  
\( r \) = reliability coefficient  
SD = standard deviation

The resulting cut-off value was calculated as 8.65 and will be referred to as the ALSWH cut-off. For the remainder of this chapter the ALSWH cut-off will be considered when describing the methodology and discussed in the results. Some comparisons will be shown for the Rogers et al cut-off, however further applications of the Rogers et al cut-off will not be shown. This is to avoid confusion in the application of the adapted HAL methodology.

### 7.3.6.2 Steps to Creating the Probability of being in Same or Better Health

Once the ALSWH cut-off was established, the next step was to generate the difference in PCS from Survey1 and Survey2. Such that:

\[
\text{PCS}_{\text{diff}} = \text{PCS}_2 - \text{PCS}_1
\]

where a negative change suggests a decline in PCS.

\text{(equation 7.6)}

The binary variable Same or Better was coded as 1, if the difference in \( \text{PCS}_{\text{diff}} > -8.65 \). The variable Same or Better was coded as 0 (Worse), if \( \text{PCS}_{\text{diff}} \leq -8.65 \).

This variable is then used as the outcome variable in the logistic regression. Those participants who did not respond or could not be contacted at Survey 2 were excluded.
7.3.6.3 Covariates for Same or Better logistic regression model

To comply with similar guidelines suggested by Selim et al (2007) the following continuous and binary covariates were used in the logistic regression models:

PCS, age at Survey, area of residence (urban/non-urban), level of education (tertiary: yes/no), living alone (yes/no), managing on income (yes/no). Selim et al (2007) had suggested the use of more covariates, however due to limitations discussed earlier in Section 7.3.5.3, only these variables were available in the ALSWH data.

\[
\text{logit}(\text{Better/Same}) = \alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{age}_1 + \beta_3 \text{area}_1 + \beta_4 \text{educ}_1 + \beta_5 \text{alone}_1 + \beta_6 \text{manage}_1
\]  

(equation 7.7)

Similar to the modelling of death above, multiple imputation was used to impute for missing covariates at baseline. The averaged regression estimates generated from the Same or Better logistic regression model from 5 imputed data sets were then compared to the regression estimates when multiple imputation was not considered.

Imputing for missing PCS values at Survey 1 and Survey 2 were part of the multiple imputation. Thus the PCS\text{diff} variable was maximized so that those women who had participated at Survey 2 but had a missing PCS could be considered in the development of the transformation equation 7.8. As mentioned previously the averaged regression results were compared to when multiple imputation had not occurred for baseline follow-up PCS and other baseline covariates.
7.3.6.4 Transformations to the Probability of Same or Better Health

The transformation equation 7.8 referred to as the predicted probability of being in Same or Better health in 3 years was then extended. So that the predicted probability could be extended for longitudinal analysis, new variables were generated using the covariates from each Survey 1, 2, 3 and 4 and the averaged coefficients. The average coefficients are from the multiple logistic regression using the combined multiple imputation datasets (M=5).

The actual transformation equation is a similar structure used in Chapter 4.

\[
\text{prob}(\text{Same \ Better}) = \frac{e^{\alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{age}_1 + \beta_3 \text{area}_1 + \beta_4 \text{educ}_1 + \beta_5 \text{alone}_1 + \beta_6 \text{manage}_1}}{1 + e^{\alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{age}_1 + \beta_3 \text{area}_1 + \beta_4 \text{educ}_1 + \beta_5 \text{alone}_1 + \beta_6 \text{manage}_1}}
\]  
\text{Equation 7.8}

The transformation used the same constant averaged regression estimates at each Survey. A probability was generated using individual covariate data at each Survey. Note that the transformation is a combination of time varying covariates (PCS, age, living alone and manage on income) and two baseline time independent covariates (area of residence and level of education).

\[
\begin{align*}
\text{prob}(\text{Same \ Better}1) &= \frac{e^{\alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{age}_1 + \beta_3 \text{area}_1 + \beta_4 \text{educ}_1 + \beta_5 \text{alone}_1 + \beta_6 \text{manage}_1}}{1 + e^{\alpha + \beta_1 \text{PCS}_1 + \beta_2 \text{age}_1 + \beta_3 \text{area}_1 + \beta_4 \text{educ}_1 + \beta_5 \text{alone}_1 + \beta_6 \text{manage}_1}} \\
\text{prob}(\text{Same \ Better}2) &= \frac{e^{\alpha + \beta_1 \text{PCS}_2 + \beta_2 \text{age}_2 + \beta_3 \text{area}_2 + \beta_4 \text{educ}_2 + \beta_5 \text{alone}_2 + \beta_6 \text{manage}_2}}{1 + e^{\alpha + \beta_1 \text{PCS}_2 + \beta_2 \text{age}_2 + \beta_3 \text{area}_2 + \beta_4 \text{educ}_2 + \beta_5 \text{alone}_2 + \beta_6 \text{manage}_2}} \\
\text{prob}(\text{Same \ Better}3) &= \frac{e^{\alpha + \beta_1 \text{PCS}_3 + \beta_2 \text{age}_3 + \beta_3 \text{area}_3 + \beta_4 \text{educ}_3 + \beta_5 \text{alone}_3 + \beta_6 \text{manage}_3}}{1 + e^{\alpha + \beta_1 \text{PCS}_3 + \beta_2 \text{age}_3 + \beta_3 \text{area}_3 + \beta_4 \text{educ}_3 + \beta_5 \text{alone}_3 + \beta_6 \text{manage}_3}}
\end{align*}
\]
\[
\text{prob(Same \_ Better)}4 = \frac{e^{\alpha + \beta_1 \text{PCS}_4 + \beta_2 \text{age}_4 + \beta_3 \text{area}_4 + \beta_4 \text{educ}_4 + \beta_5 \text{alone}_4 + \beta_6 \text{manage}_4}}{1 + e^{\alpha + \beta_1 \text{PCS}_4 + \beta_2 \text{age}_4 + \beta_3 \text{area}_4 + \beta_4 \text{educ}_4 + \beta_5 \text{alone}_4 + \beta_6 \text{manage}_4}}
\]

Note: covariate subscripts refer to the Survey number.

Note: each equation represents the predicted probability of being in the Same or Better health at the next Survey in 3 years. Therefore, the \text{prob(Same\_Better)}4 represents the predicted (expected) probability of being in Same or Better health at the time of Survey 5.

Stage Three:

7.3.7 The Conditional Probability of Being Alive and in certain Health states in three years.

The following will outline the final stage required to calculate the conditional probabilities. The conditional probabilities will involve multiplying the predicted probability of being alive and the predicted probability of being in Same or Better health in 3 years.

7.3.7.1 Probability of Being Alive and in Same or Better Health (APASB)

The final step in this method was to create the conditional probability that is the “Probability of being Alive and Same or Better health in three years”. To generate this new outcome variable the two earlier probabilities (death and Same or Better) are combined by multiplication. That is, the probability of Alive (1- probability of death) x probability of Same or Better.

\[
\text{Probability of being Alive and in Same or Better Health in 3 years} = \left[ (1 - \Pr_{\text{death}}) \times \Pr_{\text{same\_better}} \right] \times 100
\]

(equation 7.10)
The conditional probability referred to as the “Probability of being Alive and in the Same or Better health in 3 years” will be given the acronym APASB when reference is made to the ALSWH cut-off (8.65) and PASB when the Rogers et al cut-off (5.66) is referenced. Thus new variables for APASB were created at each time point t (for t=1 to 4) by multiplying the two earlier predicted probabilities for death and Same or Better for each participant.

The following equation 7.11, represent these new variables for APASB:

\[
\text{Probability of being Alive and in Same or Better Health in 3 years (APASB}_t = {} [1 - \Pr(\text{Death})_t \times \Pr(\text{Same or Better})_t] \times 100
\]

where t= 1 to 4.  

(equation 7.11)

**Stage Four:**

### 7.3.8 Using the new outcome APASB to Account for Deaths

The new outcome variable generated above represents the predicted probability of being alive and in the Same or Better health in three years. Noting that the values for APASB have been multiplied by 100 to represent a percentage (and can be thought of as a “chance in 100”). Therefore, an APASB score of 95, suggests a participant would have a 95% chance of being Alive and in the Same or Better health in 3 years. Thus if the value was a score of 5 then it would be expected that this participant would have a very low chance of being Alive and in the Same or Better health in 3 years. As in Chapter 4, these probabilities/percentages are considered as a score out of 100 for health status in the future (APCT).

Based on the notion of probability, a participant who has died could then be scored as a zero. Therefore when a participant has died, the predicted probability of being Alive and in the Same or Better health in three years will be considered as zero. Thus those
participants who had died receive a zero score for each time point after their death. This is the same concept applied in Chapter 4 for the Diehr et al (2003) method.

New variables were generated which allowed for the inclusion of zero for death. Thus the APASB is first calculated using the above method for each Survey then another variable duplicates the original data for APASB. Those who died will be missing a value and so the value of zero is imputed for the decedents. Thus, the new variable is called APASBD. Since no one was dead at Survey1 then APASB₁ = APASBD₁

7.3.9 Using ordered logistic regression to consider Better Vs Same Vs Worse

The health outcome variable (Same or Better) has been based on cut-off values. Further discussion on this topic involves whether a different combination of health status should be considered. That is, should other possibilities be considered such as combining Same and Worse rather than Same and Better. Another approach is to use ordered logistic regression. This was considered since it allows for the use of an ordinal categorical outcome rather than binary.

Thus, the ordinal outcome variable for health status considered the following categories:

Better = 1 if PCSₜdiff ≥ 8.65
Same = 0 if PCSₜdiff > -8.65 & PCSₜdiff < 8.65
Worse= -1 if PCSₜdiff ≤ -8.65

Some results from ordered logistic regression analyses using the ALSWH cut-off are reported in this chapter. This approach was also applied to the Rogers et al cut-off (5.66), however the results are not reported.
7.4 **Statistical Analysis**

In this chapter, cross-sectional comparisons of diabetes status are made at Surveys1-4. The main outcome of interest considered and reported are the APASB and APASBD. Longitudinal data analyses were conducted and details of these analyses are discussed in the following section. All other outcomes based on the different cut-off values are not reported to avoid confusion.

7.4.1 **Cross-sectional Analysis**

The cross-sectional analyses involved the use of one-way ANOVAs with Bonferroni multiple comparisons to compare the mean differences in APASBD between diabetes classifications (No diabetes, New cases and Existing cases). This same analysis was also used for APASB. Paired t-tests are used to determine whether significant changes occur between Surveys. These paired t-tests were conducted by their diabetes classifications. The cross-sectional analyses methods are viewed as a preliminary step towards assessing the impact of longitudinal data analysis methods.

7.4.2 **Longitudinal Analysis**

7.4.2.1 **Data sample analysed**

To be consistent with all the longitudinal analyses in this thesis, not all data were used in the analyses in this chapter. Although multiple imputation was used to impute values for covariates not all participants were considered. For example, the status for diabetes was not known for a proportion of participants, so they were dropped from the analyses. Since it was not clear which group they belonged to, it was decided to remove them rather than impute a category of diabetes for these unknowns.
7.4.2.2 Random Effects Models

Longitudinal data analysis using random effect models was conducted to examine the effects of using the new outcome APASBD. As with Chapter 4, a case study using diabetes was undertaken. The longitudinal models included random intercepts and reasons for this are discussed in the results. The random effect models included covariates that were both time independent (area of residence and smoking status) and time dependent (age and co-morbidities). The two models considered the inclusion and exclusion of deaths. Similar analysis was demonstrated in chapter 5. One difference in this Chapter is the use of time dependent covariates. Previously, age was considered as baseline age (0 = 69-72 years old and 1 = 73-76 years old) and the presence of co-morbidities (Yes/No) were determined retrospectively from Survey 4.

\[
\text{APASB}_i = \beta_0 + \beta_1\text{Time} + \beta_2\text{Age} + \beta_3\text{Area} + \beta_4\text{Comorb} + \beta_5\text{Smoke} + \beta_6\text{Diab}_1 + \beta_7\text{Diab}_2 + \varepsilon_{it}
\]  
(equation 7.12)

\[
\text{APASBD}_i = \beta_0 + \beta_1\text{Time} + \beta_2\text{Age} + \beta_3\text{Area} + \beta_4\text{Comorb} + \beta_5\text{Smoke} + \beta_6\text{Diab}_1 + \beta_7\text{Diab}_2 + \varepsilon_{it}
\]  
(equation 7.13)

where Diab$_1$ = new cases of Diabetes and

\[
\text{Diab}_2 = \text{existing cases of Diabetes}
\]

\[
\varepsilon_{it} = \text{error term}
\]
7.4.2.3 Time and Diabetes Status Interaction

An interaction term was applied to determine whether the difference between existing diabetes and no diabetes groups was significant over time. Adding an interaction term is an effective method to determine longitudinal change when using random intercept models. In this analysis, the time variable would be considered as a continuous variable. All statistical analyses and modeling in this chapter was conducted using Stata10.1.

The general form of the random effects model with an interaction is:

\[ Y_{it} = \beta_0i + \beta_1i \text{time} + \beta_2i \text{Diab\_retro} + \beta_3i \text{time} \times \text{Diab\_retro} + \epsilon_{it} \]

(equation 7.14)

where \( Y_{it} \) represents APASB or APASBD at the 4 time points.

\( \beta_2i \text{Diab\_retro} \) = the effect size difference between the no diabetes group and existing cases

\( \beta_3i \text{time} \times \text{Diab\_retro} \) = interaction between diabetes groups and time

Each individual was considered to have a random intercept. The variable IDalias is used to identify individuals and the inclusion of random intercepts is acknowledged in the Stata10.1 code as IDalias:

\[
\text{xi: xtmixed APASB i.diab_retro*time || IDalias:}
\]

where Diab_retro x time = interaction between diabetes groups and time.

Note: When i.diab_retro*time is used Stata will automatically include the variables Diab_retro, time and interaction terms.

A model without an interaction between Diabetes status and time seen below was also reported.

\[
\text{xi: xtmixed APASB i.diab_retro time || _IDalias:}
\]
7.4.3 Imputing for Missing Outcomes and time dependent and independent Covariates

Missing covariates at each Survey have major impacts on the generation of APASB. Thus to improve the number of APASB scores generated, multiple imputation was used in the two stage process of calculating the conditional probability of being Alive and Same or Better in 3 years. To increase the sample size for longitudinal analyses both the multiple imputation of missing covariates and the missing outcomes (APASBD) were considered. Note that a simple longitudinal imputation was applied when the participant was known to have died by “carrying forward” a zero (APASBD).

7.4.3.1 Multiple Imputation for missing covariates

As has been previously mentioned in Chapter 5, consideration was given to the missingness of the data. The multiple imputation approach used was Stata’s method of multiple multivariate data imputation using Chained Equations. The original Stata method was named Multivariate data Imputation using Chained Equations (mice). Further developments by Royston (2005) towards a more sophisticated multiple imputation approach changed the command name to ice. This method assumes MAR and can be applied more simply for the imputation of covariates.

Previously, in Chapter 5 only the longitudinal multiple imputation of the outcome variable APCTD was considered using the WinBUGS approach. This approach allowed for the assumptions of MAR and MNAR. As the WinBUGS approach is more difficult to apply to missing covariates, the missing data will be assumed to be MAR and only Stata’s ice approach will be used.
7.4.4 The Stata Code used for Longitudinal Multiple Imputation

The structure for the ice command is relatively simple. In an attempt to treat the process of multiple imputation in a longitudinal framework, the covariates were organised as shown below. Firstly, all the variables that are to be considered in the M=5 datasets are listed. Then using the cmd command: all variables which will be imputed are assigned a regression process: linear regression for a continuous variable or logistic regression for a binary variable. At the next step: for the eq command, the longitudinal nature of the dataset can be implemented as shown below. The following is one example of how the longitudinal multiple imputation could be performed. The user can also simply allow Stata 10.1 to conduct both the cmd and eq procedures. In this chapter, it was decided to take greater control of the procedures to avoid possible errors that can occur with categorical variables. If variables are included which do not have missing data then the variable will simply remain unchanged in the m=5 datasets.

The original dataset remains in the newly structured multi-levelled dataset. New variables are automatically generated, one to identify individual participants (_mi) and one to identify the new dataset (_mj). The multiple datasets including the original (_mj=0) are then saved automatically using saving and m(5) instructs Stata to conduct the process “m” times - in this case 5 times. Using regressions such as mlogit or ologit in this ice procedure can be problematic with Stata 10.1, so to avoid possible errors categorical variables were confined to binary variables. Note as mentioned earlier diabetes status (diab_retro: (1) existing cases, (2) new cases and (3) no diabetes) was considered to have no missing data since unidentified participants were removed before multiple imputation.
7.4.5 Combining MI datasets to be used in a Longitudinal Analysis

As mentioned, the Stata10.1 command known as *ice* produces multiple complete datasets. In order to summarises these datasets and conduct a longitudinal analysis of the “m” complete datasets Royston (2005) and others developed the following method *mim*:(manipulating multiply imputed datasets). The *mim*: command combines the imputed datasets and allows for longitudinal modelling with “averaged regression estimates”. The following code is an example using the ALSWH data, the variable *IDalias* identifies the individual participants’ intercepts:

```
xi: mim: xtmixed APASB i.diab_retro time || _IDalias:
```

where

```
xmixed
```

represents random effects modelling of longitudinal data.

As in Section 7.4.2.3, an interaction term was applied to determine whether the difference between existing diabetes and no diabetes groups was statistically significant over time. The following code includes the interaction between time and diabetes status:

```
xi: mim: xtmixed APASB i.diab_retro*time || _IDalias:
```
7.5 RESULTS

Earlier in the methods' section a number of strategies (i.e. ordered logistic regression and case mix models) were suggested however not all results are presented. The approach presented and discussed involves the use of the ALSWH cut-off (8.65). At times reference and comparisons to the Rogers et al cut-off (5.66) are made. Some results using ordered logistic regression are shown later.

Table 7.1 shows the proportion of missing covariates (which include PCS) at each survey. Table 7.1 highlights the extent of the amount of missing data for each of the covariates. The missing covariates were considered as time dependent (changing over time) and time independent (remain unchanged from baseline Survey). The proportions are based on the original sample size (N=12432). These covariates were considered in the modelling of death and for Same or Better. Results from an ALSWH case-mix attempt using nine logistic regressions models where different combinations of covariates were considered are shown in Appendix 7.1. The example in Appendix 7.1 was merely an attempt at using the HOS case-mix methodology and was never meant to replicate the original method, one reason being that the covariates used were not similar. The objective in this chapter is to move towards the adapted alternative Selim et al's HOS case-mix approach and the use of multiple imputation to account for missing covariates.
From Table 7.1 it appears that by Survey 4, approximately 50% of the original cohort is missing PCS and MCS scores. Reasons for missing PCS or MCS may include:

i. Non-responders

ii. unable to be contacted;

iii. withdrawn from the study before Survey 2, 3 or 4;

iv. incomplete survey items (resulting in no PCS or MCS score);

v. death.

In all the covariates the proportion of missing data increases over time. In many covariates the proportion of missing is considered high (>20%). There are many reasons (that have been mentioned earlier for the missing data), some missing data would be considered as intermittent missing (see Section 3.1.2.1).

