Williams, Claire (2018) Exploring multi-state modelling in an epidemiological and health economics context. PhD thesis.

http://theses.gla.ac.uk/8955/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

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

Enlighten:Theses
http://theses.gla.ac.uk/
theses@gla.ac.uk

Exploring multi-state modelling in an epidemiological and health economics context

Claire Williams

MSc in Medical Statistics, BSc (Hons) in Mathematics

Submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Health Economics and Health Technology Assessment

Institute of Health and Wellbeing

University of Glasgow

April 2018

Abstract

Competing risks analysis is more appropriate than standard survival analysis when there are two or more mutually exclusive possible events, and investigates which one of the events occurs first. Competing risks analysis constitutes the simplest form of multi-state modelling. Multi-state modelling more generally extends the competing risks approach to consider events of interest that can occur after the first event. However, competing risks and multi-state modelling have not been used to their full potential in health research. The aim of this thesis is to demonstrate the potential of multi-state modelling in an epidemiological and health economics context, in areas where it is not widely applied. Focus is on two case studies – one in epidemiology and one in health economics.

The first case study is in stroke epidemiology and investigates the outcomes stroke recurrence and death. The research is thought to be the first to comprehensively examine the competing risks stroke recurrence and death without recurrence. It demonstrates the clinical insights that can be gained by decomposing a composite outcome and by studying the cumulative incidence of each event alongside the hazards that drive them. Furthermore, an illustration of the flexibility in predictions of multi-state modelling is given. Predictions at the start of the study and as time progresses are demonstrated.

The second study is in health economics and is based on a technology appraisal submitted to the National Institute for Health and Care Excellence in the UK. An objective of this thesis is to compare multi-state modelling with the two common approaches of Markov decision-analytic modelling and partitioned survival. This comparison shows that the conventional decision-analytic modelling and multi-state modelling differ substantially when the assumptions vary between the approaches, but produce equivalent results when they make the same transition assumptions. Therefore, the greatest influence on the clinical and cost-effectiveness results is the choice of assumptions rather than the modelling approach used itself. The research highlights it is imperative to check that any assumptions made are realistic. The comparison of the approaches shows any output required from the conventional approaches can just as easily be produced using multi-state modelling. It is hoped this research will encourage further adoption of multi-state modelling, in the many areas where it has not yet reached its full potential.

Table of Contents

Contents

Abstract ii
Table of Contents iii
List of tables viii
List of figures x
Publications and Presentations/Invited Talks xii
Acknowledgement xiv
Author's Declaration xv
Abbreviations xvi
Chapter 1 Introduction 1
1.1 Background 1
1.2 Overall aim and objectives 3
1.3 Structure of the thesis 6
Chapter 2 Background to competing risks method 9
2.1 Introduction 9
2.2 Competing risks: definitions and common scenarios 10
2.3 Kaplan-Meier not appropriate for competing risks: introduction to competing risks cumulative incidence 14
2.3.1 Cumulative incidence for competing risks 18
2.4 Competing risks modelling approaches 20
2.4.1 Introduction 20
2.4.2 Latent failure times competing risk framework 20
2.4.3 Cause-specific hazards Cox approach to modelling competing risks 21
2.4.4 Fine and Gray's subdistribution proportional hazard modelling approach to competing risks 23
2.4.5 Other approaches to modelling competing risks 24
2.5 The effect of a covariate on the cause-specific hazard is not necessarily the same as its effect on the cumulative incidence 25
2.5.1 the covariate does not affect the cause-specific hazard of either event (scenario A1) 27
2.5.2 the covariate has an effect on one event only (scenarios B1-B4) .. 27
2.5.3 the covariate affects the cause-specific hazard of each event but in opposite directions (scenarios C1-C4) 29
2.5.4 the covariate affects the cause-specific hazard of each event in the same direction (scenarios D1-D8) 30

2.6 Derivation of the cumulative incidence using cause-specific hazard modelling 32
2.7 Variable selection strategies with competing risks 34
2.7.1 Variable selection and Fine and Gray's proportional subdistribution hazard approach 35
2.7.2 Methodological issues relating to variable selection with the cause-specific hazards approach 36
2.8 Two schools of thought on the use of two common approaches to competing risks 36
2.8.1 the research question should drive the modelling approach used .. 37
2.8.2 both the cause-specific hazard and subdistribution hazard/ cumulative incidence should be modelled for each competing risk 39
2.9 Composite outcomes vs decomposition into competing risks 40
2.10 Need for better recognition of competing risks in the clinical community 41
2.10.1 Review of recognition of the need for competing risks analyses, where appropriate 42
2.10.2 Review of competing risks analyses of stroke recurrence and death in the literature 44
2.10.3 Review of the possibility of decomposing recurrence-free survival after stroke in the literature that uses that outcome 47
2.10.4 Summary of the review 52
Chapter 3 Competing risk analysis with the Edinburgh Stroke Study 54
3.1 Overview of the Edinburgh Stroke Study 55
3.2 Analyses of the composite outcome versus decomposition of the outcome: cause-specific hazards approach to competing risks 57
3.3 Cause-specific hazards vs Fine and Gray subdistribution modelling approaches 63
3.3.1 covariates that affect the hazard of each event but in opposite directions 65
3.3.2 covariates that only affected the hazard of one event 68
3.3.3 covariates that affected the hazards of each event in the same direction 71
3.4 Consequences of not taking into account competing risks 74
3.4.1 Non-parametric naïve Kaplan-Meier versus competing risks cumulative incidence 74
3.4.2 Naïve cause-specific hazard-based cumulative incidence versus that derived appropriately using the cause-specific hazards of each of the competing risks 75
3.5 Discussion/summary of results 77
Chapter 4 Background to multi-state modelling 83
4.1 Introduction 83
4.2 Preliminary concepts 85

4.3 Markov multi-state models 89
4.3.1 Non-parametric Markov multi-state models 89
4.3.2 Semi-parametric Markov multi-state models 90
4.3.3 Parametric Markov multi-state models 92
4.4 Semi-Markov multi-state models 94
4.5 Flexibility of predictions at different times over the course of the disease 95
4.6 Broad review of multi-state models in the medical literature 97
4.7 Review of illness-death modelling of stroke, recurrence and death in the literature 99
Chapter 5 Multi-state modelling with the Edinburgh Stroke Study data 103
5.1 Description of the multi-state model and results for covariate effects 103
5.2 Results of multi-state modelling: predictions 111
5.2.1 Predictions from the start of the study 112
5.2.2 Dynamic predictions 114
5.3 Comparison with competing risks analysis 117
5.4 Discussion 118
Chapter 6 Health economic modelling: 121
6.1 Partitioned survival 122
6.2 Markov decision-analytic modelling 125
6.2.1 Define the relevant health states 125
6.2.2 (Relaxation of) the Markov property 127
6.2.3 Assignment of transition probabilities/obtaining results 128
6.3 Review of multi-state models for health economic modelling in the literature 129
6.3.1 Methodological issues: continuous-time multi-state modelling contrasted with Markov decision-analytic models 132
6.3.2 Approaches to calculating transition probabilities in continuous-time multi-state modelling 133
6.3.3 Some applied uses of multi-state modelling in the health economics literature 134
6.3.4 Multi-state modelling applied to infections 136
6.3.5 Multi-state modelling applied to HIV 138
6.3.6 Multi-state modelling contrasted with partitioned survival 138
6.3.7 Misunderstandings in the literature 139
6.3.8 Alternative approaches 140
6.3.9 Summary of the review 141
6.4 Assessment of fit over the observed period of a study 143
6.5 Extrapolation of outcomes for use in economic evaluations 145
6.5.1 Motivation for extrapolation 145

6.5.2 Conventional approach: parametric regression 146
6.5.3 Alternative approaches to extrapolation 149
Chapter 7 Comparing Markov decision-analytic modelling, partitioned survival and multi-state modelling.....150
7.1 Introduction 150
7.2 Description of the decision problem and the existing Markov decision-analytic model 151
7.2.1 The decision problem 152
7.2.2 Structure of the model.....152
7.2.3 Sensitivity analysis initiated by the Evidence Review Group 153
7.3 Description of Rituximab clinical trial dataset.....156
7.4 Partitioned survival 157
7.4.1 Progression-free survival 157
7.4.2 Overall survival 169
7.4.3 Compatibility of fits for progression-free and overall survival 182
7.5 Multi-state modelling 184
7.5.1 Introduction 184
7.5.2 Consideration of the appropriateness of the proportional hazards assumption 185
7.5.3 Deciding on the appropriateness of the Markov property 188
7.5.4 Choice of distribution for each transition.....188
7.5.5 Summary of base-case model for each transition.....198
7.6 Comparison of the three approaches 199
7.6.1 Visual assessment of the fits.....199
7.6.2 Mean Life Years/QALYs 204
7.6.3 Description of the costs 206
7.6.4 Incremental cost-effectiveness ratios with each approach 208
7.7 Sensitivity analyses for the multi-state modelling approach.....210
7.7.1 Using alternative distributions for each of the transitions in the multi-state modelling 210
7.7.2 One-way sensitivity analyses in multi-state modelling.....216
7.7.3 Probabilistic sensitivity analysis for multi-state modelling 217
7.8 Direct comparison of the manufacturer's Markov decision-analytic model with multi-state modelling 220
7.9 Summary of the results and Discussion.....225
Chapter 8 Main insights, limitations and potential areas of future research..231
8.1 Introduction 231
8.2 Main insights.....232
8.2.1 Insights gained from the competing risks analysis over and above standard survival analysis.....232