Table 7.1  Missing data - proportion of missing covariates at each survey for original cohort

<table>
<thead>
<tr>
<th>Covariates</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% missing</td>
<td>% missing</td>
<td>% missing</td>
<td>% missing</td>
</tr>
<tr>
<td>PCS</td>
<td>12.7</td>
<td>23.1</td>
<td>42.6</td>
<td>50.1</td>
</tr>
<tr>
<td>MCS</td>
<td>12.7</td>
<td>23.1</td>
<td>42.6</td>
<td>50.1</td>
</tr>
<tr>
<td>Manage on Income</td>
<td>2.1</td>
<td>24.4</td>
<td>31.0</td>
<td>43.2</td>
</tr>
<tr>
<td>Living Alone</td>
<td>6.0</td>
<td>16.6</td>
<td>31.0</td>
<td>43.0</td>
</tr>
<tr>
<td>Co-Morbidities</td>
<td>0.8</td>
<td>4.8</td>
<td>7.9</td>
<td>10.6</td>
</tr>
<tr>
<td>Cancer</td>
<td>1.9</td>
<td>1.9</td>
<td>28.9</td>
<td>40.7</td>
</tr>
<tr>
<td>Age*</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Education (baseline)</td>
<td>5.3</td>
<td>5.3</td>
<td>5.3</td>
<td>5.3</td>
</tr>
<tr>
<td>Area (baseline)</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
<td>&lt;0.1</td>
</tr>
</tbody>
</table>

*N=12432

*Age was used from baseline 70-74 years. When later used in analyses 3 years were added at each survey after baseline age for each participant.
7.5.1 Stage 1: Modelling Death – Using the ALSWH Approach:

7.5.1.1 Logit (or Log Odds) of Death in 3 years to obtain β coefficients for Transformation Equation:

The ALSWH approach to modelling death was carried out and the results shown in this section use a single multiple logistic regression model for death and multiple imputation to impute for missing covariates. Table 7.2 shows the β coefficients from a logistic regression when only complete case data were used, that is, participants who had all baseline covariates and were known to have participated at Survey 2 or had died before Survey 2. The model shown in Table 7.2 (complete case) is to compare with the estimates for the multiple imputation model in Table 7.3.

The ALSWH Approach is an adaptation of the HOS case-mix method to account for missing covariates. The use of multiple imputation allows for imputed data sets to be used where all covariates are accounted for. The results shown in Table 7.3, are considered as the “averaged” regression coefficients from the imputed datasets. There were 5 complete datasets in all and by using the Stata’s mim: command the resulting logistic regression model estimates are equivalent to the average coefficient estimates of 5 possible logistic regression models. Note for this multiple imputation process only baseline covariates were imputed. The covariates used in the regression modelling are listed in Table 7.3.

The “averaged” coefficient estimates in Table 7.3 were then used to generate a predicted probability of death at each survey. The 95% confidence intervals shown in both Tables 7.2 and 7.3 are similar in width. The regression coefficient estimates from both Tables 7.2 and 7.3 are similar in most cases. Note the standard errors (SE) are all smaller in Table 7.3 due in part to the larger sample size. The estimates for education status and living alone are reversed in direction (+/-), however the Wald statistic p-values in both Tables for these two variables are not statistically significant at level of
0.05. Also, it appears that statistical significance has changed for age, area of residence and co-morbidities. In Table 7.3, age is statistically significant whereas area of residence is not statistically significant, while the co-morbidities variable is closer to statistical significance at a level of 5%. However, as has been mentioned previously in earlier Chapters if the level of significance is set at 0.001 (large sample size) then these changes are not statistically significant.

Multiple imputation is increasingly being recommended and applied in epidemiology to dealing with missing data (Spratt et al 2010) and so only the results shown in Table 7.3 for the “averaged” β coefficients are considered for further analysis. Another reason for the preferred use of multiple imputation “averaged” β coefficients is to consider an alternative to the Selim et al approach to missing data. The “averaged” β coefficients are used in the process of calculating the predicted conditional probabilities for being alive and in Same or Better health in 3 years.

### Table 7.2 No multiple imputation (N=7010)

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.072</td>
<td>0.006</td>
<td>-0.083 , -0.061</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.033</td>
<td>0.005</td>
<td>-0.043 , -0.023</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>0.066</td>
<td>0.037</td>
<td>-0.007 , 0.139</td>
<td>0.077</td>
</tr>
<tr>
<td>Area residence</td>
<td>-0.263</td>
<td>0.115</td>
<td>-0.489 , -0.038</td>
<td>0.022</td>
</tr>
<tr>
<td>education status</td>
<td>0.034</td>
<td>0.151</td>
<td>-0.263 , 0.330</td>
<td>0.824</td>
</tr>
<tr>
<td>live alone</td>
<td>0.030</td>
<td>0.115</td>
<td>-0.195 , 0.255</td>
<td>0.795</td>
</tr>
<tr>
<td>Manage income</td>
<td>0.066</td>
<td>0.121</td>
<td>-0.172 , 0.304</td>
<td>0.586</td>
</tr>
<tr>
<td>Co-morbidities</td>
<td>0.212</td>
<td>0.140</td>
<td>-0.062 , 0.487</td>
<td>0.129</td>
</tr>
<tr>
<td>Cancer</td>
<td>0.775</td>
<td>0.150</td>
<td>0.481 , 1.069</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>intercept</td>
<td>-2.820</td>
<td>2.739</td>
<td>-8.189 , 2.549</td>
<td>0.303</td>
</tr>
</tbody>
</table>
Table 7.3 Multiple imputation “averaged coefficients” (N=8790)

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>SE</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.062</td>
<td>0.005</td>
<td>-0.072 , -0.052</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.031</td>
<td>0.005</td>
<td>-0.041 , -0.020</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>age</td>
<td>0.065</td>
<td>0.032</td>
<td>0.002 , 0.128</td>
<td>0.044</td>
</tr>
<tr>
<td>Area residence</td>
<td>-0.162</td>
<td>0.097</td>
<td>-0.354 , 0.029</td>
<td>0.096</td>
</tr>
<tr>
<td>education status</td>
<td>-0.037</td>
<td>0.137</td>
<td>-0.307 , 0.232</td>
<td>0.785</td>
</tr>
<tr>
<td>live alone</td>
<td>-0.057</td>
<td>0.106</td>
<td>-0.266 , 0.153</td>
<td>0.596</td>
</tr>
<tr>
<td>Manage income</td>
<td>0.091</td>
<td>0.108</td>
<td>-0.121 , 0.303</td>
<td>0.400</td>
</tr>
<tr>
<td>Co-morbidities</td>
<td>0.216</td>
<td>0.118</td>
<td>-0.016 , 0.447</td>
<td>0.068</td>
</tr>
<tr>
<td>cancer</td>
<td>0.686</td>
<td>0.131</td>
<td>0.430 , 0.943</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>intercept</td>
<td>-3.241</td>
<td>2.400</td>
<td>7.954 , 1.472</td>
<td>0.177</td>
</tr>
</tbody>
</table>

Equation 7.15 represents the logit (or log odds) chosen for later use in the transformation equation to calculate the predicted probability of death in 3 years at each Survey. The logit coefficients are from Table 7.3.

\[
\text{logit(death)} = -3.241 - 0.062 \times \text{PCS}_1 - 0.031 \times \text{MCS}_1 + 0.065 \times \text{age}_1 - 0.162 \times \text{area}_1 \\
- 0.037 \times \text{educ}_1 - 0.057 \times \text{alone}_1 + 0.091 \times \text{manage}_1 + 0.216 \times \text{cmorb}_1 \\
+ 0.686 \times \text{cancer}_1
\]

(equation 7.15)

7.5.2 Transformation to generate the probability of death in 3 years:

\[
\text{Probability of Death in 3 years} = \frac{e^{-3.241-0.062\text{PCS}_1-0.031\text{MCS}_1+0.065\text{age}_1-0.162\text{area}_1-0.037\text{educ}_1-0.057\text{alone}_1+0.091\text{manage}_1+0.216\text{cmorb}_1+0.686\text{cancer}_1}}{1+e^{-3.241-0.062\text{PCS}_1-0.031\text{MCS}_1+0.065\text{age}_1-0.162\text{area}_1-0.037\text{educ}_1-0.057\text{alone}_1+0.091\text{manage}_1+0.216\text{cmorb}_1+0.686\text{cancer}_1}}
\]

(equation 7.16)

The four new variables that were generated for the predicted probability of death (Survey 1-Survey 4) used the same transformation equation, shown above with the same β coefficient estimates. However, each new variable differed by the time varying
covariates of PCS, MCS, age, manage on income, living alone, co-morbidities and cancer. The variables: area of residence and education status, were baseline values and remained constant over time. The time varying age variable was based on the age range of 69 -76 years at baseline and was increased by value of 3 years at each Survey after baseline. The majority of women had reported a baseline age. Attempts were made to consider different equations for each time period, however to remain consistent with the methodological approaches in Chapters 4 and 5, only the constant equation terms will be mentioned in this Chapter.

The transformation equation 7.16 (i.e. predicted probability) requires the observed covariates in order to calculate the predicted probability. Transformation equation 7.16 has considered more covariates than transformation equation 4.3 thus there are more missing data possibilities for this approach. The number of calculated predicted probabilities of death, at each Survey, was improved by using multiple imputation. For each survey, the calculated predicted probabilities of death increased by 4%, 12%, 22% and 28%, respectively. These proportions are based on the sample of known diabetes status (N=9180) of which 1602 participants were known to have died by Survey 4. Thus, those responders who had missing covariates at each survey were included via multiple imputation.

Initially, in the construction of the transformation equation, only baseline covariates were imputed. That is, only those still participating at Survey 2 but had missing data were considered along with those who were known to have died. Those excluded were participants who were coded as:

i) Non-responders;

ii) unable to be contacted;

iii) withdrawn from Survey.
Thus applying the transformation equation 7.16 above, the predicted probabilities at each Survey were calculated and longitudinal multiple imputation of covariates at each Survey was used. The Stata code used for this procedure was shown in Appendix 7.2. At this point in the procedure those who could not be identified by diabetes status were excluded. For the missing covariates the “missingness” assumption was MAR.

Stage 2:

7.5.3 Logit for Same or Better to obtain $\beta$ coefficients for Transformation Equation:

The following section will report the results of an approach similar to the approach used above to determine a transformation equation for the binary variable Same or Better(1) vs Worse(0). Table 7.4 reported the means and standard deviations of those women who had a PCS at Survey 1 and 2. Before multiple imputation of missing PCS values the number of pairs was $N=6928$, with a difference of -1.28 (P-value<0.001) which was statistically significant. When multiple imputation was used to increase the number of pairs ($N=9180$) to calculate a change in PCS score the difference was -1.08, (P-value <0.001).

Table 7.4 Difference in PCS before and after Imputation

<table>
<thead>
<tr>
<th></th>
<th>Before Imputation</th>
<th>After Imputation (M=1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>PCS survey 1</td>
<td>8160</td>
<td>50.13</td>
</tr>
<tr>
<td>PCS survey 2</td>
<td>7650</td>
<td>49.52</td>
</tr>
<tr>
<td>PCS survey 1</td>
<td>6928</td>
<td>50.83</td>
</tr>
<tr>
<td>PCS survey 2</td>
<td>6928</td>
<td>49.55</td>
</tr>
<tr>
<td>Difference between pairs</td>
<td>6928</td>
<td>-1.28*</td>
</tr>
<tr>
<td>Difference between pairs</td>
<td>9180</td>
<td>-1.08*</td>
</tr>
</tbody>
</table>

*paired t-test statistically significant at P-value<0.001
Table 7.5 shows the two cut-offs that have been used for the binary outcome (Same / Better). Using Rogers et al’s cut-off (5.66) 75% of women are considered in Same / Better health, (without multiple imputation of PCS values) compared to the ALSWH cut-off (8.65) where the percentage was 85%. After multiple imputation of PCS values the proportions, using the different cut-offs, stayed reasonably the same for Same or Better vs Worse. The distributions of change in PCS values are shown in Figures 7.1 and 7.2 to highlight the proportions and cut-off value (8.65) when multiple imputation of PCS had been ignored and then considered.

Table 7.5    ALSWH and Rogers’ Cut-offs (Same or Better Vs Worse)

<table>
<thead>
<tr>
<th>Survey 1 to Survey 2</th>
<th>Before Imputation</th>
<th>After Imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M=1) N=6928</td>
<td>(M=1) N=8281</td>
</tr>
<tr>
<td>ALSWH’s cut off 8.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better/Same</td>
<td>5921 85%</td>
<td>Better/Same</td>
</tr>
<tr>
<td>Worse</td>
<td>1007 15%</td>
<td>Worse</td>
</tr>
<tr>
<td>Rogers’ cut off 5.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better/Same</td>
<td>5204 75%</td>
<td>Better/Same</td>
</tr>
<tr>
<td>Worse</td>
<td>1724 25%</td>
<td>Worse</td>
</tr>
</tbody>
</table>

Figure 7.1    Difference in PCS before Multiple Imputation
To generate the binary variable for Same or Better (1) versus Worse (0), multiple imputation (M=5 datasets) was conducted which allowed for the imputation of missing PCS and MCS at Survey 1 and Survey 2 for those still known to participating at Survey 2. Thus those participants who were not coded as a respondent at Survey 2 were excluded from this analysis. This allowed for those participants who had a missing PCS at baseline or at Survey 2 to still be included in the logistic regression which would be used in the transformation equation later. If a participant had withdrawn for reasons unknown or could not be contacted they were removed from this initial logistic regression analysis. Women who had died or were considered too frail to continue between Survey 1 and Survey 2 were coded as Worse (0). This binary variable Same or Better was then used to generate the necessary β coefficients along with multiple imputation to impute for missing baseline covariates of all respondents to Survey 2.

Based on the results from the 5 combined dataset, the “averaged” regression coefficients model was chosen and is shown below in Table 7.7. These “averaged” β coefficients were later used to generate a probability of being in Same or Better health.
at each Survey. Table 7.6 shows the \( \beta \) coefficients from a logistic regression when only complete case data was used.

### Table 7.6 Logistic regression model for Same or Better no multiple imputation

<table>
<thead>
<tr>
<th></th>
<th>Coeff</th>
<th>SE</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.046</td>
<td>0.004</td>
<td>-0.054, -0.038</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Area residence</td>
<td>-0.038</td>
<td>0.074</td>
<td>-0.183, 0.106</td>
<td>0.605</td>
</tr>
<tr>
<td>Education status</td>
<td>-0.035</td>
<td>0.093</td>
<td>-0.218, 0.148</td>
<td>0.708</td>
</tr>
<tr>
<td>Age</td>
<td>-0.049</td>
<td>0.025</td>
<td>-0.097, -0.001</td>
<td>0.047</td>
</tr>
<tr>
<td>Live alone</td>
<td>-0.036</td>
<td>0.076</td>
<td>-0.185, 0.114</td>
<td>0.639</td>
</tr>
<tr>
<td>Manage income</td>
<td>-0.148</td>
<td>0.087</td>
<td>-0.317, 0.022</td>
<td>0.089</td>
</tr>
<tr>
<td>intercept</td>
<td>7.798</td>
<td>1.811</td>
<td>4.249, 11.347</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### Table 7.7 Logistic regression model (Same or Better) - MI “averaged coefficients”

<table>
<thead>
<tr>
<th></th>
<th>Coeff</th>
<th>SE</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.060</td>
<td>0.004</td>
<td>-0.068, -0.051</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Area residence</td>
<td>-0.049</td>
<td>0.067</td>
<td>-0.181, 0.083</td>
<td>0.466</td>
</tr>
<tr>
<td>Education status</td>
<td>0.031</td>
<td>0.087</td>
<td>-0.140, 0.201</td>
<td>0.724</td>
</tr>
<tr>
<td>Age</td>
<td>-0.052</td>
<td>0.023</td>
<td>-0.098, -0.006</td>
<td>0.027</td>
</tr>
<tr>
<td>Live alone</td>
<td>-0.041</td>
<td>0.066</td>
<td>-0.172, 0.089</td>
<td>0.533</td>
</tr>
<tr>
<td>Manage income</td>
<td>-0.203</td>
<td>0.087</td>
<td>-0.380, -0.026</td>
<td>0.026</td>
</tr>
<tr>
<td>intercept</td>
<td>8.628</td>
<td>1.738</td>
<td>5.166, 12.091</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The “averaged” coefficient estimates in Table 7.7 were then used to generate a predicted probability of Same or Better at each Survey. The 95% confidence intervals shown in both Tables 7.6 and 7.7 are similar in width. The regression coefficient estimates from both Tables 7.6 and 7.7 are similar in most cases. Note the standard errors (SE) are smaller in Table 7.7 due in part to the larger sample size. The estimate for education status is reversed in direction (+/-), however the Wald statistic P-value in both Tables for this variable is not statistically significant at the level of 0.05. Also, it appears that statistical significance has changed for “manage on income”. In Table 7.6, the variable manage on income was not statistically significant (P-value = 0.089) whereas in Table 7.7, the variable manage on income is statistically significant (P-value < 0.026) at a significance level of 5%.
The following logit equation (7.17) was transformed (equation 7.18) then used to generate the probability of Same or Better at each Survey (1-4). The same time varying covariates were considered along with the baseline covariates for area of residence and education level. The “averaged” β coefficients shown below were generated from the model that accounted for missing covariates (See Table 7.7).

\[
\text{logit( same } \_ \text{ better }) = 8.628 - 0.060 \times \text{PCS}_t - 0.049 \times \text{area}_t + 0.031 \times \text{educ}_t - 0.052 \times \text{age}_t - 0.041 \times \text{alone}_t - 0.203 \times \text{manage}_t
\]

(equation 7.17)

### 7.5.4 Transformation to generate the probability of Same or Better:

The following transformation equation (7.18) for predicting the Probability of being in Same or Better health in 3 years was used:

\[
\text{Probability (Same or Better)}_{t+1} = \frac{e^{8.628 - 0.060 \times \text{PCS}_t - 0.049 \times \text{area}_t + 0.031 \times \text{educ}_t - 0.052 \times \text{age}_t - 0.041 \times \text{alone}_t - 0.203 \times \text{manage}_t}}{1 + e^{8.628 - 0.060 \times \text{PCS}_t - 0.049 \times \text{area}_t + 0.031 \times \text{educ}_t - 0.052 \times \text{age}_t - 0.041 \times \text{alone}_t - 0.203 \times \text{manage}_t}
\]

where \( t = 0, 1, 2, 3 \)

(equation 7.18)

Similarly, as for the death transformation equation (7.16), there are more covariates (than for the Equation 4.3) so more missing data combinations are possible at each Survey, and thus the assumptions of MAR and MNAR are much more likely than MCAR for these covariates. Thus multiple imputation (MI) was used to impute for missing covariates (assuming MAR).
To make comparisons before and after MI, the predicted probability of Same or Better health was calculated before and after MI. A number of the covariates used in equation 7.18 are the same as for equation 7.16 so similar proportions of the missing covariates occurred.

In Table 7.8 the results from using the Rogers et al's cut-off (5.66) are shown. These results represent the “averaged” coefficient estimates, similar to that shown in Table 7.7. Comparing results with Tables 7.7 and 7.8, the differences in coefficients are due to the different proportions of those women who were considered as in Same or Better category by the respective cut-offs shown in Table 7.5.