8.2.2 Insights gained from the multi-state modelling of stroke, recurrence and death 236
8.2.3 Insights gained from exploring multi-state modelling in a health economics context 237
8.3 Limitations....239
8.3.1 Limited follow-up with covariates versus extended follow-up with limited covariates....239
8.3.2 Variable selection with competing risks and multi-state modelling..240
8.4 Areas of future research....244
8.4.1 Enhancing health economic multi-state modelling by combining trial data with longer-term routine administrative data....244
8.4.2 Consideration of multiple comparators 252
8.4.3 Simulation exercise to dissect the differences in the approaches compared....252
8.4.4 Validation of models....253
8.4.5 Predictions and modifiable lifestyle factors 253
8.4.6 Widening the functionality of the coding for cost-effectiveness analysis in R using multi-state modelling....254
Appendices 256
Appendix I Literature search of reviews of extent of use of competing risks ...256
Appendix II Koller et al. (2012)'s search strategy 257
Appendix III Literature search of competing risks analyses of stroke recurrence and death....258
Appendix IV Evidence from the literature of competing risks analyses of stroke recurrence and death....259
Appendix V PRISMA diagram for literature search for use of recurrence-free survival as an outcome in stroke 267
Appendix VI Evidence of contributions to the literature using recurrence-free survival after stroke....268
Appendix VII Literature search of broad use of multi-state modelling in medical literature 277
Appendix VIII PRISMA diagram for literature search of use of multi-state modelling in health economics literature....279
Appendix IX....280
References....287

List of tables

Table 2.1 Effect on the cause-specific hazard versus that on the cumulative incidence: different scenarios 26
Table 3.1 Patient characteristics involved in analyses of Edinburgh Stroke Study 56
Table 3.2 Univariable Cox composite vs cause-specific hazard competing risks Cox modelling results 58
Table 3.3 Multivariable Cox composite vs cause-specific hazard Cox modelling results 61
Table 3.4 cause-specific hazards vs Fine and Gray's subdistribution hazards multivariable modelling 64
Table 5.1 Univariable Cox Regression for each of the transitions in the ESS multi-state model 106
Table 5.2 Unparsimonious multivariable Cox regression for each of the transitions in the ESS multi-state model 108
Table 5.3 Parsimonious multivariable Cox Regression for each of the transitions in the ESS multi-state model 109
Table 5.4 Parsimonious multivariable Cox Regression for each of the transitions in the ESS multi-state model: in initial stroke state 6 months on from stroke .. 110
Table 5.5 Parsimonious multivariable Cox Regression for each of the transitions in the ESS multi-state model: in initial stroke state 1 year on from stroke 111
Table 5.6 Risk profiles used for prediction 112
Table 6.1 Common parametric distributions 148
Table 7.1 Mean Life Years/QALYs for time horizon of 15 years: Sensitivity analyses initiated by the ERG to remove the benefit in overall survival [adapted from PenTAG (2009)] 154
Table 7.2 Cost per Life Year gained & Cost per QALY gained: manufacturers submission and ERG sensitivity analyses [adapted from PenTAG (2009)] 155
Table 7.3 ERG sensitivity analyses expressed as hazard ratios 155
Table 7.4 Summary of number of events in the CLL-8 trial 157
Table 7.5 AICs from parametric regressions models for progression-free survival 162
Table 7.6 Weibull model for progression-free survival 164
Table 7.7 Progression-free survival (years) AUCs from model with treatment as a covariate vs separate fit for RFC 166
Table 7.8 Progression-free survival (years) AUCs from model with treatment as a covariate vs separate fit for FC 166
Table 7.9 Extrapolated progression-free survival AUCs from model with treatment as a covariate vs separate fit for RFC 169
Table 7.10 Extrapolated progression-free survival AUCs from model with treatment as a covariate vs separate fit for FC 169
Table 7.11 AICs from parametric regressions models for overall survival 174
Table 7.12 Extrapolation of overall survival for FC using Weibull models - Summary using a selection of starting points 176
Table 7.13 Extrapolation of overall survival for RFC using Weibull models - Summary using a selection of starting points 177
Table 7.14 Base case models for overall survival 179
Table 7.15 Overall survival AUCs from model with treatment as a covariate vs separate fit for RFC 182

Table 7.16 Overall survival AUCs from model with treatment as a covariate vs separate fit for FC 182
Table 7.17 Effect of reducing the treatment effect on extrapolated overall survival to 15 years for RFC. 184
Table 7.18 Results of a Cox Markov state-arrival extended model for progression → death 188
Table 7.19 AICs from parametric regression models for progression -> death . 191
Table 7.20 Base case models (all Gompertz) for the three transitions using the multi-state modelling approach 199
Table 7.21 Mean Life Years and QALYs results: all three approaches 205
Table 7.22 Costs breakdown: all three approaches 207
Table 7.23 Incremental cost effectiveness ratios: all three approaches 208
Table 7.24 Mean Life Years and QALYs: trial observation period of 0-4 years . 209
Table 7.25 Mean Life Years and QALYs: extrapolation over 4-15 years 209
Table 7.26 Beta Pert distributions used in probabilistic sensitivity analysis for cost parameters 218
Table 7.27 Incremental mean Life Years in each health state: equivalent assumptions 224

List of figures

Figure 2.1 Common competing risks scenarios 12
Figure 2.2 original and updated search of competing risks literature based on the criteria of Koller et al. (2012) 43
Figure 3.1 Cumulative hazard vs cumulative incidence: lift both arms and haemorrhage on scan 67
Figure 3.2 Cumulative hazard vs cumulative incidence: independence before stroke and age 69
Figure 3.3 Cumulative hazard vs cumulative incidence: sex and stenosis 70
Figure 3.4 Cumulative hazard vs cumulative incidence: atrial fibrillation and other occlusive vascular disease 72
Figure 3.5 Cumulative hazard vs cumulative incidence: orientated speech..... 73
Figure 3.6 non-parametric naïve KM vs competing risks cumulative incidence .. 75
Figure 3.7 Cumulative incidences derived using the naïve cause-specific hazard approach versus those using the more appropriate cause-specific hazard approach. Abbreviations: CSH=cause-specific hazard, CIF=cumulative incidence function 76
Figure 4.1 State transition diagram for an illness-death model 85
Figure 4.2 Broad search of multi-state modelling in the medical literature..... 98
Figure 5.1 State-transition diagram for the illness-death model 104
Figure 5.2 Predictions, at the time of the index stroke, of being in each health state over time 113
Figure 5.3 Dynamic predictions of each event 115
Figure 5.4 comparison of predictions of recurrence: multi-state modelling vs competing risks 117
Figure 6.1 Illustration of quality-adjusted survival.....123
Figure 6.2 Illustration of partitioned survival 124
Figure 6.3 Markov decision-analytic model transition diagram 126
Figure 6.4 Illustration of tunnel states in Markov decision-analytic modelling .128
Figure 7.1 Kaplan-Meier estimates of progression-free survival for RFC and FC 158
Figure 7.2 log-log plot of progression-free survival 159
Figure 7.3 Cumulative hazard plot of progression-free survival 159
Figure 7.4 Kaplan-Meier and prediction estimates of progression-free survival: 160
Figure 7.5 Cox-Snell residuals following the parametric regression of progression-free survival 161
Figure 7.6 Extrapolated progression-free survival to 15 years: (a) RFC (b) FC 163
Figure 7.7 Predicted progression-free survival: treatment as a covariate vs separate fit for RFC. 165
Figure 7.8 Predicted progression-free survival: treatment as a covariate vs separate fit for FC.....165
Figure 7.9 Extrapolated progression-free survival: treatment as a covariate vs separate fit for RFC. 167
Figure 7.10 Extrapolated progression-free survival: treatment as a covariate vs separate fit for FC.....168
Figure 7.11 Kaplan-Meier estimates of overall survival for RFC and FC 170
Figure 7.12 Log-log plot of all-cause mortality (1-overall survival).....171
Figure 7.13 Cumulative hazard plot of all-cause mortality (1-overall survival) 171

Figure 7.14 Kaplan-Meier and prediction estimates of overall survival: (a) RFC (b) FC....172
Figure 7.15 Cox-Snell residuals following the parametric regression of overall survival....173
Figure 7.16 Extrapolated overall survival to 15 years: (a) RFC (b) FC 175
Figure 7.17 Extrapolation of overall survival for FC using Weibull models....176
Figure 7.18 Extrapolation of overall survival for RFC using Weibull models ...178
Figure 7.19 Overall survival to 15 years for RFC and FC 179
Figure 7.20 Predicted overall survival: treatment as a covariate vs separate fit for RFC. 180
Figure 7.21 Predicted overall survival: treatment as a covariate vs separate fit for FC....181
Figure 7.22 Base case partitioned survival for (a) FC and (b) RFC 183
Figure 7.23 Log-log plots for the transitions in the multi-state model 186
Figure 7.24 Cumulative hazard versus time plots for the transitions in the multi-state model....187
Figure 7.25 Progression → death for over trial observation period: (a) RFC and (b) FC....190
Figure 7.26 Progression → death : Cox-Snell residuals....191
Figure 7.27 Progression→ death extrapolated to 15 years: (a) RFC and (b) FC 192
Figure 7.28 Probability of progression over trial observation period for (a) RFC and (b) FC Different distributions for progression-free →progression are distinguished by colour Shades of each colour represent a different distribution used for progression-free → death....194
Figure 7.29 Probability of progression extrapolated to 15 years: (a) RFC and (b) FC Different distributions for progression-free →progression are distinguished by colour Shades of each colour represent a different distribution used for progression-free → death....195
Figure 7.30 Progression-free → death without progression over the trial observation period: 197
Figure 7.31 Progression-free → death without progression extrapolated to 15 years....198
Figure 7.32 Probability of being in the progression-free state using each approach....200
Figure 7.33 Probability of being in the progression state using each approach (a) RFC (b) FC 201
Figure 7.34 Probability of being in the death state using each approach (a) RFC (b) FC 202
Figure 7.35 death without progression 203
Figure 7.36 post-progression death 204
Figure 7.37 Probability of progression for RFC: extrapolation to 15 years....212
Figure 7.38 Probability of progression for FC: extrapolation to 15 years 213
Figure 7.39 Effect of fit to progression-free → death on probability of being in the: 215
Figure 7.40 Tornado plot: one-way sensitivity analyses 217
Figure 7.41 Cost-effectiveness plane 219
Figure 7.42 Cost-effectiveness acceptability curve....219
Figure 7.43 Probability of staying in the progression-free state: equivalent assumptions 221
Figure 7.44 Probability of being in the progression state: equivalent assumptions 222
Figure 7.45 Probability of being in the death state: equivalent assumptions .223

Publications and Presentations/Invited Talks

The following publications were developed as part of this thesis:

Williams, C., Lewsey, J. D., Briggs, A. H. , Mackay, D. F. (2017) Cost-effectiveness analysis in R using a multi-state modeling survival analysis framework: a tutorial. *Medical Decision Making,* 37:340-352. doi:10.1177/0272989X16651869

Williams, C., Lewsey, J. D., Mackay, D. F., Briggs, A. H. (2017) Estimation of survival probabilities for use in cost-effectiveness analyses: a comparison of a multi-state modeling survival analysis approach with partitioned survival and Markov decision-analytic modeling. *Medical Decision Making,* 37: 427-439. doi:10.1177/0272989X16670617

The following articles were also published during the period of this research:

Waduud MA, Choong WL, Ritchie M, **Williams C**, Yadavali R, Lim S, Buchanan F, Bhat R, Ramanathan K, Ingram S, Cormack L, Moss JG (2015) Endovascular Aneurysm Repair: Is Imaging Surveillance Robust, and Does It Influence Long-term Mortality? *CardioVascular and Interventional Radiology*; 38:33-39 DOI 10.1007/s00270-014-0890-5

Docking RI, Mackay A, **Williams C**, Lewsey J, Kinsella J, Booth MG (2014). Comorbidity and intensive care outcome – a multivariable analysis *Journal of the Intensive Care Society*; 15(3): 205-212 DOI: 10.1177/175114371401500306

Quasim T, **Williams C**, Riddell L, Rankin AC, Kinsella J (2013). A retrospective study on the effects of illness severity and atrial fibrillation on outcomes in the intensive care unit. *International Journal of Intensive Care*; 20(4): 119-122.

The following oral presentations were given as part of this research:

Conference Presentation 'Illustrating the usefulness of the multi-state modelling approach for health economic modelling', 34th Annual Conference of the

International Society for Clinical Biostatistics, 25 -29th August 2013, Munich, Germany

Invited Talk at Workshop 'Multi-state survival analysis modelling', Methods for extrapolation from clinical trials data to inform economic evaluation MRC Trials Hub workshop, 22 January 2013, University of Oxford, UK

Invited Talk at Seminar 'An illustration of the usefulness of the multi-state model survival analysis approach for health economic evaluation modelling', Centre for Health Economics, May 2014, University of York, UK

Lecture and Workshop presented on partitioned survival and multi-state modelling, Economic Evaluations in Clinical Trials CPD Course, 31st October 2013, University of Glasgow, UK

Acknowledgement

I am extremely grateful to my main supervisor Dr Jim Lewsey for his invaluable support, guidance and encouragement throughout this research. I would also like to express my thanks to my other supervisors Professor Olivia Wu, Professor Andrew Briggs and Dr Daniel Mackay for their helpful comments on drafts of this thesis or accompanying publications and for their advice and encouragement. My thanks also go to my colleagues in the wider HEHTA team for providing a supportive working environment.

I would also like to thank the Medical Research Council for funding this research.

I also greatly appreciate the clinical advice received from Professor Martin Dennis and Dr Will Whiteley of the University of Edinburgh on the stroke epidemiology aspects of this thesis.

Finally, I would like to thank my family and friends for their patience throughout this process.

Author's Declaration

With the exception of the parts of chapter 7 detailed below, all of this thesis is my own work and has not been submitted for the award of a different degree at the University of Glasgow, or any other institution.

With regards to chapter 7, the comparison of approaches was based on a decision-analytic model produced by Roche in a manufacturer's submission to the National Institute for Health and Care Excellence for the specific technology appraisal TA174. Furthermore, where indicated, some of the sensitivity analysis was carried out by the Evidence Review Group. All remaining analysis, that using the partitioned survival approach and multi-state modelling, was carried out by myself.

An earlier version of chapter 7 has been published (Williams et al., 2017a, Williams et al., 2017b). All drafting and analysis was undertaken by myself, and contributions by collaborating authors were in the form of comments on drafts only. The analysis and text were revised and expanded upon for this thesis, all of which was my own work.

Abbreviations

95% CI	95% confidence interval
ABCCD	score used to predict early risk of stroke after a transient ischaemic attack
AFT	accelerated failure time
AIC	Akaike information criterion
AUC	area under the curve
BIC	Bayesian information criterion
CD4	cluster of differentiation 4
CHD	coronary heart disease
CIF	cumulative incidence function
CONSORT	Consolidated Standards of Reporting Trials
CSH	cause-specific hazard
EBMT	European Group for Blood and Marrow Transplantation
ECG	electrocardiogram
ESS	Edinburgh Stroke Study
HIV	human immunodeficiency virus
HR	hazard ratio
ICA	internal carotid artery
ICER	incremental cost-effectiveness ratio
IPCW	inverse probability of censoring weighting
KM	Kaplan-Meier
lasso	least absolute shrinkage and selection operator
MAC	myeloablative conditioning
MACE	major adverse cardiovascular events
MI	myocardial infarction
mRSA	meticillin-resistant Staphylococcus aureus
NH	new regime
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NIHR	National Institute for Health Research

NRM	non-relapse mortality
OS	overall survival
PenTAG	Peninsula Technology Assessment Group
PFS	progression-free survival
PH	proportion hazards
PSM	propensity score matching
QALY	quality-adjusted life year
Q-Q plot	quantile-quantile plot
Q-TWiST	Quality-adjusted Time Without Symptoms of disease and Toxicity of treatment
RCT	randomised controlled trial
RIC	reduced intensity conditioning
RNA	ribonucleic acid
RR	relative risk
s.e.	standard error
SAS	Statistical Analysis System software
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TA	technology appraisal
TI	treatment interruption
TIA	transient ischaemic attack
TRM	treatment-related mortality

Chapter 1 Introduction

1.1 Background

Survival analysis is a well-established technique in the fields of epidemiology/medical statistics and health economics. It is widely applied in the form of standard survival analysis which involves only one event of interest. This typically involves investigating all-cause mortality as an outcome or a composite outcome comprising of a non-fatal event(s) and death. These outcomes are often studied as part of a randomised clinical trial in both medical statistics and health economics. In the former, the efficacy/effectiveness of an intervention is of interest while in the latter the cost-effectiveness of an intervention is of overall interest. However, a major component in the analysis of cost-effectiveness is effectiveness and, in health economic modelling that involves the outcomes described above, survival analysis is often used to investigate the (quality-adjusted) life years gained from an intervention.

Frequently however, standard survival analysis may be too restrictive to answer the research question of interest because more specific events are of interest. For instance, interest may be in specific causes of death rather than all-cause mortality, or in a non-fatal outcome specifically rather than a composite one. These scenarios are termed “competing risks” and require a different approach to analysis than standard survival analysis. Competing risks analysis focuses on identifying which of two or more mutually exclusive events occur first, and in that sense the risks are competing with one another to be the first to occur and hence the event. The competing risk that occurs precludes the other competing risk(s) from being the event to occur first. For example, a specific cause of death would be considered a competing risk for another cause of death. An alternative definition of a competing risk is an event that precludes an event of interest, or otherwise affects the risk of it occurring. Consequently, competing risks scenarios need not just involve death. Non-fatal outcomes can also act as competing risks. For instance, hospital discharge should be treated as a competing risk when interest is in in-hospital infection. In another common scenario that arises, death should be treated as a competing risk when analysing

a non-fatal outcome. However, Koller et al. (2012) performed a literature review that found that competing risks, and the issues that require consideration when faced with such a scenario, were not widely recognised in the clinical literature. This was particularly the case in high impact clinical journals.

Standard and competing risks survival analysis constitute the simplest forms of modelling under the more general multi-state modelling survival analysis framework. Multi-state modelling more generally extends the competing risks approach to also consider events of interest that can happen after the first event. It is therefore applicable when at least one of the events is non-fatal. In a multi-state modelling situation, a first event(s) that is studied as part of a competing risks analysis is considered to be an intermediate state(s) between the initial and final states, all of which are of interest. This is the case particularly when studying the course of a chronic and/or progressive disease. For instance, a competing risks analysis might involve following patients in an initial healthy state to see which one of the two events/states of interest non-fatal illness or death they experience first. In a multi-state model, the non-fatal illness state would act as an intermediate state between the initial healthy state and final death state.

One of the main advantages of multi-state modelling over standard and competing risks survival analysis is the flexibility in predictions it can accommodate. Predictions of being in different health states over time can be estimated as patients enter the initial health state. Furthermore, dynamic predictions can be carried out, that is to say predictions as time progresses, especially as patients progress to other health states. These latter predictions are particularly applicable in an epidemiological/medical statistics context, but do not appear to be widely applied. In a health economics context, multi-state modelling falls under the umbrella of state-transition modelling, a common approach in health economic modelling. However, as this thesis will show, multi-state modelling is under-used in the field of health economics. Multi-state modelling for cost-effectiveness analysis has great potential as an alternative to conventional spreadsheet-based approaches. It is syntax-based providing a transparent record of the analysis and it makes errors easier to spot.