Table 7.8 Using Rogers’ cut-off (5.66) Multiple imputation

<table>
<thead>
<tr>
<th>N=8281</th>
<th>Coeff</th>
<th>SE</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.053</td>
<td>0.003</td>
<td>-0.060, -0.046</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Area residence</td>
<td>-0.035</td>
<td>0.055</td>
<td>-0.143, 0.074</td>
<td>0.531</td>
</tr>
<tr>
<td>Education status</td>
<td>0.121</td>
<td>0.073</td>
<td>-0.022, 0.265</td>
<td>0.097</td>
</tr>
<tr>
<td>Age</td>
<td>-0.037</td>
<td>0.018</td>
<td>-0.073, -0.002</td>
<td>0.041</td>
</tr>
<tr>
<td>Live alone</td>
<td>-0.002</td>
<td>0.056</td>
<td>-0.111, 0.108</td>
<td>0.974</td>
</tr>
<tr>
<td>Manage income</td>
<td>-0.198</td>
<td>0.071</td>
<td>-0.341, -0.055</td>
<td>0.008</td>
</tr>
<tr>
<td>intercept</td>
<td>6.571</td>
<td>1.342</td>
<td>3.932, 9.209</td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>

The equation 7.19 represents the probability of being Same or Better health using the Rogers et al's cut-off and the estimates from Table 7.8.
Probability (Same or Better)_{t+1} = \frac{e^{6.571-0.053\times PCS_{t} -0.035\times area_{t} +0.121\times educ_{t} -0.037\times age_{t} -0.002\times alone_{t} -0.198\times manage_{t}}}{1 + e^{6.571-0.053\times PCS_{t} -0.035\times area_{t} +0.121\times educ_{t} -0.037\times age_{t} -0.002\times alone_{t} -0.198\times manage_{t}}}

where \( t = 0, 1, 2, 3 \)

(equation 7.19)

Stage 3:

7.5.5 Conditional Probability of being Alive and in Same or Better Health in 3 years

In order to generate the outcome variable to calculate the predicted probability of being alive and in Same or Better health, the predicted probability of being alive was calculated from the predicted probability of death. That is, the predicted probability of being alive equals one minus the predicted probability of death.

Probability of being Alive and in Same or Better Health in 3 years (APASB) = \[ \{ 1 - Pr_{\text{death}} \} \times Pr_{\text{same\_better}} \] \times 100

where

\[ Pr_{\text{death}} = \text{probability of death in 3 years} \]
\[ Pr_{\text{same\_better}} = \text{probability of Same or Better health in 3 years} \]

(equation 7.20)

Stage 4:

7.5.6 Imputing zeros for Death

To account for those who have died, a similar approach to that used by Diehr and colleagues in Chapter 4 was applied to this new outcome variable APASB. The predicted probability of being Alive and in the Same or Better health in 3 years (APASB) is set to zero for those who have died. In this Chapter, the new variables with the inclusion of zeros for death are known as APASBD. The value of zero is "carried forward" at each Survey after participants are known to have died. The Rogers et al
cut-off was also used, thus the predicted probability of being Alive and in Same or Better health in 3 years using this cut-off is known as \textit{PASB}. The variable that includes deaths is known as \textit{PASBD}. Tables 7.10 and 7.12 show the results of including zeros for deaths for both the ALSWH (8.65) and Rogers et al (5.66) cut-offs.

### 7.5.7 Descriptive comparisons of the two “PCS change” cut-offs

Table 7.9 shows the means and standard deviations of the Probability of being Alive and in Same or Better health in 3 years at each Survey by the two diabetes groups. Table 7.10 also shows the impact of including zeros at each Survey. Of note is the decrease in mean values, especially amongst the Existing cases group. The inclusion of zeros has greatly impacted on the standard deviations as could be seen previously with the equation 4.3 in Table 4.2. Tables 7.11 and 7.12 report the descriptive statistics when the Rogers’ et al cut-off was used to calculate the Probability of being Alive and in Same or Better health in 3 years by the two diabetes groups.

#### Table 7.9 Descriptive Statistics – ALSWH cut-off (8.65) APASB at Surveys 1-4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APASB1</td>
<td>5879</td>
<td>78.8</td>
<td>5.1</td>
<td>54.7</td>
<td>89.8</td>
</tr>
<tr>
<td>APASB2</td>
<td>5487</td>
<td>76.3</td>
<td>5.5</td>
<td>46.1</td>
<td>87.9</td>
</tr>
<tr>
<td>APASB3</td>
<td>4918</td>
<td>73.5</td>
<td>5.8</td>
<td>42.2</td>
<td>85.8</td>
</tr>
<tr>
<td>APASB4</td>
<td>4678</td>
<td>70.4</td>
<td>6.2</td>
<td>40.0</td>
<td>83.2</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APASB1</td>
<td>825</td>
<td>79.7</td>
<td>4.7</td>
<td>62.5</td>
<td>89.7</td>
</tr>
<tr>
<td>APASB2</td>
<td>588</td>
<td>77.0</td>
<td>5.4</td>
<td>52.4</td>
<td>87.3</td>
</tr>
<tr>
<td>APASB3</td>
<td>465</td>
<td>73.7</td>
<td>5.9</td>
<td>52.8</td>
<td>84.4</td>
</tr>
<tr>
<td>APASB4</td>
<td>347</td>
<td>70.0</td>
<td>7.0</td>
<td>43.5</td>
<td>82.7</td>
</tr>
</tbody>
</table>
Table 7.10  Descriptive Statistics – ALSWH cut-off (8.65) APASBD at Surveys 1-4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APASBD1</td>
<td>5879</td>
<td>78.8</td>
<td>5.1</td>
<td>54.7</td>
<td>89.8</td>
</tr>
<tr>
<td>APASBD2</td>
<td>5902</td>
<td>70.9</td>
<td>20.2</td>
<td>0</td>
<td>87.9</td>
</tr>
<tr>
<td>APASBD3</td>
<td>5722</td>
<td>63.1</td>
<td>26.1</td>
<td>0</td>
<td>85.8</td>
</tr>
<tr>
<td>APASBD4</td>
<td>5934</td>
<td>55.5</td>
<td>29.3</td>
<td>0</td>
<td>83.2</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>APASBD1</td>
<td>825</td>
<td>79.7</td>
<td>4.7</td>
<td>62.5</td>
<td>89.7</td>
</tr>
<tr>
<td>APASBD2</td>
<td>682</td>
<td>66.4</td>
<td>27.0</td>
<td>0</td>
<td>87.3</td>
</tr>
<tr>
<td>APASBD3</td>
<td>654</td>
<td>52.4</td>
<td>33.8</td>
<td>0</td>
<td>84.4</td>
</tr>
<tr>
<td>APASBD4</td>
<td>643</td>
<td>37.8</td>
<td>35.3</td>
<td>0</td>
<td>82.7</td>
</tr>
</tbody>
</table>

Comparing Tables 7.9 and 7.11, there was a similar decline in means over time when the two different cut-offs (Rogers’ 5.66) are used. The mean differences in Table 7.9 and 7.11 are due to the difference of an approximately 3 units in the cut-off values.

Table 7.11  Descriptive Statistics- Rogers’ cut-off (5.66) PASB at Surveys 1-4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASB1</td>
<td>5879</td>
<td>70.5</td>
<td>6.4</td>
<td>44.5</td>
<td>86.1</td>
</tr>
<tr>
<td>PASB2</td>
<td>5487</td>
<td>68.3</td>
<td>6.5</td>
<td>43.4</td>
<td>83.0</td>
</tr>
<tr>
<td>PASB3</td>
<td>4918</td>
<td>66.1</td>
<td>6.6</td>
<td>40.0</td>
<td>80.6</td>
</tr>
<tr>
<td>PASB4</td>
<td>4678</td>
<td>63.7</td>
<td>6.6</td>
<td>38.2</td>
<td>78.8</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASB1</td>
<td>825</td>
<td>72.4</td>
<td>5.8</td>
<td>53.9</td>
<td>85.1</td>
</tr>
<tr>
<td>PASB2</td>
<td>588</td>
<td>70.0</td>
<td>6.1</td>
<td>46.8</td>
<td>82.1</td>
</tr>
<tr>
<td>PASB3</td>
<td>465</td>
<td>67.1</td>
<td>6.2</td>
<td>49.3</td>
<td>78.9</td>
</tr>
<tr>
<td>PASB4</td>
<td>347</td>
<td>63.9</td>
<td>6.8</td>
<td>41.3</td>
<td>77.2</td>
</tr>
</tbody>
</table>
Table 7.12  Descriptive Statistics- Rogers' cut-off (5.66) PASBD at Surveys 1-4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>No diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASBD1</td>
<td>5879</td>
<td>70.5</td>
<td>6.4</td>
<td>44.5</td>
<td>86.1</td>
</tr>
<tr>
<td>PASBD2</td>
<td>5902</td>
<td>63.5</td>
<td>18.6</td>
<td>0</td>
<td>83.0</td>
</tr>
<tr>
<td>PASBD3</td>
<td>5722</td>
<td>56.8</td>
<td>23.8</td>
<td>0</td>
<td>80.6</td>
</tr>
<tr>
<td>PASBD4</td>
<td>5934</td>
<td>50.2</td>
<td>26.7</td>
<td>0</td>
<td>78.8</td>
</tr>
<tr>
<td>Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PASBD1</td>
<td>825</td>
<td>72.4</td>
<td>5.8</td>
<td>53.9</td>
<td>85.1</td>
</tr>
<tr>
<td>PASBD2</td>
<td>682</td>
<td>60.3</td>
<td>24.8</td>
<td>0</td>
<td>82.1</td>
</tr>
<tr>
<td>PASBD3</td>
<td>654</td>
<td>47.7</td>
<td>30.9</td>
<td>0</td>
<td>78.9</td>
</tr>
<tr>
<td>PASBD4</td>
<td>643</td>
<td>34.5</td>
<td>32.3</td>
<td>0</td>
<td>77.2</td>
</tr>
</tbody>
</table>

Comparing Table 7.10 and 7.12 when the different cut-offs are used the similar declines in means over time are noticeable for both cut-offs. The wider and increasing standard deviations over time are also similar for both cut-offs. Again the overall differences are due to the larger cut-off value (8.65) for ALSWH.

The following example highlights the difference in mean PASB for both cut-offs. A PCS score of 40 at Survey 1 (adjusted for all other covariates) would equate to an APASB value of 74, while the PASB value would be 68. Thus in Tables 7.9 and 7.11 for those women with existing diabetes the mean APASB at Survey 1 (mean = 78.8, sd = 5.1) is higher than the mean PASB at survey 1 (mean=70.5, sd= 6.4) which is due to the differences in estimates in equations 7.18 and 7.19, respectively. Similarly in Tables 7.9 and 7.11 for those women without diabetes, the mean APASB is higher than the mean PASB. Also, when deaths are included in Tables 7.10 and 7.12 the APASBD is again higher than PASBD for both diabetes groups.
Finally on the comparisons of cut-offs, the differences are noted, however only cross-sectional and longitudinal analyses results from the ALSWH cut-off (8.65) are reported in this chapter. Results using the Rogers et al cut-off (5.66) are not shown.

**7.5.8 A Different Approach to Better vs Same vs Worse Using Ordered Logistic Regression**

**7.5.8.1 oLogit for Better vs Same vs Worse to obtain β coefficients:**

The following section will report the results of a similar approach used above to determine a transformation equation, but this time using a categorical variable: Worse (0) vs Same (1) vs Better (2). Table 7.13 shows the two cut-offs that have been used for the ordinal outcome (Worse/Same / Better). Using Rogers et al’s cut-off (5.66), 14% of women are considered in Better health and 56% were considered to be in the Same health, (without multiple imputation of PCS values). For the ALSWH cut-off (8.65) 7% of women are considered in Better health and 72% were considered to be in the Same health. After multiple imputation of PCS values the proportions, using the different cut-offs, stayed reasonably the same for Worse vs Same vs Better. The distributions of change in PCS values are shown in Figures 7.3 to highlight the proportions and for the cut-off values when multiple imputation of PCS had not been used. The ALSWH classifies less women in Better health since the cut-off is larger than the Rogers et al's cut-off.
Table 7.13  ALSWH and Rogers’ cut-offs (Better Vs Same Vs Worse)

<table>
<thead>
<tr>
<th></th>
<th>Before Imputation</th>
<th>After Multiple Imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=7460</td>
<td>N=8813</td>
</tr>
<tr>
<td><strong>ALSWH’s cut off 8.65</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better</td>
<td>563 (7%)</td>
<td>835 (9%)</td>
</tr>
<tr>
<td>Same</td>
<td>5358 (72%)</td>
<td>6135 (70%)</td>
</tr>
<tr>
<td>Worse*</td>
<td>1539 (21%)</td>
<td>1843 (21%)</td>
</tr>
<tr>
<td><strong>Rogers’ cut off 5.66</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better</td>
<td>1055 (14%)</td>
<td>1429 (16%)</td>
</tr>
<tr>
<td>Same</td>
<td>4149 (56%)</td>
<td>4689 (53%)</td>
</tr>
<tr>
<td>Worse*</td>
<td>2256 (30%)</td>
<td>2695 (31%)</td>
</tr>
</tbody>
</table>

*Worse category is increased by those women who were too frail to continue in study or died between Survey 1 and Survey 2

Figure 7.3  Difference in PCS before Multiple Imputation
7.5.8.2 Transformation for Probability of Better using ALSWH cut-off 8.65

The following transformations were used to generate the probability of being Better at each survey (1-4). The same time varying covariates were considered along with the baseline covariates for area of residence and education level.

Where \( \text{cut}_1 = -8.58062, \text{cut}_2 = -4.26335 \),
\[
\begin{align*}
\beta_1 &= -0.08306, \\
\beta_2 &= -0.02762, \\
\beta_3 &= -0.025693, \\
\beta_4 &= -0.033556, \\
\beta_5 &= -0.02354, \\
\beta_6 &= -0.158779.
\end{align*}
\]

Probability of Better \( Pr(\text{Better}) \) = \[
\frac{e^{4.26 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}}{1 + e^{4.26 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}}
\]

Probability of Worse \( Pr(\text{Worse}) \) = \[
1 - \left( \frac{e^{8.58 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}}{1 + e^{8.58 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}} \right)
\]

Probability of Same \( Pr(\text{Same}) \) = \[
1 - \left( \frac{e^{8.58 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}}{1 + e^{8.58 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}} \right) - \left( \frac{e^{4.26 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}}{1 + e^{4.26 - 0.08 \cdot PCS - 0.02 \cdot \text{educ} - 0.03 \cdot \text{age} - 0.024 \cdot \text{manage} - 0.159 \cdot \text{manage}^2}} \right)
\]

The reason for including the results from the ordered logistic regression was to determine what may occur when the cut-off is changed. In Table 7.14 the probabilities of being Alive and in Better health are presented with and without deaths. Thus for Survey 1 the mean probability of being Alive and in better health at survey 2 is 9.6%, whereas for Survey 4 the mean probability of being Alive and in Better health at Survey 5 is 9%. However when deaths are considered, for Survey 4 the mean probability of being Alive and in Better health at Survey 5 decreases to 7%. Thus the inclusion of deaths (zeros) decreases the probability over of being Alive and in Better health.

In Table 7.15 the probabilities of being Alive and in Same health are presented with and without deaths. Thus for Survey 1 the mean probability of being Alive and in Same health at survey 2 is 69.9%, whereas for Survey 4 the mean probability of being Alive
and in Same health at Survey 5 is 64.9%. However when deaths are considered, for Survey 4 the mean probability of being Alive and in Same health at Survey 5 decreases to 50.3%. Thus similar to Table 7.14, the inclusion of deaths (zeros) decreases the probability over of being Alive and in Same health.

Table 7.14 Probability of being Alive and Better (with and without deaths) at next Survey

<table>
<thead>
<tr>
<th>Probability of being Alive and Better at next survey</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAB1</td>
<td>7291</td>
<td>9.6</td>
<td>7.2</td>
<td>1.1</td>
<td>47.5</td>
</tr>
<tr>
<td>PAB2</td>
<td>6652</td>
<td>9.2</td>
<td>6.6</td>
<td>1.2</td>
<td>40.5</td>
</tr>
<tr>
<td>PAB3</td>
<td>5885</td>
<td>9.1</td>
<td>6.4</td>
<td>1.2</td>
<td>39.7</td>
</tr>
<tr>
<td>PAB4</td>
<td>5484</td>
<td>9.0</td>
<td>5.9</td>
<td>1.2</td>
<td>35.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probability of being Alive and Better at next survey (deaths included)</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PABD1</td>
<td>7291</td>
<td>9.6</td>
<td>7.2</td>
<td>1.1</td>
<td>47.5</td>
</tr>
<tr>
<td>PABD2</td>
<td>7161</td>
<td>8.5</td>
<td>6.8</td>
<td>0</td>
<td>40.5</td>
</tr>
<tr>
<td>PABD3</td>
<td>6891</td>
<td>7.8</td>
<td>6.7</td>
<td>0</td>
<td>39.7</td>
</tr>
<tr>
<td>PABD4</td>
<td>7086</td>
<td>7.0</td>
<td>6.4</td>
<td>0</td>
<td>35.6</td>
</tr>
</tbody>
</table>

Table 7.15 Probability of being Alive and Same (with and without deaths) at next Survey

<table>
<thead>
<tr>
<th>Probability of being Alive and Same at next survey</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAS1</td>
<td>7291</td>
<td>69.9</td>
<td>6.2</td>
<td>30.7</td>
<td>77.6</td>
</tr>
<tr>
<td>PAS2</td>
<td>6652</td>
<td>68.8</td>
<td>6.3</td>
<td>21.1</td>
<td>77.0</td>
</tr>
<tr>
<td>PAS3</td>
<td>5885</td>
<td>67.0</td>
<td>6.9</td>
<td>29.0</td>
<td>76.2</td>
</tr>
<tr>
<td>PAS4</td>
<td>5484</td>
<td>64.9</td>
<td>7.5</td>
<td>28.2</td>
<td>75.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probability of being Alive and Same at next survey (deaths included)</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PASD1</td>
<td>7291</td>
<td>69.9</td>
<td>6.2</td>
<td>30.7</td>
<td>77.6</td>
</tr>
<tr>
<td>PASD2</td>
<td>7161</td>
<td>63.9</td>
<td>18.7</td>
<td>0</td>
<td>77.0</td>
</tr>
<tr>
<td>PASD3</td>
<td>6891</td>
<td>57.2</td>
<td>24.5</td>
<td>0</td>
<td>76.2</td>
</tr>
<tr>
<td>PASD4</td>
<td>7086</td>
<td>50.3</td>
<td>27.9</td>
<td>0</td>
<td>75.5</td>
</tr>
</tbody>
</table>
7.5.9 Cross-sectional Analysis

The cross-sectional analyses shown in this section, used one-way ANOVAs to make multiple comparisons of the diabetes categories at each survey. Tables 7.16 and 7.17 show the results when deaths are included and excluded from the analysis. Most noticeable is the impact of deaths on the Existing cases. The results in Table 7.16 show that the mean differences between Existing cases and No diabetes decrease over time, but by Survey 3 there is no statistically significant difference in means. However, in Table 7.17 there are statistically significant differences between No diabetes and Existing cases at each Survey. Also, in Table 7.17, comparisons between all groups at each Survey are statistically significant when deaths are included. At Survey 4 in Table 7.16, the mean difference between Existing cases and No diabetes groups without deaths is 0.35 (P-value = 0.968) compared with a mean difference of 17.69 (P-value ≤0.001) when deaths are included in Table 7.17.
Table 7.16  “Change in Health” APASB (no deaths) for Diabetes Groups at Surveys 1 to 4