Furthermore, it explicitly allows testing of the Markov property by simply adding a relevant covariate and semi-Markov modelling does not require tunnel states. In addition, the building of Markov multi-state models is less cumbersome and time-consuming; such models can be created in seconds.

1.2 Overall aim and objectives

The overall aim of this thesis is to demonstrate the potential of multi-state modelling in an epidemiological and health economics context, in areas where it is not widely applied. Two case studies are used to provide focus – one in epidemiology and one in health economics. The first case study is in stroke epidemiology with the outcomes recurrence and death.

Objectives in this stroke epidemiology context are:

- to investigate the up-to-date use in the literature of competing risks and multi-state modelling for analysing the outcomes recurrence and death

- to demonstrate the added insight gained from using competing risks analysis and multi-state modelling over and above standard survival analysis.

The second case study, which is in health economics, is based on a technology appraisal (TA) submitted to the National Institute for Health and Care Excellence (NICE) in the UK. This particular TA involved evaluating the cost-effectiveness of rituximab in combination with fludarabine and cyclophosphamide, compared to fludarabine and cyclophosphamide alone, for the first-line treatment of chronic lymphocytic leukaemia. The Evidence Review Group working on behalf of NICE had some concerns about the economic model submitted by the manufacturer and instigated some sensitivity analyses of their own. However, there was scope to show how multi-state modelling could provide an alternative to the common approaches adopted by modellers performing economic evaluations.

Objectives in the health economics context are:

- to explore the use of multi-state modelling in the health economics literature

- to compare multi-state modelling with two common approaches applied in health economic modelling, namely Markov decision-analytic modelling and partitioned survival.

In recent methodology guidance issued by the NICE Decision Support Unit [Woods et al. (2017)], the two publications that accompany this thesis have been cited to help raise awareness that multi-state modelling can provide an alternative to, or complement, cost-effectiveness analyses conducted using partitioned survival and more conventional Markov decision-analytic state-transition modelling. The guidance also outlines some of the advantages, compared to partitioned survival, of the conventional discrete-time approach to decision-analytic state-transition modelling and that of continuous-time multi-state modelling. In particular, it highlights some benefits pertaining to conceptualising the model, for a decision problem involving patients who experience distinct health states and possible transitions between them, as a discrete or continuous-time state transition model. These include, for intermediate states, incorporating sensitivity analyses of the effect of treatment and extrapolations that can be based on external data focused specifically on those health states, allowing assessment of clinical plausibility.

The “cognitive dissonance” of using the partitioned survival approach when health states and transitions are involved also means there is not a formal structural link between morality and earlier intermediate events. For example, in an illness-death model of progression-free, progression and death, mortality depends on all three individual transitions, with the rate of death reflecting the evolving proportion of patients in the progressed state and the differences in mortality between progression-free and progressed patients [Woods et al. (2017)].

Additional contributions of continuous-time multi-state modelling to the methodology and application of cost-effectiveness analysis emphasised in this thesis include:

- The ability to test the Markov property in a straightforward manner simply by including in the model a covariate that represents time in a previous state, or a function thereof. If evidence is found that the Markov property does not hold, or it is not thought to hold, then analysis can progress using a semi-Markov approach. This explicit assessment of whether the Markov property holds can often be overlooked in more conventional state-transition modelling. That possible violations of the property can be investigated with such ease, adds another tool for modellers to help them decide whether a Markov or semi-Markov is more appropriate therefore improving face validity of models.

- The implementation of semi-Markov modelling, under the continuous-time approach demonstrated in this thesis, avoids the need for some of the aspects that make the approach more complex under the more conventional discrete-time framework. For example, there is no need for tunnel states, matrix algebra involving discrete cycles or complicated microsimulation. In addition, if a modeller wants transitions to be based on trial data, unlike a Markov approach, the semi-Markov approach can be implemented without the individual patient level data. It only requires access to Kaplan-Meier curves that can be digitized if necessary.

- The analysis only involves one software package and as such is streamlined while the use of coding provides transparency. All the calculations performed on the data can be viewed together in a file that provides a traceable, annotated written record of the analysis. This can make it easier to keep track of what aspects of the code have been updated, and avoid the accidental changes that can occur if a cell is inadvertently amended when using spreadsheet-based approaches. The model building process is quick; all aspects of implementation such as estimation and prediction can be run in seconds.

The transparency of the annotated coding also better facilitates model sharing in an open-source environment. This can help with reproducibility, validation and standardisation of models and ensure that those working in a similar area do not need to start from scratch. It can also satisfy the growing culture amongst journals of requiring syntax with submissions and the increasing popularity of sharing code on websites such as Github. This thesis will demonstrate that multi-state modelling, an approach better known in other disciplines such as medical statistics, can easily be adapted for use in health economic modelling. It will illustrate that anything that can be built using the more conventional spreadsheet-based approach can also be similarly transparent, and it has provided a foundation to extend this to other health economic models in the future.

1.3 Structure of the thesis

Chapter 2 emphasises that the presence of competing risks requires a different approach to analysis than standard survival analysis, and proceeds to describe approaches to competing risks survival analysis. Intertwined with this, it also highlights key issues that should be considered when faced with a competing risks scenario. In addition, a review of the extent to which competing risks are recognised is given, both in the broad clinical literature and for stroke recurrence and death without recurrence specifically. The latter identifies gaps in the research that motivates the empirical analysis of my stroke epidemiology case study in the next chapter.

Chapter 3 presents competing risks analyses of the outcomes recurrence and death without recurrence in stroke patients. In particular, it highlights the extra insight that can be gained from decomposing a composite outcome into competing risks. Furthermore, both the cause-specific hazard and Fine and Gray (1999)'s proportional subdistribution hazard approaches to modelling are demonstrated. The greater understanding that can be gained when both the hazards and cumulative incidences/subdistribution hazards are investigated for each of the competing risks is then outlined. In addition, the bias that can be introduced when competing risks are not taken into account is highlighted.

In Chapter 4 the multi-state modelling survival analysis framework is described. In addition, a review is undertaken of multi-state modelling of stroke, recurrence and death which provides the scope for the empirical analysis in the following chapter.

Chapter 5 presents illness-death modelling of stroke, recurrence and death; a particular type of multi-state model. The results of the effects of covariates on all relevant transitions between health states are shown and interpreted. Furthermore, illustrations are given of predictions at the time of the index stroke, and dynamically taking into account the time elapsed and any subsequent events since the stroke. In addition, the extra insight gained from this multi-state modelling approach over and above the competing risks analysis in Chapter 3 is highlighted.

Chapter 6 summarises partitioned survival and Markov decision-analytic modelling, two common approaches in health economic modelling. This overview of the methods provides fundamental background prior to an analysis comparing them with multi-state modelling in the next chapter. Furthermore, a review is given of the contributions in the literature involving multi-state modelling in a health economics context. Important methodological issues and barriers to adoption of the approach are highlighted. Chapter 6 then discusses the assessment of fit of models over the observed period of a study. Finally, the important concept in health economic modelling of extrapolation of outcomes beyond the observed data in order to take a lifetime perspective is described. This will be taken into account in the comparison presented in the next chapter.

In Chapter 7 a case study in a health economics context is used to illustrate and compare the Markov decision-analytic modelling, partitioned survival and multi-state modelling approaches. Furthermore for the multi-state modelling, the primary focus of this thesis, one-way and probabilistic sensitivity analyses will be presented to explore any uncertainty in the conclusions with regards to cost-effectiveness.

Finally, Chapter 8 summarises the main insights revealed by the analyses in the empirical chapters. In addition, limitations of this thesis are discussed and possible scope for developing this research further is outlined.

Chapter 2 Background to competing risks method

2.1 Introduction

This chapter provides an overview of the competing risks methodology under the survival analysis framework, giving the background to the empirical analysis in the next chapter. As well as describing the technique, this chapter highlights key issues that need to be considered when faced with competing risks. Section 2.2 emphasises that standard survival analysis is only applicable with one event of interest. It then proceeds to point out that if patients are at risk of two or more events, that affect the risk of the other(s) occurring first, then an approach that takes into account such competing risks is required. Since the need for a competing risks approach is not always recognised, section 2.2 continues by stressing the definition of a competing risks problem and outlines common scenarios when an approach that takes into account competing risks is required. Section 2.3 then explains why the Kaplan-Meier approach in standard survival analysis is not appropriate when competing risks are present. In sub-section 2.3.1 the more appropriate competing risks cumulative incidence approach is introduced. Section 2.4 concentrates on describing modelling approaches with competing risks data. Its main focus is on the two most widely-applied methods of cause-specific hazards modelling and Fine and Gray's proportional subdistribution hazards model. Then, section 2.5 discusses the issue that the effect of a covariate on the cause-specific hazard is not necessarily the same as its effect on the cumulative incidence. Section 2.6 is devoted to highlighting that the cumulative incidence can be derived from cause-specific hazard modelling. Next, section 2.7 considers issues related to variable selection with competing risks approaches. In section 2.8, two schools of thought with regard to the use of the two common approaches to modelling competing risks are outlined. The first is that the research question should drive the modelling approach used. The second strategy is that both the cause-specific hazard and cumulative incidence should be considered for each competing risk, to gain a full understanding of the competing risks scenario under consideration. Next, section 2.9 emphasises the extra insight that could be gained by decomposing a composite outcome into each of its competing risks component parts.

Contributions from the literature are used in the aforementioned sections of this chapter to help describe the competing risks method and highlight issues that arise over and above standard survival analysis. Because these sections are discussing the key methodological issues that repeatedly appear in the literature, a robust search strategy was not developed to review all such contributions to the literature. However, for the final section of this chapter section 2.10, a search strategy was developed in order to review the extent of use of competing risks in the literature. This had a particular focus on stroke epidemiology to identify gaps in the existing research to motivate the empirical analysis of the stroke case study in the next chapter. Furthermore, throughout this current chapter – i.e. the background to the method, the methodological issues raised and the review of the extent of use in the literature – misconceptions and misunderstandings relating to some of the methodological issues are highlighted. The purpose of this is to alleviate the confusion that can arise from some of the conflicting messages in the literature.

This chapter is concerned with competing risks in the standard case when all event times are observed exactly or are right-censored. Readers interested in competing risks involving interval-censored and truncated data may find the contributions by Hudgens et al. (2001), Hudgens et al. (2014) or Li (2016) useful.

2.2 Competing risks: definitions and common scenarios

Survival analysis is a commonly-applied statistical method in medical research. It is used for time-to-event analysis where patients are followed up to see whether, and when, they experience an event of interest. In the standard (non-competing risks) survival analysis setting there is one event of interest, such as any-cause mortality or a composite outcome combining a non-fatal event and death. Typical approaches used for analysis include the Kaplan-Meier survival estimator and Cox proportional hazard regression modelling. However, often there are situations where patients are at risk of two or more mutually exclusive events, which affect the risk of each other, and this requires a different approach. In such situations there is a competing risks scenario. The risks are said to be “competing” with each other to be the first event. For instance, two

different causes of death act as competing risks because only one of them can occur. Another example of competing risks are hospital discharge and in-hospital infection, in that discharge affects the risk of in-hospital infection by preventing it occurring first.

There has been a wealth of literature that provides an overview of the competing risks method in both the medical statistics/clinical epidemiology field [e.g. Fine and Gray (1999), Putter et al. (2007), Pintilie (2007a), Lau et al. (2009), Varadhan et al. (2010), Allignol et al. (2011), Tai et al. (2011), Andersen et al. (2012), Bakoyannis and Touloumi (2012), Koller et al. (2012), Latouche et al. (2013)] and in the wider general clinical literature [e.g. Satagopan et al. (2004), Kim (2007), Dignam and Kocherginsky (2008), Pfirrmann et al. (2011), Dignam et al. (2012), Austin et al. (2016)]. Most of the contributions giving such an overview provide a formal definition of a competing risk. A commonly used definition of a competing risk is that it is an event that precludes an event of interest. This is the sole definition used in the work by Lau et al. (2009) and Koller et al. (2012). However, this definition does not convey every scenario in which competing risks can be present. It implies only deaths can be competing risks. Gooley et al. (1999), Satagopan et al. (2004), Mell and Jeong (2010), Chappell (2012) and Wolbers et al. (2014) use the more comprehensive definition that a competing risk is an event that precludes the event of interest, or otherwise modifies the probability of experiencing the event of interest. Therefore, they recognise that competing risks need not be limited to deaths, and that non-fatal events can also act as competing risks. Similarly, Bakoyannis and Touloumi (2012) and Austin et al. (2016) use the definition that a competing risk prevents the event of interest occurring first, acknowledging that competing risks consist of non-fatal events and/or deaths.

To help fully understand the definition of a competing risk, common competing risks scenarios are illustrated in Figure 2.1. The scenarios outlined are primarily based on the scenarios described in the tutorial by Putter et al. (2007).

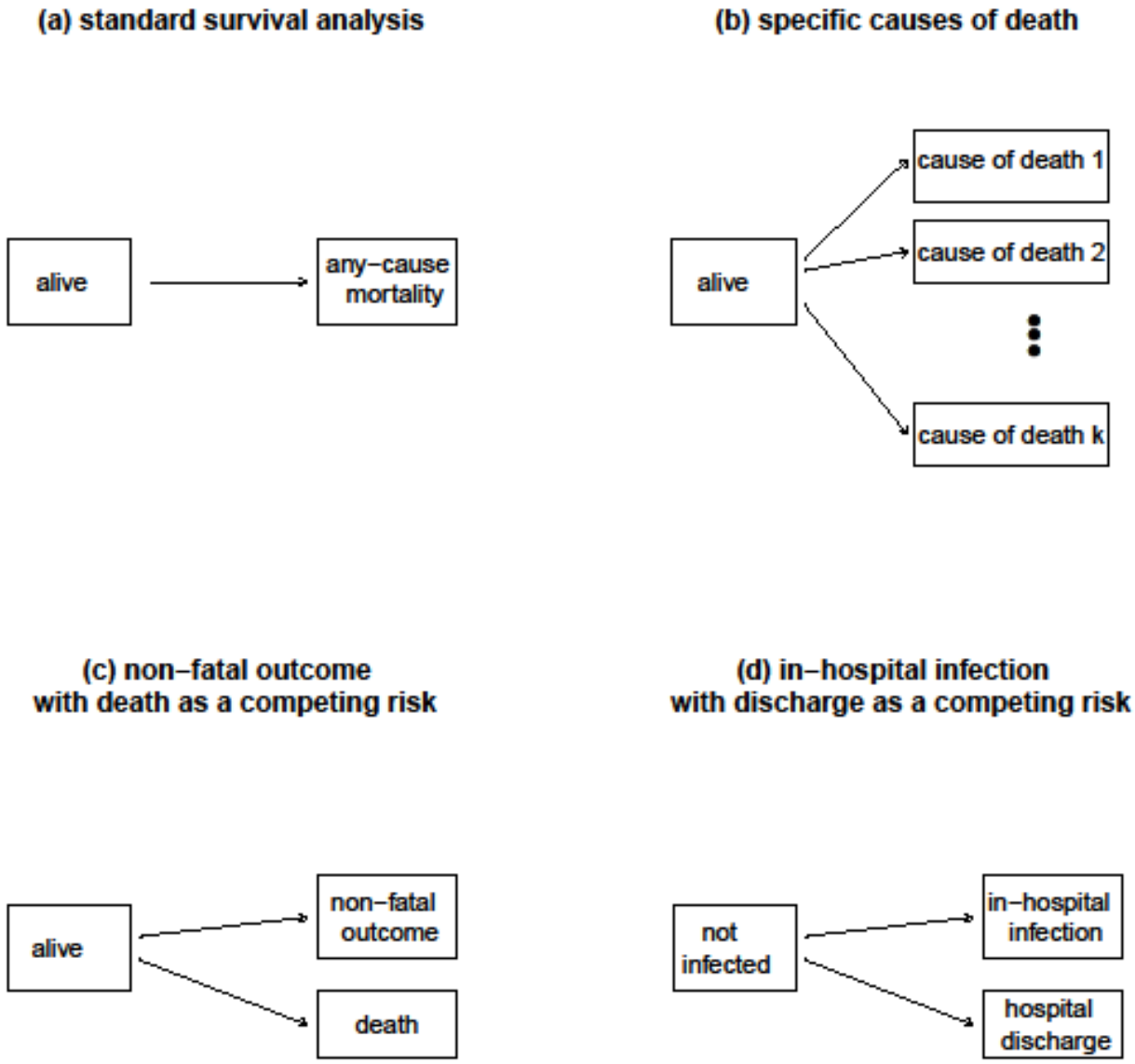


Figure 2.1 Common competing risks scenarios

Figure 2.1 (a) shows a diagram of the standard (non-competing risks) survival analysis model for comparison purposes. The typical model has every patient in an alive/healthy state at the beginning of a study. A sole transition is of interest and this is often to death, also known as all-cause or any-cause mortality. More generally, the event in standard survival analysis need not be death but could be a composite outcome comprising of a non-fatal event(s) or death.

In contrast, competing risks scenarios involve two or more mutually exclusive events. Figure 2.1 (b) depicts the situation when any-cause mortality is split into specific causes of death. This could form the basis for two different competing risks scenarios. Firstly, interest might be in disease-specific mortality and other causes of death act as competing risks. Secondly, each specific cause of death may be of interest in its own right. In that situation, each cause could be seen to be competing with the others to become the event. It might be that interest is in how frequently each cause occurred, or in assessing the influence of risk factors for each cause separately allowing them to be compared.

Figure 2.1 (c) shows another scenario that requires taking competing risks into consideration. It may be that a non-fatal outcome is the event of interest. Deaths would then need to be taken into account as a competing risk, because they would preclude the event of interest. (The only exception to this would be if no, or very few, deaths were observed. Then the situation could be thought of as involving a composite outcome of a non-fatal outcome and death, albeit with the number of deaths being zero. As an analysis of a composite outcome, standard survival analysis techniques could be employed). However, a competing risk can arise even without preclusion. For instance, as Putter et al. (2007) point out, hospital discharge can act as a competing risk when interest is in in-hospital infection (Figure 2.1 (d)). Hence, in that example, the competing risk is preventing the event of interest from occurring first.

It is important to recognise a competing risks scenario, when one exists, as this requires a different approach to standard survival analysis, as explained in the rest of this chapter. Before leaving this section it should be noted that, while there are many useful contributions that define and illustrate competing risks scenarios, there are other articles in the literature that portray competing risks in a confusing and often misleading way. For example, Tai et al. (2011) state that “the occurrence of a specific event would preclude the competing risks from being observed”. This is not inaccurate in itself because a specific event is considered a competing risk when in fact the competing risk is treated as the event of interest. However, it does not follow the usual convention that it is the competing risk that precludes, or otherwise alters the probability of, the event

of interest and not the other way round. Another piece, a commentary by Chappell (2012), emphasises that administrative censoring can be a competing risk. However, reaching the end of a study does not typically involve an event that precludes, or otherwise alters the probability of occurrence of, an outcome of interest. Therefore, this is a misleading statement. In other work, Varadhan et al. (2010) include the semi-competing risks approach in their review of statistical methods for competing risks. This introduces confusion as the semi-competing risks approach they present is not used for competing risks scenarios. It does not just consider the first event to occur from two or more mutually exclusive events. Instead, this approach also considers subsequent events. It is often known as an "illness-death model", part of the more general multi-state modelling framework. This method is the focus of Chapter 4 (methodological aspects) and Chapter 5 (empirical analysis).