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean (SD)</th>
<th>**Bonferroni P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Survey 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>825</td>
<td>79.7 (4.7)</td>
<td>Exist v New 0.02 (1.00)</td>
</tr>
<tr>
<td>New Case</td>
<td>587</td>
<td>79.7 (4.9)</td>
<td>No D v New -0.84 (P&lt;0.001)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5879</td>
<td>78.8 (5.1)</td>
<td>No D v Exist -0.86 (P&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>588</td>
<td>77.0 (5.4)</td>
<td>Exist v New -0.23 (1.00)</td>
</tr>
<tr>
<td>New Case</td>
<td>577</td>
<td>77.3 (5.3)</td>
<td>No D v New -0.97 (P&lt;0.001)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5487</td>
<td>76.3 (5.5)</td>
<td>No D v Exist -0.74 (P&lt;0.006)</td>
</tr>
<tr>
<td><strong>Survey 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>465</td>
<td>73.7 (5.9)</td>
<td>Exist v New -0.35 (1.00)</td>
</tr>
<tr>
<td>New Case</td>
<td>502</td>
<td>74.1 (5.7)</td>
<td>No D v New -0.62 (0.07)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>4918</td>
<td>73.5 (5.8)</td>
<td>No D v Exist -0.28 (0.993)</td>
</tr>
<tr>
<td><strong>Survey 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>347</td>
<td>70.0 (6.9)</td>
<td>Exist v New -0.87 (0.159)</td>
</tr>
<tr>
<td>New Cases</td>
<td>459</td>
<td>70.9 (6.7)</td>
<td>No D v New -0.52 (0.274)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>4678</td>
<td>70.4 (6.2)</td>
<td>No D v Exist 0.35 (0.968)</td>
</tr>
</tbody>
</table>

*Statistically Significant at alpha level of 0.001

Figure 7.4 and Figure 7.5 show cross-sectional comparisons between the three diabetes groups. In Figure 7.4 when deaths are not included the differences between the means across all three groups appears quite close. In Table 7.16 there are some statistically significant differences (P<0.001) between groups but these differences are not obvious in Figure 7.4 because of the scale on the Y axis. In Figures 7.4 and 7.5 the same scale was used to highlight the impact of including deaths on the means in all three groups. This is clearly seen in Figure 7.5 where there are obvious differences between means at Surveys 2, 3 and 4 for Existing cases and the No diabetes groups.
Figure 7.4  ALSWH Mean Probability of being Alive and in Same or Better health at each Survey by diabetes groups without deaths

Figure 7.5  ALSWH Mean Probability of being Alive and in Same or Better health at each Survey by diabetes groups with deaths
Table 7.17  “Change in Health” APASBD (deaths included) for Diabetes at Surveys 1 to 4

<table>
<thead>
<tr>
<th>Groups</th>
<th>N</th>
<th>Mean (SD)</th>
<th><strong>Bonferroni P-value</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Survey 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>825</td>
<td>79.7 (4.7)</td>
<td>Exist v New 0.02 (1.00)</td>
</tr>
<tr>
<td>New Case</td>
<td>587</td>
<td>79.7 (4.9)</td>
<td>No D v New -0.84 (P&lt;0.001)</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5879</td>
<td>78.8 (5.1)</td>
<td>No D v Exist -0.86 (P&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>682</td>
<td>66.4 (27.0)</td>
<td>Exist v New -10.85(P&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>577</td>
<td>77.3 (5.3)</td>
<td>No D v New -4.51 (P&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5902</td>
<td>70.9 (20.2)</td>
<td>No D v Exist 6.33 (P&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>654</td>
<td>52.4 (33.8)</td>
<td>Exist v New -19.8 (P&lt;0.001)*</td>
</tr>
<tr>
<td>New Case</td>
<td>515</td>
<td>72.2 (12.9)</td>
<td>No D v New -9.07 (P&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5722</td>
<td>63.1 (4.4)</td>
<td>No D v Exist 10.71 (P&lt;0.001)*</td>
</tr>
<tr>
<td><strong>Survey 4</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Cases</td>
<td>643</td>
<td>37.8 (35.3)</td>
<td>Exist v New -26.14 (P&lt;0.001)</td>
</tr>
<tr>
<td>New Cases</td>
<td>509</td>
<td>63.9 (22.1)</td>
<td>No D v New -8.45 (P&lt;0.001)*</td>
</tr>
<tr>
<td>No Diabetes</td>
<td>5934</td>
<td>55.5 (29.3)</td>
<td>No D v Exist 17.69 (P&lt;0.001)*</td>
</tr>
</tbody>
</table>

** New = New cases, No D = No diabetes, Exist=Existing cases
*Statistically Significant at alpha level of 0.001

For cross-sectional analysis no imputation, apart from including zeros for deaths, were considered. As shown in earlier in Chapter 5, when imputing zeros for deaths, some compensation needs to be granted for missing data. Since the sicker (diabetes) smaller sample size group will be more greatly impacted by the reinstatement of those who died. As has been highlighted earlier, the existing diabetes group has a higher proportion of deaths and therefore the impact of the zeros lowers the mean value for this sicker group. Thus for the longitudinal analysis which follows, multiple imputation of missing covariates and/or outcomes (APASBD) will also be reported.
7.5.10 Longitudinal Analysis

7.5.10.1 Longitudinal analysis Random intercept models adjusting for an interaction between time and diabetes status

In this analysis, the time variable was considered as a continuous variable. There has been discussion in the literature (Twisk 2003) as to when to consider “time” as a continuous or categorical variable. This especially becomes an issue when the time intervals are not equal. However, in this ALSWH dataset the time intervals are equal (3 years).

The general form of the random effects model with an interaction between diabetes classification and time was the following:

\[ Y_{it} = \beta_0 + \beta_1 \text{time} + \beta_2 \text{Diab_retro} + \beta_3 \text{time} \times \text{Diab_retro} + \epsilon_{it} \]

where \( Y_{it} \) represents APASB or APASBD at the 4 time points.

\( \beta_2 \text{Diab_retro} = \) the effect size difference between the no diabetes and existing cases

\( \beta_3 \text{time} \times \text{Diab_retro} = \) interaction between diabetes groups and time

(equation 7.21)

The following Stata10.1 command was performed to generate a model when no deaths or imputation of missing covariates were considered. Table 7.18 shows the results of this first random effects model. The adjusted mean results of this model in Table 7.18 are also displayed below in Figure 7.6.

\[ \text{xi: xtmixed APASB i.diab_retro*time || IDalias:} \]

The following Stata10.1 command developed by Royston (2005) was used to combine all 5 complete datasets generated using multiple imputation to produce averaged beta coefficients. The random intercepts is determined by a variable called \_mi:, it is used instead of the variable IDalias.
According to Table 7.18, when deaths and missing covariates are not considered the adjusted mean APASB difference between Existing cases and No diabetes is 0.41, 95% CI: (0.08, 0.75) which is statistically significant (P-value = 0.015). This small mean difference appears to be significant over time since when an interaction between time and diabetes groups is included the Wald statistics for the interaction are statistically significant (P-values = 0.04 and <0.001). Figures 7.6 and 7.7 are drawn to match the scale shown later Figures 7.8 and 7.9. Figure 7.6 shows what this constant difference would be like over time but the statistically significant result from Table 7.18 is not apparent due to the scale used. Since the interaction is significant Figure 7.7 would be a better indication of what is occurring over time. Again the scale shows how close the two groups are over time when deaths are not considered. The slopes of the two groups are crossing over. Interesting to note that the existing diabetes group has a higher starting point (mean = 79.7) as was seen in Tables 7.16. This may be due to some artifact from the data. Also, the results of Table 7.18 report only the survivors whom have available data at each Survey.

Table 7.18  (Same or Better) Random intercept model No Deaths (no MI) and with a time interaction

<table>
<thead>
<tr>
<th>N obs</th>
<th>N grps</th>
<th>Outcome= APASB</th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>25312</td>
<td>8328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Diabetes (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New cases</td>
<td>1.12</td>
<td>0.59, 1.65</td>
<td>P&lt;0.001</td>
<td>0.73</td>
<td>0.36, 1.10</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing cases</td>
<td>1.41</td>
<td>0.93, 1.89</td>
<td>P&lt;0.001</td>
<td>0.41</td>
<td>0.08, 0.75</td>
<td>0.015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-2.81</td>
<td>-2.86, -2.76</td>
<td>P&lt;0.001</td>
<td>-2.86</td>
<td>-2.90, -2.82</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Cases X time</td>
<td>-0.17</td>
<td>-0.33, -0.01</td>
<td>0.040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing X time</td>
<td>-0.49</td>
<td>-0.66, -0.32</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>81.7</td>
<td>81.6, 81.9</td>
<td>P&lt;0.001</td>
<td>81.8</td>
<td>81.7, 81.9</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7.6  Adjusted mean probabilities of being Alive and in Same or Better health over time (No deaths and no interaction)

Figure 7.7  Adjusted mean probabilities of being Alive and in Same or Better health over time (No deaths but an interaction)
Table 7.19 shows the results when deaths but no missing covariates or outcomes are considered. As in Table 7.18 there is a model which includes an interaction term for time and diabetes status. In Table 7.19 when deaths were included the mean difference in Existing cases and No diabetes groups (-6.88, 95%CI: (-8.12, -5.64)) is greater than in Table 7.18 and statistically significant (P-value <0.001). There also appears to be a significant interaction, thus the mean difference in Existing cases and No diabetes groups appears to be showing a greater change over time. Figures 7.8 and 7.9 show how firstly the gap between the slopes has widen compared to Figures 7.6 and 7.7. The slopes of the graphs in Figure 7.9 are much further apart towards Surveys 3 and 4 and are not crossing over as in Figure 7.7.

Comparing results from Table 7.18 to Table 7.19, the impact of adding deaths on the two groups of interest (Existing diabetes and No diabetes) is evident. The zeros have widened the gap between the two groups and this is clearly shown in Figure 7.9 This result may be a stronger indication of the “true decline” in the Existing cases group.

Table 7.19 (Same or Better) Random intercept model (Deaths) without multiple imputation

<table>
<thead>
<tr>
<th>Outcome= APASBD</th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Diabetes (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New cases</td>
<td>0.38</td>
<td>-1.71, 2.46</td>
<td>0.725</td>
<td>5.85</td>
<td>4.41, 7.28</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Existing cases</td>
<td>5.60</td>
<td>3.78, 7.42</td>
<td>P&lt;0.001</td>
<td>-6.88</td>
<td>-8.12, -5.64</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Time</td>
<td>-7.32</td>
<td>-7.50, -7.14</td>
<td>P&lt;0.001</td>
<td>-7.65</td>
<td>-7.82, -7.48</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>New Cases X time</td>
<td>2.30</td>
<td>1.66, 2.93</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing X time</td>
<td>-5.36</td>
<td>-5.93, -4.78</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>85.3</td>
<td>84.6, 85.9</td>
<td>P&lt;0.001</td>
<td>86.1</td>
<td>85.5, 86.7</td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 7.8  Adjusted mean probabilities of being Alive and in Same or Better health over time (Deaths included but no interaction)

Figure 7.9  Adjusted mean probabilities of being Alive and in Same or Better health over time (Deaths included and time interaction)
Table 7.20 shows the results from two random intercept models which include deaths and imputed outcome values. These models are to demonstrate the impact of reinstating deaths and missing outcome data. The models were generated using mim: command and reported the combined analyses from these multiply imputed models. Since interest at present is whether there are differences APASBD in Existing cases and No diabetes over time, the outcome variable will include imputed values. This has increased the sample size and allows for comparisons to when only deaths have been accounted for (see Table 7.19).

In Table 7.20 when no interaction between diabetes group and time is considered the mean difference between Existing cases and No diabetes is (-3.17, 95% CI: (-4.13, -2.20)) which is statistically significant (P-value <0.001). Figure 7.10 illustrates how this mean difference between the two groups lies between the results shown in Figures 7.6 and 7.8. In Table 7.20 the interaction term is also statistically significant (P-value <0.001) which similar to earlier results before imputation of missing outcomes data. Figure 7.11 illustrates this interaction by the different slopes of the two diabetes groups. The results suggest that imputing outcome data (APASBD) and including zeros for deaths may in fact be a more appropriate way to estimate the true mean differences in the groups.

Table 7.20 (Same or Better) Random intercept model (Deaths) with multiple imputation of APASBD

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Outcome= APASBD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Diabetes (ref)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New cases</td>
<td>-0.29</td>
<td>-2.02, 1.45</td>
<td>0.746</td>
<td>5.30</td>
<td>4.14, 6.46</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Existing cases</td>
<td>3.35</td>
<td>1.90, 4.79</td>
<td>P&lt;0.001</td>
<td>-3.17</td>
<td>-4.13, -2.20</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>Time</td>
<td>-6.85</td>
<td>-7.00, -6.69</td>
<td>P&lt;0.001</td>
<td>-6.99</td>
<td>-7.13, -6.84</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>New Cases Xtime</td>
<td>2.24</td>
<td>1.72, 2.75</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Existing Xtime</td>
<td>-2.61</td>
<td>-3.04, -2.18</td>
<td>P&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>85.6</td>
<td>85.1, 86.2</td>
<td>P&lt;0.001</td>
<td>85.9</td>
<td>85.5, 86.5</td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 7.10  Mean probabilities of being Alive and in Same or Better health over time (with Deaths and imputed APASBD no interaction)

Figure 7.11  Mean probabilities of being Alive and in Same or Better health over time (Deaths and imputed APASBD and an interaction)
7.5.10.2 Longitudinal analysis – Random slopes and Intercepts models adjusting for independent covariates

In a previous chapter, the approach to handling missing data was by using WinBUGS. This approach involved imputing only missing outcomes and allowed for the possible assumptions of missing at random (MAR) or missing not at random (MNAR). The imputation of covariates was not considered in Chapter 5 since the focus at that stage was considering the implications of MAR and MNAR for the newly created outcome APCTD over time. Also, in earlier Chapters the covariates used previously were for those participants who had certain baseline covariates

In Tables 7.21 and 7.22 the results presented are from random slopes and random intercepts models which included baseline covariates. In Table 7.21, the model included deaths as zeros but no multiple imputation was used. In Table 7.22 missing baseline covariates were imputed as well as the missing outcome APASBD using the mim: command in Stata10.1. The Stata code used for the multiple imputation can be found in Appendix 7.2.

Table 7.21  (Same or Better) Random slopes and intercepts (Deaths) without MI

<table>
<thead>
<tr>
<th>N = 8083</th>
<th>Random intercept</th>
<th>Random Intercept and slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome= APASBD</td>
<td>β</td>
<td>95% CI</td>
</tr>
<tr>
<td>No Diabetes (ref)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New cases</td>
<td>5.34</td>
<td>3.89, 6.79</td>
</tr>
<tr>
<td>Existing cases</td>
<td>-7.08</td>
<td>-8.35, -5.81</td>
</tr>
<tr>
<td>Baseline Age group</td>
<td>-4.18</td>
<td>-5.03, -3.33</td>
</tr>
<tr>
<td>Baseline co-morbidities</td>
<td>3.51</td>
<td>2.68, 4.34</td>
</tr>
<tr>
<td>Baseline Smoke</td>
<td>-8.11</td>
<td>-9.63, -6.59</td>
</tr>
<tr>
<td>Baseline Area</td>
<td>1.03</td>
<td>0.23, 1.82</td>
</tr>
<tr>
<td>Time</td>
<td>-7.94</td>
<td>-8.13, -7.75</td>
</tr>
<tr>
<td>Intercept</td>
<td>86.8</td>
<td>85.8, 87.9</td>
</tr>
</tbody>
</table>
In Table 7.21, when similar baseline covariates are considered as in Chapter 5, the random intercepts model shows a similarly larger adjusted mean difference for Existing cases and No diabetes groups (-7.08, 95% CI: (-8.35, -5.81)) when deaths are included, compared to that shown in earlier models. In Table 7.21 the sample size is smaller because of the increase in use of baseline covariates. When the random slope and intercepts were considered as in earlier chapters, the adjusted mean difference in the two groups is statistically significant, at a 5% level, however this adjusted mean difference is not as large when only random intercepts were considered.

Table 7.22  (Same or Better) Random intercepts and slopes (Deaths) (with MI)

<table>
<thead>
<tr>
<th>N obs = 36696</th>
<th>Random intercepts</th>
<th>Random intercept and slopes</th>
</tr>
</thead>
<tbody>
<tr>
<td>M=5 N grps = 9174</td>
<td>β</td>
<td>95% CI</td>
</tr>
<tr>
<td>Outcome= APASBD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Diabetes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>New cases</td>
<td>4.47</td>
<td>3.28, 5.65</td>
</tr>
<tr>
<td>Exist cases</td>
<td>-4.71</td>
<td>-5.70, -3.73</td>
</tr>
<tr>
<td>Baseline Age</td>
<td>-4.19</td>
<td>-4.90, -3.49</td>
</tr>
<tr>
<td>Co-Morbidities</td>
<td>3.17</td>
<td>2.36, 3.98</td>
</tr>
<tr>
<td>Baseline Area</td>
<td>1.21</td>
<td>0.56, 1.87</td>
</tr>
<tr>
<td>Baseline Smoke</td>
<td>-5.54</td>
<td>-6.80, -4.29</td>
</tr>
<tr>
<td>Time</td>
<td>-8.59</td>
<td>-8.77, -8.40</td>
</tr>
<tr>
<td>Intercept</td>
<td>88.4</td>
<td>87.4, 89.3</td>
</tr>
</tbody>
</table>
In Table 7.2, the results presented are from two models: random slopes and intercepts model and random intercepts model. The main differences to the earlier models are the use of the time varying co-morbidities variable. The remaining variables are baseline covariates previously considered. In Table 7.2, for the random intercepts model, the adjusted mean difference between existing diabetes and no diabetes groups is -4.71 (95% CI: (-5.70, -3.73)) and is also statistically significantly (P-value <0.001) when compared to the result in Table 7.21, the mean difference is -7.08 (95% CI: (-8.35, -5.81)). The results from Tables 7.21 and 7.22 suggest that imputation may provide a more acceptable adjusted mean difference and generalisable results than just imputing zeros for deaths and ignoring missing data.
Figure 7.13 highlights the impact of including imputed data for missing covariates and outcome (APASBD). That is, the adjusted mean difference between groups over time has narrowed compared to results in Figure 7.12.

Figure 7.13  Adjusted mean probabilities of being Alive and in Same or Better health over time (Deaths and imputed APASBD and no interaction)
7.5.11 Cross-Sectional Analysis for ordered logistic regression

Figures 7.14 and 7.15 are shown to highlight the separation of Better and Same health. Figure 7.14 shows a cross sectional comparison mean probabilities of being Alive and in Better health for Existing cases and No diabetes at each survey. This figure shows that those women with diabetes on average had a slightly higher probability of being Alive and in Better health. When deaths are included (0) by survey 4 the mean probability of being Alive and in Better health is lower than those without diabetes. Note Figures 7.14 and 7.15 are shown simply to highlight when Better and Same are separated, no cross sectional statistical tests are included in this chapter.