2.3 Kaplan-Meier not appropriate for competing risks: introduction to competing risks cumulative incidence

The previous section highlighted that there is a need to recognise competing risks, when they exist, because a different approach to analysis is required. There is a fundamental difference between standard and competing risks survival analysis that gives rise to this. With standard survival analysis, there is a one-to-one correspondence between the hazard and survival functions. This means that when the hazard increases so does the probability (1-survival) of the event occurring. Similarly, a decrease in the hazard leads to a decrease in the probability of the event occurring. However, this one-to-one relationship does not necessarily hold in the presence of competing risks. One of the consequences of this is that the standard Kaplan-Meier analysis is no longer appropriate with competing risks, as this section will explain.

A contentious issue with competing risks is whether survival is an appropriate quantity to estimate. A few authors stipulate that the ideal estimand is *marginal* or *net* survival [e.g. (Dignam and Kocherginsky, 2008)]. Lambert et al. (2010) mention that net survival is a "measure of patient survival corrected for the effect of other causes". The authors also explain that, in the context of cancer, the net probability of death due to cancer allows comparisons to be made over

time or between places; which is of particular relevance because deaths due to other causes can also vary over time or between places and it is important that survival/death estimates are not influenced by these changes in mortality due to other causes. However the authors also mention that the net probability of death due to cancer (1- net survival) is not a true measure of the probability of death due to cancer as it assumes that deaths due to other causes do not exist. Therefore, in a competing risks setting, the term survival (from any one of the competing events) is said to only apply in a hypothetical situation. It does not have a meaningful interpretation in the real world scenarios that medicine presents, and therefore there is a general consensus that survival is not an appropriate term to use for any specific competing risk e.g. Bakoyannis and Touloumi (2012).

The marginal or net survival alluded to above could be estimated using the Kaplan-Meier method commonly applied in standard survival analysis. However, as well as only being appropriate in a hypothetical scenario, it is widely recognised in the literature that the Kaplan-Meier approach is not appropriate in the competing risks context for the reasons explained below [e.g. Gooley et al. (1999); Fine and Gray (1999), Satagopan et al. (2004), Kim (2007), Putter et al. (2007), Dignam and Kocherginsky (2008), Lau et al. (2009), Varadhan et al. (2010), Pfirrmann et al. (2011), Tai et al. (2011), Andersen et al. (2012), Bakoyannis and Touloumi (2012), Chappell (2012), Dignam et al. (2012), Koller et al. (2012), Latouche et al. (2013), Wolbers et al. (2014), Austin et al. (2016)]. Andersen et al. (2012) emphasise that this is one of the consequences of the one-to-one correspondence between the hazard and survival not holding for competing risks.

In standard survival analysis, because there is only one possible event, patients either experience the event or are censored at the time they were last known not to have the event. In a competing risks scenario, analysis is undertaken for each specific event by treating it as the event of interest. Competing events are censored because they are not the event of interest. The Kaplan-Meier approach relies on an assumption of non-informative censoring. That is to say, at the time of censoring, for patients who are censored the risk of the event should not be

any different than that for the patients who are at risk but are still in the study (i.e. uncensored). In other words, the censoring mechanism should not provide any information that influences the distribution of event times. However, by censoring any competing risks, informative censoring is introduced by the very definition of a competing risk. This is particularly evident when death is a competing risk as patients censored for this reason categorically cannot experience the event of interest, and therefore their risk is different from the uncensored patients. More specifically, if the Kaplan-Meier method is used, bias will be introduced that inflates the probability of the event of interest. This is because the uncensored individuals left in the dataset are not representative of all those who have not experienced the event. Those not at risk of experiencing the event first have been removed from the risk set through censoring. Hence the risk set contains more patients at risk of the event than it should and therefore the probability of the event of interest is inflated.

However, just like standard survival analysis, the probability at any given time of a specific event occurring is of interest when competing risks are present. An appropriate estimand for this is the cumulative incidence that takes into account competing risks. Briefly, the calculation of this involves not only the hazard of the specific event but also the hazard of each competing risk. It is explained in more detail in the next sub-section. There are many contributions in the literature that demonstrate the bias in the Kaplan-Meier approach by contrasting (the complement of) a Kaplan-Meier curve with a corresponding cumulative incidence curve that takes into account competing risks [e.g. Gooley et al. (1999), Kim (2007), Varadhan et al. (2010), Andersen et al. (2012), Bakoyannis and Touloumi (2012), Austin et al. (2016)]. By comparing the two curves, each of these authors illustrate the inflation of the cumulative incidence of the specific event that can arise if the competing risks are not taken into account appropriately.

This section has emphasised that the Kaplan-Meier (KM) approach is inappropriate in the presence of competing risks. While there is a wealth of useful contributions to the literature that raise awareness of this, there are also articles that contain misleading statements. The papers by Kim (2007) and

Satagopan et al. (2004) contain statements which are representative of the typical errors that are made. These were educational papers and have received much attention, having being cited 172 (257) and 311 (431) times respectively up to the end of 2016 according to Web of Science (Google Scholar).

The work by Kim (2007) begins by explaining that the Kaplan-Meier method is inappropriate when competing risks are present. Furthermore, it recognises that relapse and transplant-related mortality are two competing risks when studying allogeneic hematopoietic stem cell transplantation. However, the paper then makes the contradictory statement that "The KM estimate of cumulative incidence function is simple and useful for a single end point such as relapse". This is however later refuted by clarifying that the cumulative incidence of relapse is overestimated using the KM method because there are also transplant-related deaths. However, the contradictory statement could still lead to confusion. It does acknowledge that the Kaplan-Meier approach is useful for single end points. However, crucially, it fails to note that in a Kaplan-Meier analysis non-fatal outcomes such as relapse need to be combined with deaths if they occur. Related to this, the paper uses the term "relapse-free survival" when referring to the estimate using the Kaplan-Meier approach. This is an inaccurate description when the event of interest is relapse alone as the term should only be used with the composite outcome relapse or death. In addition, the illustration of the Kaplan-Meier method is that of the "naïve" Kaplan-Meier method but it is not labelled as such. Naïve Kaplan-Meier method is well-recognised terminology in the competing risks literature when demonstrating Kaplan-Meier analysis in the presence of competing risks. Therefore this contribution by Kim (2007) is inconsistent with other research. Another piece that does not use the term naïve with their illustration of the Kaplan-Meier method is the work by Satagopan et al. (2004). The paper does however explicitly convey that the analysis is not taking into account the competing risk. In a similar manner to Kim (2007) mentioned above, this work uses the term haematologic malignancy-free survival when in fact the event of interest is haematologic malignancy alone.

2.3.1 Cumulative incidence for competing risks

Earlier in this section it was highlighted that there is a need to appropriately take into account any competing risks when calculating the cumulative incidence of an event of interest. It alluded to the fact that the cumulative incidence of a specific event must take into account the hazards of each competing event, as well as that for the event of interest. They are many contributions that describe how to calculate the cumulative incidence function (e.g. Putter et al. (2007), Dignam and Kocherginsky (2008), Varadhan et al. (2010), Allignol et al. (2011), Pfirrmann et al. (2011), Andersen et al. (2012), Bakoyannis and Touloumi (2012), Dignam et al. (2012)). The formal equation to calculate the cumulative incidence, $F_k(t)$, i.e. the cumulative probability of event k having occurred in the presence of other competing events, is:

$F_k(t)$ = Pr(failure time T ≤ t, cause= k) = $\int_o^t S(u)\lambda_k(u)du$

where $S(t)$ =survival free from *any* of the events up to time t and $\lambda_k(t)$ = cause-specific hazard for the event of interest

(Dignam and Kocherginsky, 2008).

The cause-specific hazard is the key driving force behind the cumulative incidence in the presence of competing risks. Dignam et al. (2012) describe it as representing the probability of failure due to cause k at a moment in time, given that no failure of any kind has occurred thus far. More formally it is

$$\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{\text{Prob}(t \leq T < t + \Delta t, D = k | T \geq t)}{\Delta t}$$

where T is the time of failure and
D is the cause of failure (Putter et al., 2007)

It is the instantaneous rate of failing from cause k in the small time interval $[t, t + \Delta t)$, given that no failure of any kind occurred prior to time t.

The $S(t)$ element of the cumulative incidence also involves the cause-specific hazards. As outlined by Putter et al. (2007), $S(t)$ =exp(- $\sum_{k=1}^{K} \Lambda_k(t)$) where $\Lambda_k(t) = \int_0^t \lambda_k(s)ds$.

Therefore, $S(t)$ incorporates the cause-specific hazard of each event i.e. the event of interest and any competing events.

For estimation purposes, the following non-parametric plug-in estimators can be used (Bakoyannis and Touloumi, 2012):

$$\hat{F}_k(t) = \sum_{m:\, t_m \le t} \hat{\lambda}_k(t_m)\hat{S}(t_{m-1}), \qquad \text{where } \hat{\lambda}_k(t_m) = \frac{d_{km}}{n_m}, \hat{S}(t_{m-1}) = \prod_{i=1}^{m-1}\left(1 - \frac{d_i}{n_i}\right),$$

t_m is the mth-ordered failure time
d_{km} is the number of failures from cause k at t_m
d_m is the total number of failures (from any cause) and
n_m is the number of subjects at risk at t_m.