![Graph showing mean probability of being Alive and Better Health at each survey by diabetes groups with Deaths included](image)

Figure 7.14  Mean Probability of being Alive and Better Health at each survey by diabetes groups with Deaths included
In Figure 7.15 the cross-sectional comparisons appear to behave similarly to when Better and Same health were combined earlier in the chapter. This figure suggests that a greater proportion of women in both diabetes groups are classified as being Same health. Also, a greater proportion of women who were classified as Same health died between surveys. Figure 7.15 suggests that gap between women with Existing cases and No diabetes appears to be widening over time for those classified as Same health.

![Probability of being Alive and Same Health by Diabetes Groups](image)

Figure 7.15 Mean Probability of being Alive and Same Health by Diabetes Groups (Deaths included)
7.5.12 Longitudinal Analysis for ordered logistic regression

Results in Table 7.23 include random intercept models which include interactions between time and Diabetes groups. The models are similar to that produced earlier however the outcome this time is for the Probability of being Alive and Better. The second model includes multiple imputation of the outcome variable. The sample sizes used to models differ (N=597). Table 7.23 shows the results of two models. The first model has average adjusted coefficients where the second model has coefficients produced without imputation. The coefficients for the imputation model are all slightly smaller in this model and MAR was assumed for missing Probability of being Alive and Better (APABD) values. Results from Probability of being Alive and Same were considered but are not shown. Figures 7.16 and 7.17 illustrate the impact of imputation on the results shown in Table 7.23. Interesting to note that when the Probability of being Alive and Better is considered there is a cross-over of the fitted lines for the groups. When the outcome was Same or Better (combined) as shown in earlier analyses, this is not as apparent as shown in Figure 7.16 (although the scale of the graph does slightly over emphasise this point).
Figure 7.16  Adjusted Probability of being Alive and Better by diabetes groups with Deaths included and time interaction without multiple imputation

Figure 7.17  Mean Probability of being Alive and Better by diabetes groups with Deaths included and time interaction with MI
Table 7.23  (Better) Random intercept model (Deaths) with and without multiple imputation

<table>
<thead>
<tr>
<th>APABD</th>
<th>β</th>
<th>95% CI</th>
<th>P-value</th>
<th>N obs = 28429 N grps= 8583</th>
<th>M=5 N obs= 36720 N grps= 9180</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Diabetes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New cases</td>
<td>1.03</td>
<td>0.38,</td>
<td>P&lt;0.001</td>
<td>0.94 0.36, 1.51 0.001</td>
<td></td>
</tr>
<tr>
<td>Existing cases</td>
<td>4.82</td>
<td>4.25,</td>
<td>P&lt;0.001</td>
<td>4.08 3.60, 4.55 P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-0.71</td>
<td>-0.77,</td>
<td>P&lt;0.001</td>
<td>-0.65 -0.70, -0.60 P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>New CasesXtime</td>
<td>0.39</td>
<td>0.19,</td>
<td>P&lt;0.001</td>
<td>0.36 0.19, 0.54 P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>ExistingXtime</td>
<td>-1.42</td>
<td>-1.61,</td>
<td>P&lt;0.001</td>
<td>-0.92 -1.06, -0.77 P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>9.65</td>
<td>9.46,</td>
<td>P&lt;0.001</td>
<td>9.69 9.52, 9.87 P&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>
7.6 DISCUSSION

In this chapter attempts have been made to determine whether the HOS case-mix methodology developed by the HAL group could be adapted and used with the ALSWH data. The alternative HOS Case-Mix modelling approach suggested by Selim et al (2007) was seen as more applicable for the ALSWH data. Thus the Case-Mix methodology was considered and applied here in a number of different ways to the method’s original intention. The basis of the Case-Mix methodology to modelling deaths and combining with being alive and in Same or Better health was seen as a useful approach to accounting for deaths in longitudinal studies of elderly populations. Further, once the predicted probabilities of being Alive and in Same or Better health in three years are generated at each time point, those who have died are “reinstated” and “carried forward” in the analyses as a zero probability.

As mentioned throughout this chapter there were two cut-offs for Same or Better considered. Using either cut-off will provide slightly different regression estimates, however the end result appears to be similar. That is, the impact of deaths on the diabetes groups over time cannot be ignored. The decision to choose a different cut-off was so that comparisons could be made with the original cut-off suggested by Rogers et al (5.66). Although Rogers (personal correspondence, July 2009) criticised the use of the cut-off value of 8.65 it has still been a “useful educational exercise” in developing a different cut-off. Since the ALSWH PCS has been developed and standardised to an ageing Australian population of women, then attempts to develop a different cut-off are justifiable.
There has also been discussion as to whether different combinations of Better versus Same versus Worse could be applied. Ordered logistic regressions were conducted and some results were shown suggest that the choice of Better versus Same/Worse is worthwhile considering, but with an ageing population the expectation of health improving is less likely. The results from the ordered logistic regression cut-offs (Better versus Same versus Worse) suggest that the majority of elderly women are classified as having the Same health (see Table 7.13). Another consideration may be to focus on Worse versus Same/Better however this could be calculated from the structure already in place (i.e. Same or Better versus Worse).

The strengths and the limitations of this case study using the conditional probability of being Alive and in the Same or Better health in 3 years will be discussed in the Summary Chapter 8. Also, in the final chapter recommendations about the possible and plausible usage of this method will be discussed. There will also be suggestions about future research which could be developed as a result of this method.
CHAPTER 8:

Summary
Strengths and Limitations
Recommendations
Future Research
Conclusions
### List of Abbreviations used in Chapter 8

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSWH</td>
<td>Australian Longitudinal Study on Women’s Health</td>
</tr>
<tr>
<td>APASB</td>
<td>(Australian) Probability of being Alive and in Same or Better health in three years</td>
</tr>
<tr>
<td>APASBD</td>
<td>(Australian) Probability of being Alive and in Same or Better health in three years (Includes deaths)</td>
</tr>
<tr>
<td>APCT</td>
<td>(Australian) Probability of being healthy in three years</td>
</tr>
<tr>
<td>APCTD</td>
<td>(Australian) Probability of being healthy in three years (includes deaths)</td>
</tr>
<tr>
<td>BMI</td>
<td>Body Mass Index</td>
</tr>
<tr>
<td>HALex</td>
<td>Health and Activity Limitation Index</td>
</tr>
<tr>
<td>HR-QOL</td>
<td>Health Related Quality of Life</td>
</tr>
<tr>
<td>MCS</td>
<td>Mental Component Summary score (derived from SF-36)</td>
</tr>
<tr>
<td>MCTD</td>
<td>Probability of being mentally healthy in three years (includes deaths)</td>
</tr>
<tr>
<td>PCS</td>
<td>Physical Component Summary score (derived from SF-36)</td>
</tr>
<tr>
<td>PCT</td>
<td>Probability of being healthy in two years</td>
</tr>
<tr>
<td>PCTD</td>
<td>Probability of being healthy in two years (includes deaths)</td>
</tr>
<tr>
<td>PCTDI</td>
<td>Probability of being healthy in two years (includes deaths and imputed values)</td>
</tr>
<tr>
<td>QALY</td>
<td>Quality Adjusted Life Years</td>
</tr>
<tr>
<td>SF-6D</td>
<td>Short Form - 6 Domains</td>
</tr>
<tr>
<td>SF-36</td>
<td>Short Form 36</td>
</tr>
<tr>
<td>SF-36 v1</td>
<td>Short Form 36 Version 1</td>
</tr>
<tr>
<td>WinBUGS</td>
<td>Windows - Bayesian inference using Gibbs Sampling</td>
</tr>
</tbody>
</table>
INTRODUCTION

This chapter will summarise the four methods which were applied in Chapters 4, 5, 6, and 7. In each of the four method sub-sections the strengths and limitations will be discussed. Recommendations will be made about each method along with suggestions for future research and further applications of these methods. In the final section, conclusions about the statistical methodologies explored in this thesis will be presented.

8.1 METHOD ONE

Transforming the SF-36 to account for death in longitudinal studies with three year follow-up

In Chapter 4 a method was adapted from Diehr et al (2003) which transformed the SF-36 PCS into a probability of being healthy in the future. Logistic regression was used to derive a new equation to estimate the probability of being ‘healthy’ in three years as a function of the current PCS score. To account for those women who died, the APCT score is set to zero at all time points after their death, as there is no chance they will be healthy at any future follow-up. The set of transformed scores that include values for death are labelled APCTD.

8.1.1 Strengths and Limitations

Revicki et al (2001) and Kazis et al (2007) suggest that conceptually the original approach presented by Diehr et al (2003) is reasonable and it has a valid interpretation. A limitation of the methodology in Chapter 4 is that no imputation of missing data was conducted. Diehr et al (2001) argued for the superiority of their PCTDI measure that included deaths and also imputed scores when PCS are missing for reasons other than
death. People who died may have too much influence on the PCTD and missing data may not be missing at random. In the case study presented in Chapter 4, women who had missing data (but were not known to be dead) tended to be in poorer health at the previous survey than other respondents. This highlighted the need to address the issue of missing data in longitudinal data analyses. At the time consideration was not given to missing data even though Diehr had argued for their imputation approach. However, their approach was too simplistic and not considered sufficient.

The ALSWH elderly cohort has a number of strengths that can be drawn upon for this case study. Diehr et al (2003) suggest that a limitation of previous studies was that the response rates were low. One of the strengths of the ALSWH is having a national representative sample of community dwelling women with high retention rates at follow-up surveys. The women in the ALSWH were also in better health than the United States Veterans.

Finally, the purpose of developing and validating the ALSWH transformation method was for the following reasons. Firstly there was uncertainty of the effect of the length of time between follow-up surveys on the transformed scores (issue raised by Diehr et al 2001). Also, previous studies questioned whether investigators should calculate their own transformations for their specific datasets (Diehr et al 2001). The findings of this case study provide evidence to support this point when the follow-up period is not one year. Secondly, the ALSWH PCS has been standardised using an Australian population of older women, making the Diehr equation less applicable. The results in Chapter 4 suggest that the ALSWH transformation is justifiable because of these differences.
8.1.2 Recommendations

For longitudinal studies involving the SF-36 in which subjects have died, the findings from Chapter 4 support the recommendation made by Diehr et al (2003) that both the PCS and the PCTD should be analysed to examine the influence of deaths on the study conclusions. Further, when follow-up surveys are not conducted annually, or other population norms are used to calculate the PCS, then the study data should be used to derive empirical parameters for the transformation. However, if the aim of a research exercise is to make comparisons across studies (with equivalent time frames) then using the same parameters for the transformation would be preferable.

Note that this approach for accounting for death is not restricted to the PCS. It could also be applied to the SF-36 Mental Health Summary Score (MCS) and the eight subscales (such as: Physical Functioning). This transformation approach may be useful for studies dealing with other continuous outcomes (such as body mass index (BMI)) that have lost participants over time to death. Researchers could generate the “probability of having healthy weight” at the next study in three years. Participants who are considered obese would have a low probability scores (0-100). That is, if a participant was considered morbidly obese and died before the next survey the participant would have scored a low probability at Survey 1 then since died would score 0 at Survey 2.

8.1.3 Future Research

The results from Chapter 4 shown in Tables 4.5 and 4.6 suggest that there were missing data for both groups at the two surveys. Missing data are discussed in Chapter 3 and as noted, are a complex issue. The main purpose of Tables 4.5 and 4.6 was to highlight the differences between groups and within groups when deaths are
considered. However, due to the structure of paired t-tests only complete cases were used in these analyses. Thus in Chapter 5 the analyses was extended to include data from Surveys 3 and 4 and through the use of longitudinal analysis methods explored approaches that account for missing data and the complex nature of why data may have been missing.

In summing up, Kazis et al (2007) mentioned the plotting the area under the curve of PCTD and MCTD from baseline to follow-up as an addition use of the PCTD. They suggested that this could be interpreted as change in years of healthy life. Applying this idea to more time points is also a possibility with the ALSWH data and will be considered for future research.

Finally, the focus of this thesis has been the SF-36 PCS (physical health). In future other attempts to generate transformations may include the MCTD (mental health) and the SF-36 eight sub-scales.

8.2 Method Two

Accounting for deaths in longitudinal studies – Generating a PCS value for death using a health utility index (SF-6D)

In Chapter 6, a method presented by Trisolini et al (2002) and Trisolini et al (2005) was adapted to generate a PCS value for death. The original method regressed the HALex utility score and PCS values to extrapolate a PCS value for death when the HALex index was zero. In Chapter 6, using the ASLWH data, the SF-36v1 was firstly converted to SF-6D at surveys 1 to 4. The resulting PCS death values that were extrapolated from the regression analyses were then used in longitudinal data analysis.
Consideration was given to the “missingness” of the data and so MAR and MNAR random effects mixed models were compared. Other comparisons such as complete case for PCS and PCSD were also considered for the diabetes case study.

8.2.1 Strengths and Limitations

Converting the SF-36 into the SF-6D opens up a number of options for those using the SF-36. Firstly, the SF-6D can be regressed against the PCS or even MCS to extrapolate a death value. Secondly, the newly derived SF-6D can be used for health economic evaluations since a health utility score becomes available.

A limitation of this SF-6D approach, as is the case for the HALex approach, is the actual extrapolation. The floor (lowest) value (0.301) of the SF-6D was constructed using the work by Brazier et al (1998). This result meant that the linear relationship used had a slightly further distance to extrapolate than the HALex, where the floor value was 0.1. However, results from the observed data suggest that the value of 9 was reasonable or even a value of 7 and 5 for death. However, the use of a linear relationship between SF-6D and PCS appeared to be more appropriate as discussed in Chapter 6.

8.2.2 Recommendations

Trisolini et al (2002) suggested that their “method is innovative, but as such it should be further tested and validated in studies conducted by other researchers”. They also recommended that the method be considered “using a range of other data sets and utility measures”. As has been stated in Chapter 6, there were limitations to using the HALex with the ALSWH data so the SF-6D utility index was seen as an acceptable alternative. The SF-6D comes highly recommended by Fryback (2003).
In Chapter 6, two approaches were considered for the PCS value for death. The first option was to use a constant PCS value of 9 over the time points. The second approach was to consider the value of death as a decreasing value from 7 to 5. Trisolini et al (2005) suggest a value of 14 which they applied to a single follow-up when deaths had occurred. This thesis has considered more follow-ups and also raised the question of whether it should remain a constant value. The second approach of a declining death score is an interesting possibility. A value below 10 and not 14 (i.e. 9 or 7 or 5) also seems reasonable since the lowest obtainable ALSWH PCS is 10.62 (see Section 6.5.5). As a recommendation for the use of this method to dealing with deaths in longitudinal analysis, it would be worth pursuing both a constant and declining PCS death value and then comparing results using ALSWH data and/or a different population.

Finally, if another longitudinal study of an elderly population intended to use this approach, then the recommendation would be to apply the method to their own data. That is, determine the PCS value(s) for death based on the standardised PCS that was initially used in their study.

8.2.3 Future Research

The next step would be to validate this approach in a similar way to that used in Chapter 4. Also, the ALSWH data for Survey 5 has recently been made available so the fifth wave of data would be included and further longitudinal analyses conducted. It would be interesting to determine whether the PCS death value using observed PCS and SF-6D for the survivors’ decreases below a value of 5.

The conversion of the SF-36 to the SF-6D may also open more options for answering research questions for the ALSWH data. Although it was not an initial intention of this
thesis, converting the SF-36 and then applying the SF-6D may be a possibility for future research for the ALSWH data. As Brazier et al (2007) suggest “the SF-6D allows the analyst to obtain quality adjusted life years (QALYs) from the SF-36 for use in cost utility analysis”. Therefore, there is an opportunity for greater research scope (i.e. QALYs) with the analysis of ALSWH data.

8.3 Method Three

Accounting for deaths in longitudinal studies – Using the adapted HOS Methodology (Same or Better Vs Worse)

In Chapter 7, the HOS case-mix method was adapted from work by Rogers et al (2004) and Selim et al (2007). The aim of this chapter is to assess the effects of using the adapted HOS methodology to account for deaths in the ALSWH data. Firstly, an ALSWH cut-off point for Same or Better Vs Worse was determined. Then a conditional probability of being alive and in same or better health was developed. This method differs in a number of ways from the original approach and the reasons have been detailed in Chapter 7.

The new outcome variable was known as APASB. As was the case in Chapter 4, those women who died were reinstated with a value of zero (APASB). Longitudinal analyses were conducted and consideration was given to the “missingness” of the data and so MAR and MNAR random effects mixed models were compared. Other comparisons such as complete case for APASB and APASBD were also considered for the diabetes case study.
8.3.1 Strengths and Limitations

Ware and colleagues have suggested a longer follow-up period may be required to identify clinically important differences in PCS (Ware et al 1996b; Trisolini et al 2002). Trisolini et al (2002) suggested that extending to four years follow-up period may be useful to reflect clinical quality of care in average health status outcomes. However, both research groups have mostly focused on changes from baseline and just one 2 year follow-up. The adapted approach discussed in Chapter 7 has not only used a longer follow-up period to the original method but has extended to more than two time points.

In Chapter 4, comparisons between the Diehr transformation equation and the ALSWH transformation equation were made since follow-up times were different. In Chapter 6, this was not possible since the transformations equations for deaths and Same or Better were not comparable to the earlier work by Rogers et al (2004) and Selim et al (2007).

The ALSWH data for the older cohort provided an opportunity to follow-up on the HOS Case-Mix methodology and to extend the possible ways to account for deaths in longitudinal studies. In the white paper, Kazis et al (2007) highlighted the HOS Case-Mix methodology as a useful approach to account for deaths. Kazis et al (2007) suggested that “it not only allows for deaths to be incorporated in health status it provides a separate model for death. Thus, allowing the user of the method to adjust for covariates associated with death separately to those associated with changes in health status”.

The adapted ALSWH approach used in Chapter 7 has similarities to the transformation equation developed in Chapter 4. However, this adapted ALSWH approach involves more covariates and more than one transformation equation (conditional probabilities). The ALSWH approach also considers the imputation of missing covariate data in the early stages of development. Also, it is the use of multiple imputation in the development of the transformation equations for death and Same or Better that has extended the HAL groups usage of the conditional probability of being alive and in Same or Better health. The ALSWH approach has also extended the time frame from a single two year change to multiple three year time periods.

### 8.3.2 Recommendations

From the results in Chapter 7, there is merit in this adapted ALSWH approach. As has been seen in analyses conducted in Chapter 4, failing to account for deaths in analyses may result in misleading differences in groups over time, especially in ageing populations.

### 8.3.3 Future Research

It would be worthwhile pursuing this approach, and the next step would be to validate this approach in a similar way to that used in Chapter 4. Also, since the next wave of ALSWH data is available these data would be included. Although only briefly considered in Chapter 7, there is scope for further consideration of separating Better and Same and analysing with the next wave of data using possibly ordered logistic regression.
8.4 **METHOD FOUR**

**Analysing longitudinal changes in health related quality of life: A method to adjust for longitudinal missing data**

To extend on the Chapter 4 method, in Chapter 5 Fully Bayesian random effects models were developed to account for missing outcome data (APCTD) which may be missing at random (MAR) or missing not at random (MNAR). The models were used to compare the longitudinal changes in health-related quality of life for older women with and without diabetes.