Therefore, $S(t)$ can be estimated with the widely-applied Kaplan-Meier estimator.

Furthermore, as Pfirrmann et al. (2011) emphasise, the cumulative incidence estimator above can be expressed as

$$F_k(t_m) = F_k(t_{m-1}) + \hat{p}_k(t_m) = F_k(t_{m-1}) + \hat{\lambda}_k(t_m) \times \hat{S}(t_{m-1})$$

explicitly emphasising its cumulative nature.

Many of the authors that outline how to calculate the cumulative incidence also offer a step-by-step demonstration of such a calculation [e.g. Kim (2007), Putter et al. (2007), Pfirrmann et al. (2011), Tai et al. (2011), Andersen et al. (2012)]. Each of these contributions provide a useful insight into how the cumulative incidence is influenced by the cause-specific hazard of each event. Satagopan et al. (2004) also provide a step-by-step illustration of the cumulative incidence

calculation. In their discussion, the authors state: “One minus the cumulative incidence is the probability of surviving the event of interest up to a specific time”. However, as previously mentioned on page 14, survival is not an appropriate term to use for a specific event in a competing risks context. Instead, focus should be on the cumulative incidence, i.e. probability of having the event taking into account any competing risks, such as that described in this sub-section.

Interest is often in comparing cumulative incidence curves between different groups. Gray (1988) developed a K-test for this purpose, analogous to the log-rank text for comparing Kaplan-Meier survival curves.

2.4 Competing risks modelling approaches

2.4.1 Introduction

The previous section described a non-parametric estimator of the cumulative incidence under the competing risks framework. It was non-parametric in the sense that no covariates were considered. This section describes approaches to modelling competing risks that can accommodate covariates. It begins by briefly summarising the classical latent failure times competing risks framework and some of its problems. This approach has largely been superseded by approaches based on observable quantities, a framework which is now widely considered to be the most acceptable way to carry out competing risks analysis. The section then continues by summarising the two most commonly used approaches under this framework, namely the cause-specific hazard Cox approach and Fine and Gray (1999)’s proportional subdistribution hazards model. Finally, some of the alternative approaches that can be taken are emphasised.

2.4.2 Latent failure times competing risk framework

With the latent failure times approach to competing risks, each patient is assumed to have a potential failure time for each type of failure. Only the failure that occurs first is actually observed, with the other failure times deemed to be latent. More formally, under the latent failure times framework,

it is assumed there are latent or potential failure times $Y_1, \ldots, Y_m$ for each of m failure types and that only $T=\min(Y_1, \ldots, Y_m)$ is observed.

A fundamental concept in the latent failure times approach is that of the joint survivor function $Q(y_1, \ldots, y_m : \mathbf{z}) = P(Y_1 > y_1, \ldots, Y_m > y_m ; \mathbf{z})$ where $\mathbf{z}$ is a covariate vector and Y_i is as defined above. A related concept is the marginal "survival" distribution $Q_j(y_j: \mathbf{z}) = Q(0, \ldots, 0, y_j, \ldots, 0 : \mathbf{z})$.

However, both the joint survivor function and the marginal distribution suffer from the problem of identifiability. They cannot be identified from the observed data without additional assumptions (Bakoyannis and Touloumi, 2012). One such assumption is independence of the different latent failure times. Furthermore, the assumptions are not testable. Further details of the problems with the latent failure times approach are detailed in the contributions by Tsiatis (1975), Gail (1975) and Prentice et al. (1978).

Therefore, owing to such problems with the latent failure times approach, it has been largely superseded by approaches based on observable quantities. Two such approaches are described in the following two parts of this section.

2.4.3 Cause-specific hazards Cox approach to modelling competing risks

Cox proportional hazards regression modelling is a well-recognised method in standard survival analysis. It involves modelling, for a patient with covariate values $\mathbf{Z} = (Z_1, \ldots, Z_p)$, the hazard $\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\boldsymbol{\beta}^T \mathbf{Z})$

where $\boldsymbol{\beta}$ is a vector of regression coefficients and $\lambda_0(t)$ is the baseline hazard. $\boldsymbol{\beta}^T\mathbf{Z}$ is shorthand for $\sum_{k=1}^{p} \beta_k \times Z_k$ where p = number of parameters in the model.

An analogous Cox regression approach can also be applied when competing risks are present. The approach is then called cause-specific hazards Cox regression. It involves fitting a separate Cox regression model of the hazard for each of the competing events (causes). The cause-specific hazard of cause k, for a patient with covariate vector $\mathbf{Z}$, is

$$\lambda_k(t|\mathbf{Z}) = \lambda_{k,0}(t)\exp(\boldsymbol{\beta}_k^T \mathbf{Z})$$

where $\lambda_{k,0}(t)$ is the baseline cause-specific hazard of cause k

and $\boldsymbol{\beta}_k$ represents the covariate effects on cause k.

The cause-specific hazard is the instantaneous rate of experiencing cause k amongst those who are event-free (i.e. not yet had cause k or any of the competing events). A straightforward way of applying this cause-specific hazards approach is to fit a separate Cox model for each cause, censoring any competing events at their time of occurrence. An alternative way to implement cause-specific hazard modelling is to use a "data-augmentation" method. This offers more flexibility and may help to overcome problems of overfitting. It allows inclusion of an interaction between a covariate and the cause, thereby enabling direct evaluation of the difference in the (relative) effect of a covariate between the causes. In particular, it allows for hypothesis testing of an effect of an interaction, aiding a decision of the modeller as to whether a common effect of the covariate across all causes may suffice. It also allows the effect of a covariate on one cause to be proportional to the effect on another cause, providing another solution to overfitting. The data-augmentation approach involves setting up the data in long format, with each patient having as many rows as the number of causes. Since patients can only experience one cause in a competing risks scenario, any competing causes are censored at the time of occurrence of the cause experienced. Any patients who do not experience any of the events have censored times in each of their rows, with the censoring occurring at the end of their follow-up. Covariate information is recorded for each cause. The original covariate values in the data are used for the cause that is actually experienced. For competing causes, all covariate values are set to zero. These cause-specific covariates allow for cause-specific modelling to be undertaken using one model rather than separate modelling for each cause. The modelling also has the flexibility to consider different covariates for each cause-specific hazard. More details of the implementation and flexibility offered by the data-augmentation/cause-specific hazards approach is available in the work by Lunn and Mcneil (1995) and Putter et al. (2007).

2.4.4 Fine and Gray's subdistribution proportional hazard modelling approach to competing risks

As previously mentioned in section 2.3, the one-to-one correspondence between the hazard and survival that exists with standard survival analysis does not necessarily hold when competing risks are present. As a consequence, the effect of a covariate on the cause-specific hazard for a particular cause may be different from its corresponding effect on the probability of the event occurring. To overcome the related problems with interpretation with the cause-specific hazards approach, Fine and Gray (1999) developed an alternative method that retains a one-to-one link. Instead of the hazard, they introduce the concept of a subdistribution hazard that has a one-to-one correspondence with the cumulative incidence of the event. The subdistribution hazard of cause k is defined as:

$$\lambda_k(t;\mathbf{Z}) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{t \le T \le t + \Delta t, \varepsilon = k \mid T \ge t \cup (T \le t \cap \varepsilon \ne k), \mathbf{Z}\}$$

$$= \mathrm{d}F_k(t;\mathbf{Z})/\mathrm{d}t\ \}\ /\{1 - F_k(t;\mathbf{Z})\}$$

$$= -\mathrm{dlog}\ \{1 - F_k(t;\mathbf{Z})\}/\mathrm{dt}$$

where T is the failure time

$\varepsilon \in (1, \ldots, K)$ is the cause of failure

$\mathbf{Z}$ is the $p \times 1$ bounded time − independent covariate vector

and $F_k(t;\mathbf{Z}) = \Pr\{t \le T, \varepsilon = k \mid \mathbf{Z})$

The risk set associated with this subdistribution hazard is somewhat counter-intuitive. Patients who experience a competing event are retained in the risk set, even though the occurrence of their event results in zero, or otherwise altered, probability of the event under consideration. This is in contrast to the cause-specific hazards approach which censors such patients at the time of occurrence of the competing event. However, just like standard survival analysis and the cause-specific hazard approach, the subdistribution hazard can be modelled in a proportional hazards framework via:

$\lambda_k(t;\mathbf{Z}) = \lambda_{k0}(t)\exp(\mathbf{Z}^T\boldsymbol{\beta})$, where $\lambda_{k0}(t)$ is completely unspecified.

It can also be extended to time-varying covariates **Z(**t), using

$$\lambda_k(t; \mathbf{Z}) = \lambda_{k0}(t)\exp(\mathbf{Z}^T(t)\,\boldsymbol{\beta})$$

with $F_k(t; \mathbf{Z}) = 1 - \exp[-\int_0^t \lambda_{k0}(s)\exp\{\mathbf{Z}^{\mathbf{T}}(s)\boldsymbol{\beta}\}\,ds\,]$

In practical terms, the method of implementation of the modelling of the subdistribution hazard for a particular cause depends on the reason for any censoring that occurs. If the censoring is purely administrative, then the modelling can be carried out using Cox proportional hazards regression. Any competing events are censored at a time just after the last observed event of the type under consideration. This way the competing events are retained in the risk set with a time for the event under consideration that is essentially infinity. This corresponds to the representation that the subdistribution hazard λ_k can be thought of as acting on the improper random variable T* = $\mathrm{I}\,(\varepsilon = \mathrm{k}) \times \mathrm{T} + \{1 - \mathrm{I}(\varepsilon = \mathrm{k})\} \times \infty$. The cumulative incidence function $F_k(t; \mathbf{Z})$ is itself an improper probability distribution in that $F_k(\infty; \mathbf{Z})$ < 1 i.e. it never reaches 1.