**8.4.1 Strengths and Limitations**

The results from the diabetes case study in Chapters 4 and 5 suggest that imputation of missing data for reasons other than death is important and may reduce error. Simply using complete case analysis resulted in higher coefficients for the Existing cases of diabetes groups. Comparing the results for the coefficients in the models (complete case and both imputed models) suggests that the difference in health outcomes between women with No diabetes and Existing cases becomes larger when imputed APCTD values are included. The results highlight improvements, such as the increase in sample size due to imputation resulting in greater precision of the coefficient estimates and less biased Least Squares Means (when adjusted for all other covariates). Although MNAR cannot truly be determined (since it is dependent on unobserved data), this approach has attempted to consider non-ignorable missingness and compares the MAR and MNAR random effects models. The results from the MAR and MNAR imputed models are similar, in general the MNAR model shows tighter confidence limits which is possibly due to the addition of the longitudinal logistic modeling used. The results from the two complete case models justify the need to
impute values for women who died. Not imputing values for APCTD may have an impact on conclusions regardless of whether we assume MAR or MNAR.

It would have been reasonable to expect to find that mean APCTD for the diabetes (sicker) group would have been significantly lower than APCTDI values at each survey because of the expected higher number of deaths in the group with diabetes. The introduction of zeros for deaths has naturally caused the mean values to become lower. For survey 1 mean APCTD and APCTDI were similar, while at surveys 2, 3 and 4 there appears to be an upward shift in mean APCTDI compared to APCTD for the group with diabetes. These results highlight the significance of longitudinal multiple imputation, especially as more and more elderly participants drop out over time.

A limitation of this WinBUGS approach is the computational time required to run models with very large datasets. In Chapter 5 all available data was used in the modelling which meant that the data set was quite large. All available data were used to simulate a real world situation in order to push the limits of the program. A possible way to increase computing time speed would be to consider random samples from the whole data. Another limitation of this approach was the use of only baseline covariates that had complete data. The approach was not considered for missing covariates, only missing outcomes because the main interest was to evaluate the impact of including an outcome value (0) for death.

The WinBUGS approach to account for missing data was also applied in Chapter 6. In this chapter, a PCS value was calculated for death. Thus as mentioned the value(s) for death were imputed and then the WinBUGS method was applied to account for other missing PCS values. The results shown in Chapter 6, Tables 6.9 and 6.10 report similar findings to those mentioned above from Chapter 5, that is the β coefficients are similar in the MAR and MNAR models.
8.4.2 Recommendations

This study has highlighted the importance of assessing the impact of missing data, in this case by using longitudinal multiple imputation. It appears that the “missingness” mechanics (ignoreable or non-ignoreable) may not be a major concern for this specific case study. One reason being that those who have died were more likely to be non-ignoreable missing and they have been reinstated with either a value of zero when using APCTD or a PCS value of 9, 7 or 5 when using the outcome PCS. Thus when the MNAR and MAR models are conducted the remaining missing outcomes are mostly intermittent missing data or drop-outs which no longer include deaths. Although this may be a criticism of the MNAR model, it has at least provided future research with a useful starting point to modelling when there is the possibility non-ignoreable missing.

In general, failure to consider the patterns of missingness and deal with the effects of missing data may lead to less reliable estimates of changes in HR-QoL particularly among an ageing cohort over the course of a longitudinal study. Finally, it is strongly recommended that an assessment of MAR and MNAR is made when dealing with missing outcome data in longitudinal settings where the amount of missing data is likely to be increasing over time. It would also be of benefit to use a longitudinal approach to imputing missing data.

Kazis et al (2007) suggest that if the HOS was changed in the future to include more than two waves of data at follow-up then they recommend the use of the HALex and the probability of being healthy in the future. This thesis has throughout attempted to adapt and apply such methods to a real longitudinal study (ALSWH). Thus further recommendations are the use of the transformation of PCS (PCTD) and the use of the SF-6D and the adapted case-mix approach which were discussed in Chapters 4, 6 and
7. To adjust results due to the reinstatement of deaths, multiple imputation should also be considered. The Fully Bayesian approach using WinBUGS discussed in Chapter 5 is strongly recommended, or as an alternative the use of other more user friendly MI approaches such as Stata’s *ice* procedure.

### 8.4.3 Further Research

This Fully Bayesian method could be extended to consider missing covariates. The method has been applied to real datasets in order to get a sense of the difficulties which may be encountered by an end-user of this method. The main concern with using WinBUGS with large longitudinal datasets (that would assume MNAR) is the computer processing speed required and the ability to converge (especially if missing covariates are considered). In later analyses smaller datasets could be used for simulation purposes.

Mentioned above was the disadvantage of the MNAR model for longitudinal multiple imputation. Possible future research may be to apply both the MAR and MNAR models where the outcome is PCS and consider the types of drop-out such as: frailty and death. Rather than impute a PCS value for death (as in Chapter 6) consideration may be given to what occurs when the longitudinal logistic regression model used in the MNAR model allows for most types of drop-out. Simulations using this extended approach with smaller datasets and the allowance of missing time-dependent covariates could also be considered for future research.
8.5 CONCLUSIONS

This thesis has explored three possible ways to account for deaths and a method for dealing with missing data which may be considered MNAR. The methods for accounting for deaths have been adapted from the current methods that are considered in the literature. This thesis is an attempt to address the issues and provide further advancements and more validation of the current methods. The main purpose of this thesis was to provide researchers with possible ways to account for the increasing number of deaths in longitudinal analyses. At this point in time none of the current methods have been applied to more than two waves of data. This thesis has adapted three methods to a completely new setting and applied these methods to a wider time period (3 years) and more time points (4 surveys). This body of work has also extended a Fully Bayesian method which can consider the missingness mechanisms of MAR and MNAR and longitudinal missing data. Therefore without recommending a single method to account for deaths it would be more appropriate to conclude that these methods are further additions to possible solutions to dealing with the problem of deaths in longitudinal studies. There is no simple solution to dealing with deaths in longitudinal studies.

Finally, the results from the various methods suggest that using complete case analysis will bias results. Therefore, it is essential in longitudinal studies of elderly populations that at least one of the suggested approaches be adapted and applied to the data. Simply reporting the results of the analyses which excludes deaths will soon be considered in the literature as inadequate. As a prediction for the future, there will be a growing expectation for the reporting of the statistical methods for dealing with deaths and missing data in longitudinal studies where the focus is the change in HR-QOL of elderly populations.
REFERENCES
References


Bradley C, Plowright R, Stewart J, et al. (2007) The Diabetes Treatment Satisfaction Questionnaire change version (DTSQc) evaluated in insulin glargine trials shows greater responsiveness to improvements than the original DTSQ. Health and Quality of Life Outcomes 5(57).


References


References


Trisolini MG, McCall NT, Pope GC, et al. (2002) Evaluating the Two-Year Follow-up Health Status of Medicare Fee-For-Service Beneficiaries Using the Health Outcomes Survey-Final Report.


Ware JE. (1993) SF-36 Health Survey Manual and Interpretation Guide. *The Health Institute, New England Medical Center Boston*.


Ware JE, Bayliss M, Rogers W, et al. (1996b) Differences in 4-year health outcomes for elderly and poor, chronically ill patients treated in HMO and fee-for-service systems. Results from the Medical Outcomes Study JAMA. 276(13):1039-47.


WinBUGS1.4.1. (2003) with DoodleBUGS. 1.4 Imperial College & Medical Research Council UK.


APPENDIX 2.1

SAS code for calculating ALSWH PCS and MCS:

/*The first step is to compute Z-scores. The observed values are the sample data. Mean and Standard Deviation is from the Australian general population factor analytic sample (Old age group – 65+). Figures were taken from Stevenson (1996, Pg 26) */

PF_ZA=(PF-57.3)/28.8;
RP_ZA=(RP-56.0)/42.8;
BP_ZA=(BP-65.4)/28.6;
GH_ZA=(GH-61.1)/22.4;
VT_ZA=(VT-57.4)/21.4;
SF_ZA=(SF-77.3)/27.7;
RE_ZA=(RE-72.1)/37.0;
MH_ZA=(MH-75.3)/17.3;

/* Sample raw factor scores are now computed. The Z scores are used from the code in the box above. Scoring coefficients are from Australian women aged 70-74 (WHA). */

prawa=(PF_ZA *.44) + (RP_ZA * .41) + (BP_ZA * .32) + (SF_ZA *.02) + (MH_ZA * -.31) + (RE_ZA * .004) + (VT_ZA * -.08) + (GH_ZA * .14);
mrawa=(PF_ZA * -.20) + (RP_ZA * -.17) + (BP_ZA * -.08) + (SF_ZA * .24) + (MH_ZA * .54) + (RE_ZA * .22) + (VT_ZA * .34) + (GH_ZA * .11);

/* Standardised scores are computed from Australian data. */

PCSA = (prawa*10) + 50;
MCSA = (mrawa*10) + 50;

/* A second set of Z scores are calculated. The observed values are still the sample data but the mean and standard deviation is derived from the WHA population. */

FF_ZB=(FF-63.35)/25.94;
RP_ZB=(RP-57.39)/43.24;
BP_ZB=(BP-65.09)/26.68;
## APPENDIX 2.2

### Table A2.1 Characteristics of the instruments discussed Table 1 of Fryback (2003)

<table>
<thead>
<tr>
<th>Item</th>
<th>SF-36v2</th>
<th>WHOQOL-BREF</th>
<th>QWB-SA</th>
<th>HUI2/3</th>
<th>EQ-5D</th>
<th>SF-6D</th>
</tr>
</thead>
<tbody>
<tr>
<td>National origin</td>
<td>USA</td>
<td>International</td>
<td>USA</td>
<td>Canada</td>
<td>Europe</td>
<td>USA</td>
</tr>
<tr>
<td>Individual questions in the instrument</td>
<td>36</td>
<td>26</td>
<td>73</td>
<td>15 or 40*</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Alternative dimensional structures</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Structure 1: Dimensions to represent health state</td>
<td>(Summated scales) 8</td>
<td>4</td>
<td>4</td>
<td>(HUI2) 7</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Structure 2: Dimensions to represent health state</td>
<td>(Factor scores) 2</td>
<td>-</td>
<td>-</td>
<td>(HUI3) 8</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Single summary scalar score to represent overall health?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes**</td>
</tr>
</tbody>
</table>

*Length depends on mode of administration: 15 item self administered questionnaire or 40 item branching interviewer-administered questionnaire.

** In theory, this will yield a single score, but as yet there is no final score among several alternative functions that have been derived.
## APPENDIX 2.3

Table A2.2  from Fryback (2003) Nominal content of the six instruments

<table>
<thead>
<tr>
<th>SF-36v2</th>
<th>WHOQOL-BREF</th>
<th>QWB-SA</th>
<th>HUI2</th>
<th>HUI3</th>
<th>EQ-5D</th>
<th>SF-6D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Physical function</td>
<td>* Pain and discomfort</td>
<td>* Sleep and rest</td>
<td>* Energy and fatigue</td>
<td>* Mobility</td>
<td>* Activities of daily living</td>
<td>* Dependence on medicinal and medical aids</td>
</tr>
<tr>
<td>Mental Health</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Vitality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>* Social functioning</td>
<td>* Role limitation due to emotional functioning</td>
<td>* mental health</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Psychological
- * positive feelings
- * Thinking, learning, memory, and concentration
- * Self-esteem
- * Bodily image
- * negative feelings
- * Spirituality

### Social Relationships
- * Personal relationships
- * Social Support
- * Sexual activity

### Environment
- * Freedom, physical safety
- * Home environment
- * Financial resources
- * Access to health and social care
- * Opportunities to acquire new information and skills
- * Participation & opportunities for recreation and leisure
- * Physical environment (pollution, noise, traffic, climate)
- * Transport

Overall health Assessment
### APPENDIX 4.1


<table>
<thead>
<tr>
<th>“Healthy”</th>
<th>P-Alive</th>
<th>P-E/NG/G</th>
<th>P-Top 75%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td>PCS</td>
<td>1.36</td>
<td>0.0496</td>
<td>-3.77</td>
</tr>
<tr>
<td>MCS</td>
<td>1.97</td>
<td>0.0178</td>
<td>-3.44</td>
</tr>
<tr>
<td>Role-emotional</td>
<td>2.38</td>
<td>0.0078</td>
<td>-1.28</td>
</tr>
<tr>
<td>General health</td>
<td>1.80</td>
<td>0.0240</td>
<td>-4.14</td>
</tr>
<tr>
<td>Mental health</td>
<td>2.22</td>
<td>0.0086</td>
<td>-3.03</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>2.16</td>
<td>0.0125</td>
<td>-2.17</td>
</tr>
<tr>
<td>Physical function</td>
<td>1.81</td>
<td>0.0238</td>
<td>-2.28</td>
</tr>
<tr>
<td>Role-physical</td>
<td>2.41</td>
<td>0.0150</td>
<td>-1.11</td>
</tr>
<tr>
<td>Social function</td>
<td>1.94</td>
<td>0.0150</td>
<td>-2.62</td>
</tr>
<tr>
<td>Vitality</td>
<td>2.05</td>
<td>0.0191</td>
<td>-2.71</td>
</tr>
</tbody>
</table>

a = intercept; b = slope.
APPENDIX 4.2

Figures A4.2a-c Distributions of PCS at Survey 1-3

Figures A4.3a-c Distributions of PCT (Diehr Equation) at Survey 1-3

Figures A4.4a-c Distributions of PCTD (Diehr Equation) at Survey 1-3
Figures A4.5a-c Distributions of APCT (ALSWH Equation) at Survey 1-3

Figures A4.6a-c Distributions of APCTD (ALSWH Equation) at Survey 1-3
APPENDIX 4.3

Table A4.1 Logistic model—goodness-of-fit test

<table>
<thead>
<tr>
<th>Table collapsed on quantiles of estimated probabilities</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>number of observations =</td>
<td>9676</td>
</tr>
<tr>
<td>number of groups =</td>
<td>10</td>
</tr>
<tr>
<td>Hosmer-Lemeshow chi2(8) =</td>
<td>9.08</td>
</tr>
<tr>
<td>Prob &gt; chi2 =</td>
<td>0.3360</td>
</tr>
</tbody>
</table>

Table A4.2 Logistic Regression – Diagnostic test

<table>
<thead>
<tr>
<th></th>
<th>Healthy</th>
<th>Not Healthy</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>4454</td>
<td>678</td>
<td>5132</td>
</tr>
<tr>
<td>Negative</td>
<td>2231</td>
<td>2313</td>
<td>4544</td>
</tr>
<tr>
<td>Total</td>
<td>6685</td>
<td>2991</td>
<td>9676</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pr( + D)</th>
<th>Pr( ~D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>66.63%</td>
<td>77.33%</td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>Pr( D +)</td>
<td>86.79%</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>Pr(~D -)</td>
<td>50.90%</td>
</tr>
<tr>
<td>False + rate for true ~D</td>
<td>Pr( +~D)</td>
<td>22.67%</td>
</tr>
<tr>
<td>False - rate for true D</td>
<td>Pr( - D)</td>
<td>33.37%</td>
</tr>
<tr>
<td>False + rate for classified +</td>
<td>Pr(~D +)</td>
<td>13.21%</td>
</tr>
<tr>
<td>False - rate for classified -</td>
<td>Pr( D -)</td>
<td>49.10%</td>
</tr>
<tr>
<td>Correctly classified</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>69.94%</td>
<td></td>
</tr>
</tbody>
</table>
In Figure A4.8 the plot of $\Delta X^2$ influence statistic against predicted probability shows that there are a few points in the top right hand corner (corresponding to large values of $\Delta X^2$). This indicates that these points are fitted poorly by the model.

In Figure A4.9 the plot of $\Delta D$ influence statistic and predicted probability shows that there are points that are not fitted well by the model. The plot shows quite a number of points...
falling above 4 (approx $\chi^2 = 3.84$, 1 degree of freedom) which suggests the model may not be a good fit.

**Figure A4.9 the plot of $\Delta$D influence statistic and predicted probability**

In Figure A4.10 the plot of Pregibon $\Delta$ influence statistic against predicted probability shows that, quite a few points lie away from the majority of the data.
Examination of the covariates shows that some of the covariate patterns may be expected to induce changes in the estimated coefficients if removed from the data.

```
list p_evgg2 ps sd2 sddev sbeta shat if sd2>4 & sd2~=. count if sd2>4 & sd2~=.
```

```
+-------------------------------------------------------------+  
<table>
<thead>
<tr>
<th>p_evgg2</th>
<th>ps</th>
<th>sd2</th>
<th>sddev</th>
<th>sbeta</th>
<th>shat</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>0</td>
<td>.9178507</td>
<td>11.17495</td>
<td>4.999324</td>
<td>.0019904</td>
</tr>
<tr>
<td>3.</td>
<td>0</td>
<td>.8061848</td>
<td>4.160176</td>
<td>3.282191</td>
<td>.0006227</td>
</tr>
<tr>
<td>16.</td>
<td>0</td>
<td>.8264695</td>
<td>4.763417</td>
<td>3.503349</td>
<td>.0007411</td>
</tr>
<tr>
<td>18.</td>
<td>0</td>
<td>.8461744</td>
<td>5.501759</td>
<td>3.744777</td>
<td>.00089</td>
</tr>
<tr>
<td>21.</td>
<td>0</td>
<td>.8942514</td>
<td>8.457873</td>
<td>4.494169</td>
<td>.001483</td>
</tr>
<tr>
<td>29.</td>
<td>0</td>
<td>.9095209</td>
<td>4.077873</td>
<td>2.223149</td>
<td>.0014491</td>
</tr>
<tr>
<td>49.</td>
<td>0</td>
<td>.8638722</td>
<td>6.347098</td>
<td>3.988989</td>
<td>.0010619</td>
</tr>
<tr>
<td>60.</td>
<td>0</td>
<td>.8813661</td>
<td>7.42631</td>
<td>4.263149</td>
<td>.0012795</td>
</tr>
<tr>
<td>115.</td>
<td>0</td>
<td>.8846563</td>
<td>7.671063</td>
<td>4.320425</td>
<td>.0013282</td>
</tr>
<tr>
<td>128.</td>
<td>0</td>
<td>.9294148</td>
<td>13.16961</td>
<td>5.302809</td>
<td>.0023354</td>
</tr>
<tr>
<td>388.</td>
<td>0</td>
<td>.9249233</td>
<td>12.3219</td>
<td>5.17941</td>
<td>.0021916</td>
</tr>
<tr>
<td>394.</td>
<td>0</td>
<td>.8715192</td>
<td>6.784414</td>
<td>4.104647</td>
<td>.0011505</td>
</tr>
<tr>
<td>407.</td>
<td>0</td>
<td>.8315105</td>
<td>4.935863</td>
<td>3.562323</td>
<td>.0007756</td>
</tr>
<tr>
<td>411.</td>
<td>0</td>
<td>.8200039</td>
<td>4.556375</td>
<td>3.430167</td>
<td>.0007</td>
</tr>
<tr>
<td>416.</td>
<td>0</td>
<td>.8638487</td>
<td>6.345835</td>
<td>3.988645</td>
<td>.0010616</td>
</tr>
</tbody>
</table>
+-------------------------------------------------------------+
```
APPENDIX 5.1

WinBUGS code:

**Model 1 Complete Case**

model
{
    for (i in 1:5451) {  # 5451 women, indexed by i

        # APCTD, model of interest
        for (s in 1:4){
            cut.apctd[i,s]<-cut(apctd[i,s]);
            apctd[i,s]<-dnorm(apctd.mu[i,s],apctd.tau);
            apctd.mu[i,s] <- i.reg[i] + b.slo[i]*(s-2.5); # centred
        }

        # subject-level regression for APCTD model
        i.reg[i]<-b.int[i] + b.age*age[i] + b.area*area[i] + (b.diab[1]*equals(diab[i],2))
        +(b.diab[2]*equals(diab[i],1)) + b.com*comorb[i] +b.smok*smoke[i];

        b.int[i] ~ dnorm(b0.mu, b0.tau); # random subject intercept
        b.slo[i] ~ dnorm(b1.mu, b1.tau); # random time slope
    } # end if i loop

    # priors for regression parameters in model of interest
    b.diab[1] ~ dnorm(0.0,0.01E-4);
    b.diab[2] ~ dnorm(0.0,0.01E-4);
    b.com ~ dnorm(0.0,0.01E-4);
    b.age ~ dnorm(0.0,0.01E-4);
    b.area ~ dnorm(0.0,0.01E-4);
    b.smok ~ dnorm(0.0,0.01E-4);

    # random intercept
    b0.mu ~ dnorm(0.0,0.01E-4);
    b0.tau ~ dgamma(0.01, 0.01); # variance of the random intercepts for each individual

    # random slope
    b1.mu ~ dnorm(0.0,0.01E-4);
    b1.tau ~ dgamma(0.01, 0.01);

    # precision
    apctd.tau ~ dgamma(0.01, 0.01);

    # Variance = 1/precision
    apctd.sig <- sqrt(1/apctd.tau); # variance of APCTD distribution
    apctd.var <- 1/apctd.tau;
    b.sig <- sqrt(1/b0.tau); # variance of the random intercepts for each individual
    b.var <- 1/b0.tau;
    rho <- b.var/(b.var+apctd.var);
}
Model 2 Ignorable Missing Data Mechanism (MAR)

model
{
for (i in 1:8368) {  # 8368 women, indexed by i
  # APCTD, model of interest
  for (s in 1:4){
    #cut.apctd[i,s]<-cut(apctd[i,s]);
    apctd[i,s]~dnorm(apctd.mu[i,s],apctd.tau);
    apctd.mu[i,s] <- i.reg[i] + b.slo[i]*(s - 2.5);  # centred
  }
  # subject-level regression for APCTD model
  i.reg[i]<-b.int[i] + b.age*age[i] + b.area*area[i] + (b.diab[1]*equals(diab[i],2))
   +(b.diab[2]*equals(diab[i],1)) + b.com*comorb[i] +b.smok*smoke[i];

  b.int[i] ~ dnorm(b0.mu, b0.tau);  # random subject intercept
  b.slo[i] ~ dnorm(b1.mu, b1.tau);  # random time slope
}  # end if i loop

# priors for regression parameters in model of interest
b.diab[1] ~ dnorm(0.0,1.0E-4);
b.diab[2] ~ dnorm(0.0,1.0E-4);
b.com ~ dnorm(0.0,1.0E-4);
b.age ~ dnorm(0.0,1.0E-4);
b.area ~ dnorm(0.0,1.0E-4);
b.smok ~ dnorm(0.0,1.0E-4);

# random intercept
b0.mu ~ dnorm(0.0,1.0E-4);
b0.tau ~ dgamma(0.01, 0.01);
# random slope
b1.mu ~ dnorm(0.0,1.0E-4);
b1.tau ~ dgamma(0.01, 0.01);
# precision
apctd.tau ~ dgamma(0.01, 0.01);
# Variance = 1/precision
apctd.sig <- sqrt(1/apctd.tau);  # variance of APCTD distribution
apctd.var <- 1/apctd.tau;
b.sig <- sqrt(1/b0.tau);  # variance of the random intercepts for each individual
b.var <- 1/b0.tau;
rho <- b.var/(b.var+apctd.var);
}
}
Model 3 Non-ignorable Missing Data Mechanism (MNAR)

model
{
  for (i in 1:8368) {  # 8368 women, indexed by i
    # APCTD, model of interest
    for (s in 1:4){
      cut.apctd[i,s] <- cut(apctd[i,s]);
      apctd[i,s] ~ dnorm(apctd.mu[i,s],apctd.tau);
      apctd.mu[i,s] <- i.reg[i] + b.slo[i]*(s-2.5); # centred
      # NMAR Logistic model for the probability
      obs[i,s] ~ dbern(p.obs[i,s]);
      logit(p.obs[i,s]) <- a.cons[s]+0.0001*cut.apctd[i,s]+i.missing[i];
    }
    # subject-level regression for missing model
    i.missing[i]<-a.diab[1]*equals(diab[i],2)+(a.diab[2]*equals(diab[i],1)) +
      a.com*comorb[i] + a.smok*smoke[i] + a.age*age[i] + a.area*area[i];
    # subject-level regression for APCTD model
    i.reg[i]<-b.int[i] + b.age*age[i] + b.area*area[i] + (b.diab[1]*equals(diab[i],2))
      +(b.diab[2]*equals(diab[i],1)) + b.com*comorb[i] + b.smok*smoke[i];
    b.int[i] ~ dnorm(b0.mu, b0.tau); # random subject intercept
    b.slo[i] ~ dnorm(b1.mu, b1.tau); # random time slope
  }
  # priors for regression parameters in model of interest
  b.diab[1] ~ dnorm(0.0,1.0E-4);
  b.diab[2] ~ dnorm(0.0,1.0E-4);
  b.age ~ dnorm(0.0,1.0E-4);
  b.area ~ dnorm(0.0,1.0E-4);
  b.smok ~ dnorm(0.0,1.0E-4);
  a.diab[1] ~ dnorm(0.0,1.0E-4);
  a.diab[2] ~ dnorm(0.0,1.0E-4);
  a.com ~ dnorm(0.0,1.0E-4);
  a.age ~ dnorm(0.0,1.0E-4);
  a.area ~ dnorm(0.0,1.0E-4);
  a.smok ~ dnorm(0.0,1.0E-4);
  # random intercept
  b0.mu ~ dnorm(0.0,1.0E-4);
  b0.tau ~ dgamma(0.01, 0.01);
  # random slope
  b1.mu ~ dnorm(0.0,1.0E-4);
  b1.tau ~ dgamma(0.01, 0.01);
  # precision
  apctd.tau ~ dgamma(0.01, 0.01);
  # Variance = 1/precision
  apctd.sig <- sqrt(1/apctd.tau); # variance of APCTD distribution
  apctd.var <- 1/apctd.tau;
  b.sig <- sqrt(1/b0.tau); # variance of the random intercepts for each individual
}
b.var <- 1/b0.tau;
rho <- b.var/(b.var+apctd.var);
# Priors for coefficients in the logistic model
for (s in 1:4){
  a.cons[s] ~ dnorm(0.0,1.0E-4);
}
}
APPENDIX 5.2

Output examples from Model 2 Ignorable Missing Data Mechanism (MAR)
Output examples from Model 2 Ignorable Missing Data Mechanism (MAR)

b.diab[1] sample: 25000
-6.0 -4.0 -2.0 0.0
0.0
0.2
0.4
0.6

-20.0 -18.0 -16.0 -14.0
0.0
0.2
0.4
0.6

b.diab[1] lag
0 20 40
-1.0 -0.5 0.0 0.5 1.0

b.diab[2] lag
0 20 40
-1.0 -0.5 0.0 0.5 1.0

b.diab[1] iteration
6001 20000
-6.0 -4.0 -2.0 0.0 2.0

b.diab[2] iteration
6001 20000
-18.0 -17.0 -16.0 -15.0 -14.0 -13.0
APPENDIX 6.1

HALex questionnaire for calculating the Healthy People 2000 Years of Healthy Life

Adult version
Based on the Behavioral Risk Factor Surveillance Survey: CDC Telephone Administered Version.
1. Would you say that in general your health is:
   a. Excellent 1
   b. Very good 2
   c. Good
   d. Fair 4
   e. Poor 5
   Do not know/not sure 7
   Refused 9

Section A: ages 18–69 years of age
1. What were you doing most of the past 12 months?
   a. Working at a job or business 1
   b. Keeping house Go to Q. 4 2
   c. Going to school Go to Q. 6 3
   d. Something else Go to Q. 6 4
   Do not know/not sure 7
   Refused 9

2. Does any impairment or health problem now keep you from working at a job or business?
   a. Yes Go to Q. 9 1
   b. No 2
   Do not know/not sure 7
   Refused 9

3. Are you limited in the kind or amount of work you can do because of any impairment or health problem?
   a. Yes Go to Q. 9 1
   b. No Go to Q. 8 2
   Do not know/not sure 7
   Refused 9

4. Does any impairment or health problem now keep you from doing any housework at all?
   a. Yes Go to Q. 6 1
   b. No 2
   Do not know/not sure 7
   Refused 9

5. Are you limited in the kind or amount of housework you can do because of any impairment or health problem?
   a. Yes 1
   b. No 2
   Do not know/not sure 7
   Refused 9

6. Does any impairment or health problem keep you from working at a job or business?
   a. Yes Go to Q. 9 1
   b. No 2
   Do not know/not sure 7
   Refused 9
7. Are you limited in the kind or amount of work you could do because of any impairment or health problem?
   a. Yes Go to Q. 9 1
   b. No 2
   Do not know/not sure 7
   Refused 9

   If Yes to Q. 4 or Yes to Q. 5, go to Q. 9.

8. Are you limited in any way in any activities because of any impairment or health problem?
   a. Yes 1
   b. No Go to closing statement 2
   Do not know/not sure 7
   Refused 9

9. Because of any impairment or health problem, do you need the help of other persons with your personal care needs, such as eating, bathing, dressing or getting around the house?
   a. Yes 1
   b. No 2
   Do not know/not sure 7
   Refused 9

10. Because of any impairment or health problem, do you need the help of other persons in handling your routine needs, such as everyday household chores, doing necessary business, shopping or getting around for other purposes?
    a. Yes Go to closing statement 1
    b. No Go to closing statement 2
    Do not know/not sure Go to closing statement 7
    Refused Go to closing statement 9

Section B: ages 70 and older

11. Because of any impairment or health problem, do you need the help of other persons with your personal care needs, such as eating, bathing, dressing or getting around the house?
    a. Yes 1
    b. No 2
    Do not know/not sure 7
    Refused 9

12. Because of any impairment or health problem, do you need the help of other persons in handling your routine needs, such as everyday household chores, doing necessary business, shopping, or getting around for other purposes?
    a. Yes Go to closing statement 1
    b. No 2
    Do not know/not sure 7
    Refused 9

13. Are you limited in any way in any activities because of any impairment or health problem?
    a. Yes 1
    b. No 2
    Do not know/not sure 7
    Refused 9
### Appendix 6.2

#### Table A6.1

*Brazier and Roberts*  
*Medical Care* • *Volume 42, Number 9, September 2004*

<table>
<thead>
<tr>
<th>sf-6d(sf-36)</th>
<th>sf-6d(sf-12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>c (1.000)</td>
<td>c (1.000)</td>
</tr>
<tr>
<td>PF2 -0.051</td>
<td>PF23 -0.035</td>
</tr>
<tr>
<td>PF3 -0.011</td>
<td>PF4 -0.044</td>
</tr>
<tr>
<td>PF4 -0.040</td>
<td>PF5 -0.056</td>
</tr>
<tr>
<td>PF5 -0.054</td>
<td></td>
</tr>
<tr>
<td>RL2 -0.053</td>
<td>RL3 -0.061</td>
</tr>
<tr>
<td>RL3 -0.055</td>
<td>RL234 -0.053</td>
</tr>
<tr>
<td>RL4 -0.056</td>
<td>RL4 -0.054</td>
</tr>
<tr>
<td>SF2 -0.055</td>
<td>SF2 -0.057</td>
</tr>
<tr>
<td>SF3 -0.067</td>
<td>SF3 -0.059</td>
</tr>
<tr>
<td>SF4 -0.070</td>
<td>SF14 -0.072</td>
</tr>
<tr>
<td>SF5 -0.087</td>
<td></td>
</tr>
<tr>
<td>PAIN2 -0.047</td>
<td></td>
</tr>
<tr>
<td>PAIN3 -0.025</td>
<td>PAIN23 -0.042</td>
</tr>
<tr>
<td>PAIN4 -0.056</td>
<td>PAIN4 -0.065</td>
</tr>
<tr>
<td>PAIN5 -0.091</td>
<td>PAIN5 -0.102</td>
</tr>
<tr>
<td>PAIN6 -0.167</td>
<td>PAIN6 -0.171</td>
</tr>
<tr>
<td>MH2 -0.049</td>
<td></td>
</tr>
<tr>
<td>MH3 -0.042</td>
<td>MH3 -0.055</td>
</tr>
<tr>
<td>MH4 -0.109</td>
<td>MH4 -0.119</td>
</tr>
<tr>
<td>MH5 -0.128</td>
<td>MH5 -0.140</td>
</tr>
<tr>
<td>VIT2 -0.086</td>
<td></td>
</tr>
<tr>
<td>VIT3 -0.061</td>
<td></td>
</tr>
<tr>
<td>VIT4 -0.054</td>
<td>VIT4 -0.059</td>
</tr>
<tr>
<td>VIT5 -0.091</td>
<td>VIT5 -0.103</td>
</tr>
<tr>
<td>MOST -0.070</td>
<td>MOST -0.085</td>
</tr>
<tr>
<td>n 249</td>
<td>249</td>
</tr>
<tr>
<td>adj $R^2$ 0.526</td>
<td>0.472</td>
</tr>
<tr>
<td>Predictive Ability</td>
<td></td>
</tr>
<tr>
<td>MAE 0.073</td>
<td>0.074 0.079</td>
</tr>
<tr>
<td>No. &gt; 0.05</td>
<td>120</td>
</tr>
<tr>
<td>No. &gt; 0.10</td>
<td>51 58</td>
</tr>
<tr>
<td>t -1.146 -1.317 -1.647 -1.780</td>
<td></td>
</tr>
<tr>
<td>JB 0.173</td>
<td>1.169 0.964 1.650</td>
</tr>
<tr>
<td>LB 189.87</td>
<td>163.87 122.49 111.57</td>
</tr>
</tbody>
</table>

Models are estimated with White heteroscedasticity consistent standard errors.  
Figures in bold are significant at $p < 0.10$.  
MAE indicates mean absolute error.  
$t$ is a test that the mean of the prediction errors is zero.  
JB, Jarque-Bera test for normality of errors; LB, Ljung-Box test for autocorrelation of errors.
Items from ALSWH and how they match up with Brazier’s SF-6D questions.

```
o1q3a  =       'Health limit vigorous activities' (SF3)
o1q3b  =       'Health limit moderate activities' (SF4)
o1q3j  =       'Health limit bathing or dressing' (SF12)
o1q4c  =       'Physical Health/limited in type work' (SF15)
o1q5b  =       'Emotional Health/accomplish less'(SF18)
o1q7   =       'How much bodily pain'(SF21)
o1q8   =       'Pain interfered with work'(SF22)
o1q9b  =       'How often/feel nervous person'(SF24)
o1q9e  =       'How often/have a lot of energy'(SF27)
o1q9f  =       'How often/felt down'(SF28)
o1q10  =       'How much time/interfere social life'(SF32)
```

The code will convert the ALSWH SF-36 items to SF-6D index using Survey 1 data.

*The following Stata code has been taken from the Hanmer website:

This code uses the Consistent model (Column 2, Table 4) from Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care. 2004; 42: 851-859. Model 2 from Table 4.

```
sf3 = vigorous activities where 1="limited a lot" to 3="not limited at all"
sf4 = moderate activities where 1="limited a lot" to 3="not limited at all"
sf12 = bathing & dressing where 1="limited a lot" to 3="not limited at all"
sf15 = physical limited work where 1="all of the time" to 5="none of the time"
sf18 = emotional accomplish less where 1="all of the time" to 5="none of the time"
sf21 = bodily pain where 1="none" to 6="very severe"
sf22 = pain interferes with work where 1="not at all" to 5="extremely"
sf24 = nervous where 1="all the time" to 6="none of the time"
sf27 = energy where 1="all the time" to 6="none of the time"
sf28 = downhearted & blue where 1="all the time" to 6="none of the time"
sf32 = social activities where 1="all the time" to 5="none of the time"
```

gen    sf1_3   =   o1q3a
ngen   sf1_4   =   o1q3b
ngen   sf1_12  =   o1q3j
ngen   sf1_15  =   o1q4c
ngen   sf1_18  =   o1q5b
ngen   sf1_21  =   o1q7
ngen   sf1_22  =   o1q8
ngen   sf1_24  =   o1q9b
ngen   sf1_27  =   o1q9e
ngen   sf1_28  =   o1q9f
ngen   sf1_32  =   o1q10

tab sf1_15
tab sf1_18

gen rand=uniform()
gen sf1_24r=sf1_24
replace sf1_24r=2 if sf1_24==2
replace sf1_24r=2 if sf1_24==3 & rand<.5
replace sf1_24r=3 if sf1_24==3 & rand>.5
replace sf1_24r=4 if sf1_24==4
replace sf1_24r=4 if sf1_24==5
replace sf1_24r=5 if sf1_24==5
replace rand=uniform()
gen sf1_27r=sf1_27
replace sf1_27r=2 if sf1_27==2
replace sf1_27r=2 if sf1_27==3 & rand<.5
replace sf1_27r=3 if sf1_27==3 & rand>.5
replace sf1_27r=4 if sf1_27==4
replace sf1_27r=5 if sf1_27==5
replace sf1_27r=6 if sf1_27==6

replace rand=uniform()

gen sf1_28r=sf1_28
replace sf1_28r=2 if sf1_28==2
replace sf1_28r=2 if sf1_28==3 & rand<.5
replace sf1_28r=3 if sf1_28==3 & rand>.5
replace sf1_28r=4 if sf1_28==4
replace sf1_28r=5 if sf1_28==5
replace sf1_28r=6 if sf1_28==6

/*form the dimensions from Table 1 */

gen sf1_Phys=
replace sf1_Phys=1 if sf1_3==3 & sf1_4==3 & sf1_12==3
replace sf1_Phys=2 if sf1_3==2 & sf1_4==3 & sf1_12==3
replace sf1_Phys=2 if sf1_3==1 & sf1_4==3 & sf1_12==3
replace sf1_Phys=3 if sf1_4==2 & sf1_12==3
replace sf1_Phys=4 if sf1_4==1 & sf1_12==3
replace sf1_Phys=5 if sf1_12==2
replace sf1_Phys=6 if sf1_12==1

gen sf1_Role=
replace sf1_Role=1 if sf1_15==2 & sf1_18==2
replace sf1_Role=2 if sf1_15==1 & sf1_18==2
replace sf1_Role=3 if sf1_18==1
/* if sf1_18=1, then same SF-6D score were assigned regardless of the response to sf-15, so I have collapsed Role levels 3 and 4 */

gen sf1_Social=
replace sf1_Social=1 if sf1_32==5
replace sf1_Social=2 if sf1_32==4
replace sf1_Social=3 if sf1_32==3
replace sf1_Social=4 if sf1_32==2
replace sf1_Social=5 if sf1_32==1

gen sf1_Pain=
replace sf1_Pain=1 if sf1_21==1 & sf1_22==1
replace sf1_Pain=2 if sf1_21>1 & sf1_22==1
replace sf1_Pain=3 if sf1_22==2
replace sf1_Pain=4 if sf1_22==3
replace sf1_Pain=5 if sf1_22==4
replace sf1_Pain=6 if sf1_22==5

gen sf1_Mental=
replace sf1_Mental=1 if sf1_24r==5 | sf1_28r==5
replace sf1_Mental=1 if sf1_24r==4 | sf1_28r==4
replace sf1_Mental=3 if sf1_24r==3 | sf1_28r==3
replace sf1_Mental=4 if sf1_24r==2 | sf1_28r==2
replace sf1_Mental=5 if sf1_24r==1 | sf1_28r==1
*replace sf1_Mental=. if sf1_24<1 | sf1_24>5
*replace sf1_Mental=. if sf1_28<1 | sf1_28>5

gen sf1_Vital=
replace sf1_Vital=1 if sf1_27r==1
replace sf1_Vital=2 if sf1_27r==2
replace sf1_Vital=3 if sf1_27r==3
replace sf1_Vital=4 if sf1_27r==4
replace sf1_Vital=5 if sf1_27r==5

/*create MOST category if any dimension is at its worst state */
gen most1=0
/*assign decrements based on level from consistent in Brazier et al (2004) Table 4 and calculate score*/

gen pf1_1=.
replace pf1_1=0 if sf1_Phys==1
replace pf1_1=-.035 if sf1_Phys==2
replace pf1_1=-.035 if sf1_Phys==3
replace pf1_1=-.044 if sf1_Phys==4
replace pf1_1=-.056 if sf1_Phys==5
replace pf1_1=-.117 if sf1_Phys==6

gen rl1_1=.
replace rl1_1=0 if sf1_Role==1
replace rl1_1=-.053 if sf1_Role>1

gen sc1_1=.
replace sc1_1=0 if sf1_Social==1
replace sc1_1=-.057 if sf1_Social==2
replace sc1_1=-.059 if sf1_Social==3
replace sc1_1=-.072 if sf1_Social==4
replace sc1_1=-.087 if sf1_Social==5

gen pn1_1=.
replace pn1_1=0 if sf1_Pain==1
replace pn1_1=-.042 if sf1_Pain==2
replace pn1_1=-.042 if sf1_Pain==3
replace pn1_1=-.065 if sf1_Pain==4
replace pn1_1=-.102 if sf1_Pain==5
replace pn1_1=-.171 if sf1_Pain==6

gen mh1_1=.
replace mh1_1=0 if sf1_Mental==1
replace mh1_1=-.042 if sf1_Mental==2
replace mh1_1=-.042 if sf1_Mental==3
replace mh1_1=-.100 if sf1_Mental==4
replace mh1_1=-.118 if sf1_Mental==5

gen v1_1=.
replace v1_1=0 if sf1_Vital==1
replace v1_1=-.071 if sf1_Vital==2
replace v1_1=-.071 if sf1_Vital==3
replace v1_1=-.071 if sf1_Vital==4
replace v1_1=-.092 if sf1_Vital==5

gen mst1_1=.
replace mst1_1=0 if most1==0
replace mst1_1=-.061 if most1==1

gsf1_6D = 1 + pf1_1 + rl1_1 + sc1_1 + pn1_1 + mh1_1 + v1_1 + mst1_1
replace sf1_6D=. if pf1_1==. | rl1_1==. | sc1_1==. | pn1_1==. | mh1_1==. | v1_1==. | mst1_1==.
APPENDIX 6.3

Quadratic Regression Equations

As previously mentioned, the development of a death value by Trisolini et al (2005) involved the use of a quadratic model. In the graph shown by Smith et al (2005), the data were simplified to a “fitted” number of points to highlight the curvilinear shape of the relationship between PCS and SF-6D. It was decided that quadratic equations should also be examined. Using JMP8.0 the following equations were generated for surveys 1, 2, 3 and 4. Likewise in the simple linear regressions, if X=0 is substituted into the equations values for Y can be found. The quadratic equations contain a term \((X-C)^2\) and this constant \(C\) is adjusted at each survey depending on the data and the fitted equation. The \(C\) values were calculated by the statistical software JMP 8.0.

When \(SF_1 6D = 0\)

\[
\hat{Y}_{pcs1} = \alpha + \beta_1 \times SF_1 6D + \beta_2 \times (SF_1 6D - C_1)^2
\]

\[
= 9.34 + 55.85 \times SF_1 6D - 4.38 \times (SF_1 6D - 0.724)^2 = 7.04
\]

When \(SF_2 6D = 0\)

\[
\hat{Y}_{pcs2} = \alpha + \beta_1 \times SF_2 6D + \beta_2 \times (SF_2 6D - C_2)^2
\]

\[
= 6.86 + 58.10 \times SF_2 6D + 9.85 \times (SF_2 6D - 0.731)^2 = 12.11
\]

When \(SF_3 6D = 0\)

\[
\hat{Y}_{pcs3} = \alpha + \beta_1 \times SF_3 6D + \beta_2 \times (SF_3 6D - C_3)^2
\]

\[
= 5.02 + 59.68 \times SF_3 6D + 9.77 \times (SF_3 6D - 0.713)^2 = 9.98
\]

When \(SF_4 6D = 0\)

\[
\hat{Y}_{pcs4} = \alpha + \beta_1 \times SF_4 6D + \beta_2 \times (SF_4 6D - C_4)^2
\]

\[
= 4.67 + 59.01 \times SF_4 6D + 21.16 \times (SF_4 6D - 0.694)^2 = 14.86
\]

Inserting X=0 for each equation where X and \((X-C)^2\) occurs results in values of 7.0, 12.1, 10.0 and 14.9 (rounded to 1 decimal place) for Y. Later it will be discussed whether the use of the quadratic equations is justifiable in this case study.
**APPENDIX 6.4**

**Cubic Regression Equations**

when $SF_16D = 0$

$$\hat{Y}_{pcs} = \alpha + \beta_1 \times SF_16D + \beta_2 \times (SF_16D - C_2)^2 - \beta_3 \times (SF_16D - C_2)^3$$

$$= 8.48 + 57.09 \times SF_16D - 8.36 \times (SF_16D - 0.724)^2 - 36.00 \times (SF_16D - 0.724)^3 = 17.76$$

when $SF_26D = 0$

$$\hat{Y}_{pcs2} = \alpha + \beta_1 \times SF_26D + \beta_2 \times (SF_26D - C_2)^2 - \beta_3 \times (SF_26D - C_2)^3$$

$$= 5.72 + 59.75 \times SF_26D + 3.58 \times (SF_26D - 0.731)^2 - 47.72 \times (SF_26D - 0.731)^3 = 26.24$$

when $SF_36D = 0$

$$\hat{Y}_{pcs3} = \alpha + \beta_1 \times SF_36D + \beta_2 \times (SF_36D - C_3)^2 - \beta_3 \times (SF_36D - C_3)^3$$

$$= 4.14 + 60.98 \times SF_36D + 6.04 \times (SF_36D - 0.713)^2 - 36.87 \times (SF_36D - 0.713)^3 = 20.57$$

when $SF_46D = 0$

$$\hat{Y}_{pcs4} = \alpha + \beta_1 \times SF_46D + \beta_2 \times (SF_46D - C_4)^2 - \beta_3 \times (SF_46D - C_4)^3$$

$$= 2.81 + 61.78 \times SF_46D + 16.30 \times (SF_46D - 0.694)^2 - 74.02 \times (SF_46D - 0.694)^3 = 35.42$$

The cubic equations contain the terms $(X-C)^2$ and $(X-C)^3$ and the constants 'C' are adjusted at each survey depending on the data and the fitted equations. Note the C's are the same for $(X-C)^2$ and $(X-C)^3$ at each survey. The C values for the quadratic and cubic terms were calculated by the statistical software JMP 8.0.

Inserting $X=0$ for X, $(X-C)^2$ and $(X-C)^3$ results in values of 17.8, 26.2, 20.6 and 35.4 (rounded to 1 decimal place) for Y at each survey. Later it will be discussed whether or not the use of the cubic equations is justifiable in this case study.
APPENDIX 6.5

Regression Equations using Stevenson (1996) subscales

Linear
when SF1 6D = 0
\[ \hat{Y}_{\text{pcst}} = \alpha + \beta_1 \times \text{SF}_1 6D = 14.07 + 50.62 \times \text{SF}_1 6D \]
\[ = 14.07 \]

Quadratic
when SF1 6D = 0
\[ \hat{Y}_{\text{pcst}} = \alpha + \beta_1 \times \text{SF}_1 6D + \beta_2 \times (\text{SF}_1 6D - C_1)^2 = \]
\[ = 14.22 + 50.48 \times \text{SF}_1 6D - 3.54 \times (\text{SF}_1 6D - 0.724) \]
\[ = 12.37 \]

Cubic
when SF1 6D = 0
\[ \hat{Y}_{\text{pcst}} = \alpha + \beta_1 \times \text{SF}_1 6D + \beta_2 \times (\text{SF}_1 6D - C_1)^2 + \beta_3 \times (\text{SF}_1 6D - C_1)^3 \]
\[ = 9.44 + 57.43 \times \text{SF}_1 6D - 25.81 \times (\text{SF}_1 6D - 0.724)^2 - 201.17 \times (\text{SF}_1 6D - 0.724)^3 \]
\[ = 72.29 \]
APPENDIX 6.6

Figure A6.1 - From Trisolini et al (2002) – Fig 3.1 Regression Results for PCS and MCS scores on HALex Utility Scores

Figure A6.2 Regression results at Survey 1 for PCS on SF-6D using all 4 equations
APPENDIX 7.1

Table A7.1 ALSWH Case-Mix death models –
From each model a predicted value for death can be calculated based on the covariates.

<table>
<thead>
<tr>
<th>Model 1</th>
<th>coeff</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.071</td>
<td>-0.081, -0.060</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.028</td>
<td>-0.038, -0.019</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>age1</td>
<td>0.049</td>
<td>-0.021, 0.120</td>
<td>0.168</td>
</tr>
<tr>
<td>area</td>
<td>-0.299</td>
<td>-0.520, -0.078</td>
<td>0.008</td>
</tr>
<tr>
<td>educ</td>
<td>0.093</td>
<td>-0.199, 0.384</td>
<td>0.533</td>
</tr>
<tr>
<td>alone1</td>
<td>0.049</td>
<td>-0.170, 0.269</td>
<td>0.659</td>
</tr>
<tr>
<td>manage1</td>
<td>0.087</td>
<td>-0.143, 0.317</td>
<td>0.459</td>
</tr>
<tr>
<td>cmorb1</td>
<td>0.234</td>
<td>-0.033, 0.501</td>
<td>0.086</td>
</tr>
<tr>
<td>cancer1</td>
<td>0.779</td>
<td>0.492, 1.067</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>constant</td>
<td>-2.153</td>
<td>-7.326, 3.020</td>
<td>0.415</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 2</th>
<th>coeff</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.072</td>
<td>-0.083, -0.062</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.028</td>
<td>-0.037, -0.018</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>age1</td>
<td>0.055</td>
<td>-0.015, 0.124</td>
<td>0.124</td>
</tr>
<tr>
<td>area</td>
<td>-0.299</td>
<td>-0.518, -0.080</td>
<td>0.007</td>
</tr>
<tr>
<td>educ</td>
<td>0.128</td>
<td>-0.159, 0.414</td>
<td>0.384</td>
</tr>
<tr>
<td>alone1</td>
<td>0.028</td>
<td>-0.190, 0.245</td>
<td>0.804</td>
</tr>
<tr>
<td>manage1</td>
<td>0.074</td>
<td>-0.154, 0.303</td>
<td>0.523</td>
</tr>
<tr>
<td>cmorb1</td>
<td>0.221</td>
<td>-0.045, 0.487</td>
<td>0.103</td>
</tr>
<tr>
<td>constant</td>
<td>-2.366</td>
<td>-7.493, 2.761</td>
<td>0.366</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 3</th>
<th>coeff</th>
<th>95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.074</td>
<td>-0.084, -0.064</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.029</td>
<td>-0.038, -0.019</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>age1</td>
<td>0.055</td>
<td>-0.015, 0.124</td>
<td>0.123</td>
</tr>
<tr>
<td>area</td>
<td>-0.302</td>
<td>-0.520, -0.084</td>
<td>0.007</td>
</tr>
<tr>
<td>educ</td>
<td>0.117</td>
<td>-0.169, 0.404</td>
<td>0.422</td>
</tr>
<tr>
<td>alone1</td>
<td>0.013</td>
<td>-0.204, 0.230</td>
<td>0.904</td>
</tr>
<tr>
<td>manage1</td>
<td>0.067</td>
<td>-0.161, 0.295</td>
<td>0.566</td>
</tr>
<tr>
<td>constant</td>
<td>-2.064</td>
<td>-7.162, 3.035</td>
<td>0.428</td>
</tr>
<tr>
<td>Model 4</td>
<td>coeff</td>
<td>95%</td>
<td>CI</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>-----------</td>
<td>---------------</td>
</tr>
<tr>
<td>o1pcswha</td>
<td>-0.075</td>
<td>-0.085, -0.065</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.030</td>
<td>-0.039, -0.020</td>
<td>P&lt;0.001</td>
</tr>
<tr>
<td>age1</td>
<td>0.059</td>
<td>-0.010, 0.128</td>
<td>0.093</td>
</tr>
<tr>
<td>area</td>
<td>-0.294</td>
<td>-0.510, -0.078</td>
<td>0.008</td>
</tr>
<tr>
<td>educ</td>
<td>0.124</td>
<td>-0.160, 0.409</td>
<td>0.392</td>
</tr>
<tr>
<td>alone1</td>
<td>0.013</td>
<td>-0.202, 0.229</td>
<td>0.903</td>
</tr>
<tr>
<td>constant</td>
<td>-2.265</td>
<td>-7.317, 2.786</td>
<td>0.379</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 5</th>
<th>coeff</th>
<th>95%</th>
<th>CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.077</td>
<td>-0.087, -0.067</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.032</td>
<td>-0.041, -0.023</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>age1</td>
<td>0.063</td>
<td>-0.005, 0.130</td>
<td>0.068</td>
<td></td>
</tr>
<tr>
<td>area</td>
<td>-0.274</td>
<td>-0.484, -0.063</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td>educ</td>
<td>0.113</td>
<td>-0.168, 0.395</td>
<td>0.431</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-2.359</td>
<td>-7.292, 2.575</td>
<td>0.349</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 6</th>
<th>coeff</th>
<th>95%</th>
<th>CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.074</td>
<td>-0.083, -0.064</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.033</td>
<td>-0.042, -0.024</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>age1</td>
<td>0.060</td>
<td>-0.006, 0.126</td>
<td>0.075</td>
<td></td>
</tr>
<tr>
<td>area</td>
<td>-0.243</td>
<td>-0.446, -0.039</td>
<td>0.019</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-2.229</td>
<td>-7.055, 2.598</td>
<td>0.365</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 7</th>
<th>coeff</th>
<th>95%</th>
<th>CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.074</td>
<td>-0.083, -0.064</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.033</td>
<td>-0.041, -0.024</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>age1</td>
<td>0.061</td>
<td>-0.005, 0.126</td>
<td>0.071</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-2.388</td>
<td>-7.211, 2.435</td>
<td>0.332</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 8</th>
<th>coeff</th>
<th>95%</th>
<th>CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.074</td>
<td>-0.083, -0.064</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>o1mcswha</td>
<td>-0.032</td>
<td>-0.041, -0.024</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>2.012</td>
<td>1.406, 2.619</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 9</th>
<th>coeff</th>
<th>95%</th>
<th>CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>o1pcswha</td>
<td>-0.072</td>
<td>-0.082, -0.063</td>
<td>P&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>0.342</td>
<td>-0.076, 0.760</td>
<td>0.109</td>
<td></td>
</tr>
</tbody>
</table>
**APPENDIX 7.2**

Stata multiple imputation code:

```stata
ice diab_retro o1pcswha o1mcswha o2pcswha o2mcswha o3pcswha
    o3mcswha
    o4pcswha o4mcswha age1 age2 age3 age4 area educ smoke ///
cancer1 cancer2 cancer3 cancer4 manage1 manage2 manage3 ///
manage4 alone1 alone2 alone3 alone4 cmorb1 cmorb2 cmorb3
    cmorb4 ///
attrition12 attrition13 attrition14 IDalias, ///
saving(impAPASB) m(5) ///
    cmd( o1mcswha: regress, ///
        o1pcswha: regress, ///
        o2mcswha: regress , ///
        o2pcswha: regress, ///
        o3mcswha: regress, ///
        o3pcswha: regress, ///
        o4mcswha: regress, ///
        o4pcswha: regress, ///
        age1: regress, ///
        age2: regress, ///
        age3: regress, ///
        age4: regress, ///
        educ: logit, ///
        smoke: logit, ///
        area: logit: ///
        cancer1: logit, ///
cancer2: logit, ///
cancer3: logit, ///
cancer4: logit, ///
manage1: logit, ///
manage2: logit, ///
manage3: logit, ///
alone1: logit, ///
alone2: logit, ///
alone3: logit, ///
alone4: logit, ///
cmorb1: logit, ///
cmorb2: logit, ///
cmorb3: logit, ///
cmorb4: logit) ///
eq(o1pcswha: age1 area educ alone1 manage1 cmorb1, ///
    o1mcswha: age1 area educ alone1 manage1 cmorb1, ///
    o2pcswha: o1pcswha age2 area educ alone2 manage2, ///
    o2mcswha: o1mcswha age2 area educ alone2 manage2, ///
    o3pcswha: o2pcswha age3 area educ alone3 manage3, ///
    o3mcswha: o2mcswha age3 area educ alone3 manage3, ///
    o4pcswha: o3pcswha age4 area educ alone4 manage4, ///
```

---

Appendices
o4mcswha: o3mcswha age4 area educ alone4 manage4, ///
  age1: educ manage1 alone1 area, ///
  age2: educ manage2 alone2 area, ///
  age3: educ manage3 alone3 area, ///
  age4: educ manage4 alone4 area, ///
  educ: age1 manage1 alone1 area, ///
  smoke: age1 manage1 alone1 area, ///
cancer1: age1 manage1 alone1 area, ///
cancer2: age2 manage2 alone2 area, ///
cancer3: age3 manage3 alone3 area, ///
cancer4: age4 manage4 alone4 area, ///
manage1: age1 alone1 area smoke, ///
manage2: age2 manage1 alone2 area, ///
manage3: age3 manage1 alone3 area, ///
manage4: age4 manage3 alone4 area, ///
alone1: age1 manage1 area smoke, ///
alone2: age2 manage2 alone1 area, ///
alone3: age3 manage3 alone1 area, ///
alone4: age4 manage4 alone1 area, ///
cmorb1: age1 manage1 alone1 area, ///
cmorb2: age2 manage2 alone2 area, ///
cmorb3: age3 manage3 alone3 area, ///
cmorb4: age4 manage4 alone4 area, ///
area: age1 alone1 manage1 educ) ///

seed(1285964) replace
APPENDIX 8.1

i) **Publication from the method developed in Chapter 4**

Bowe SJ, Young AF, Sibbritt DW & Furuya H (2006) Transforming the SF-36 to account for death in longitudinal studies with three year follow-up Med Care 44: 956-959

ii) **Publication related to the statistical analysis applied in Chapters 6 and 7**