When the censoring is instead right-censoring due to loss of follow-up, weighting of the censoring is used. Fine and Gray (1999) adapt inverse probability of censoring weighting (IPCW) techniques and incorporate them into their modelling of the subdistribution hazard. An estimate of the survivor function of the censoring distribution is used to reweight contributions to the risk set for events due to competing causes.

2.4.5 Other approaches to modelling competing risks

This thesis applies the two main approaches to competing risks regression described above. Alternative approaches and extensions to modelling of cause-specific hazards and of cumulative incidences are described elsewhere. Contributions that may be of interest include those involving extending the standard case, when all event times are observed exactly or right-censored, to interval-censored and possibly truncated data [Hudgens et al. (2001), Hudgens et al. (2014) and Li (2016)]. In addition, alternative approaches proposed include vertical modelling (Nicolaie et al., 2010) and mixture models (Lau et al., 2008).

Furthermore, methods that can be used when the proportional hazards assumption is not met include those involving flexible modelling of competing risks [Scheike and Zhang (2008) and Belot et al. (2010)], stratified competing risks regression (Zhou et al., 2011) and modelling of time-varying cause-specific hazard ratios (Sun et al., 2008). Lastly, extensions to modelling of the cumulative incidence function include using pseudo values (Klein and Andersen, 2005) and addressing missing causes of failure (Bakoyannis et al., 2010).

2.5 The effect of a covariate on the cause-specific hazard is not necessarily the same as its effect on the cumulative incidence

As previously mentioned, the one-to-one correspondence between the hazard and survival that exists with standard survival analysis may no longer hold when competing risks are present. This is a fundamental difference that affects the analysis and interpretation of competing risks, and it has two main consequences. Section 2.3 focused on the first consequence that Kaplan-Meier analysis is inappropriate with competing risks. This section focuses on the second consequence, that the effect of a covariate on the cause-specific hazard may be different from the corresponding effect on the cumulative incidence. There are many contributions in the literature that illustrate this important aspect of competing risks, e.g. (Putter et al., 2007, Lau et al., 2009, Allignol et al., 2011, Latouche et al., 2013, Wolbers et al., 2014, Austin et al., 2016).

Table 2.1 summarises, for two competing events, the probable effects of a covariate on the cumulative incidences, under different scenarios of effects of the same covariate on the cause-specific hazards.

scenarios		effect of covariate on cause-specific hazard (csh) of:		probable impact on cumulative incidence of:	
		event of interest	competing event	event of interest	competing event
covariate has no effect on csh of either event	A1	0	0	0	0
covariate has an effect on csh of one event only	B1	↓(↓↓)	0	↓(↓↓)	↑
	B2	↑(↑↑)	0	↑(↑↑)	↓
	B3	0	↓(↓↓)	↑	↓(↓↓)
	B4	0	↑(↑↑)	↓	↑(↑↑)
covariate affects csh of each event but in opposite directions	C1	↑	↓(↓↓)	↑	↓(↓↓)
	C2	↓	↑(↑↑)	↓	↑(↑↑)
	C3	↑↑	↓(↓↓)	↑↑	↓(↓↓)
	C4	↓↓	↑(↑↑)	↓↓	↑(↑↑)
covariate affects csh of each event but in the same direction	D1	↓	↓↓	↑	↓↓
	D2	↓↓	↓	↓↓	↑
	D3	↑↑	↑	↑↑	↓
	D4	↑	↑↑	↓	↑↑
	D5	↓	↓	↓	↓
	D6	↓↓	↓↓	↓↓	↓↓
	D7	↑	↑	↑	↑
	D8	↑↑	↑↑	↑↑	↑↑

Abbreviations used:
csh = cause-specific hazard
0 = covariate has no effect
↑ = covariate associated with an increased effect
↑↑ = covariate associated with a more pronounced increase
↓ = covariate associated with a reduced effect
↓↓ = covariate associated with a more pronounced reduction

Table 2.1 Effect on the cause-specific hazard versus that on the cumulative incidence: different scenarios

Throughout this thesis, the hazard ratio is the effect that is being represented by any mention of the phrase "the effect of a covariate on the cause-specific hazard". That is to say that phrase, and similar wording, is being used as shorthand for the effect on the cause-specific hazard of a particular level of covariate relative to the reference level of that covariate.

The scenarios can be classified as follows:

2.5.1 the covariate does not affect the cause-specific hazard of either event (scenario A1)

In a situation where a covariate has no effect on either the cause-specific hazard of the event of interest or the competing event then, obviously, the covariate will not have any effect on the cumulative incidence of either event (Dignam et al. (2012)). However, this scenario is rarely encountered because a covariate that has no effect on either event is unlikely to be included in any modelling.

2.5.2 the covariate has an effect on one event only (scenarios B1-B4)

The 'B' section of Table 2.1 outlines scenarios where a covariate affects the cause-specific hazard of one event but not the other. Each row in the B section condenses together two different scenarios for brevity, distinguishable by whether the bracketed or unbracketed term is considered in those cells which contain both. For instance in row B1, ignoring the bracketed terms outlines the following scenario: the covariate decreases the hazard of the event of interest and has no effect on the hazard of the competing event, which in turn results in that covariate being associated with a probable decrease in the cumulative incidence of the event of interest and a probable increase in the cumulative incidence of the competing event. Considering the bracketed term instead of the unbracketed term results in the aforementioned effects on the event of interest being replaced by those that are more pronounced.

When a covariate has no effect on the cause-specific hazard of a competing event, Allignol et al. (2011) stress that the interpretation is straightforward of how the effect of the covariate on the cause-specific hazard of an *event of interest* translates to a corresponding effect on its cumulative incidence (scenarios B1 and B2). When outlining this aspect in their simulations, the authors state that the direction of the effect on the cause-specific hazard determines the direction of the effect on the cumulative incidence. When the effect on the hazard is more pronounced, the impact of the effect on the

cumulative incidence is also larger (bracketed B1 and B2 scenarios in Table 2.1). However this ease of interpretation only applies to the event of interest.

Wolbers et al. (2014) emphasise that, even though there is no effect on the cause-specific hazard of the competing event, the covariate *can* have an effect on the *cumulative incidence* of the competing event. The authors provide a useful clinical illustration to help explain scenarios B1 in Table 2.1. Wolbers et al. (2014) present an example concerned with the effect of the binary treatment covariate (intervention versus control) on two events. Coronary heart disease (CHD) death is the event of interest and non-CHD death is the competing event. The intervention is found to reduce the cause-specific hazard of CHD death but to have no effect on the cause-specific hazard of non-CHD death. The authors explain that the reduction seen in the cause-specific hazard of CHD death for those with the intervention leaves more patients vulnerable to the force that draws them towards non-CHD death. Therefore, even though the intervention does not affect the cause-specific hazard of non–CHD death, it is expected to increase the cumulative incidence (absolute risk) of non-CHD death.

A similar argument explains scenarios B2 ((Lau et al., 2009, Dignam et al., 2012). Dignam et al. (2012) present the results of a prostate cancer trial focused on the effect of treatment, age and tumour grade on death due to prostate cancer and death due to other causes. Compared to a grade 1 tumour, having a grade 3 tumour appears at first to have a protective effect on the cumulative incidence of death due to other causes. This effect is seen even though tumour grade is not found to have an association with the cause-specific hazard of death due to other causes. However, a grade 3 tumour is associated with an increase in the cause-specific hazard of prostate cancer death. Therefore, the lack of increase in the absolute risk of death due to other causes is actually due to there being fewer patients surviving prostate cancer to be at risk of death due to other causes.

There are analogous arguments when a covariate has no effect on the cause-specific hazard on the event of interest but does affect the cause-specific hazard of the competing event (scenarios B3 and B4). Specifically, the direction

of the effect on the cause-specific hazard of the competing event is mirrored in the corresponding effect on the cumulative incidence of the competing event. Furthermore, the covariate can have an effect on the cumulative incidence of the event of interest, even when it has no effect on its cause-specific hazard. This is typically in the opposite direction to that seen on the cumulative incidence of the competing event.

Andersen et al. (2012) present an example that aids understanding by providing an explanation in terms of risk sets. Their example involved patients with chronic myeloid leukaemia who had an allogeneic stem cell transplant and originated from European Group for Blood and Marrow Transplantation (EBMT) data. The illustration investigated the effect of EBMT risk score on the events relapse and non-relapse mortality (NRM). It was found that there was no evidence of a difference between those with score 0-1 and 2 in the cause-specific hazard of relapse, while there was an increased cause-specific hazard of NRM for those with score 2 compared to those with score 0-1. Of note, while there was no effect on the cause-specific hazard of relapse, there was a lower cumulative incidence of relapse for those with score 2 compared to those with score 0-1. The authors point out that the cause-specific hazard acts on those patients still in the risk set. Their explanation continues by outlining that because more with score 2 experience NRM, over time, the risk set of those with score 2 decreases faster than the risk set of those with score 0-1. Consequently, in absolute terms, there are fewer with score 2 having relapse and therefore the cumulative incidence of relapse is lower in that group.

2.5.3 the covariate affects the cause-specific hazard of each event but in opposite directions (scenarios C1-C4*)*

Allignol et al. (2011) stress that the interpretation of how the effect of a covariate on the cause-specific hazards translates to effects on the cumulative incidences is again straightforward when the former have opposing effects for the two events. For each of the two events, the direction of the effect on the cause-specific hazard is mirrored in the direction of the effect on the cumulative incidence (Allignol et al., 2011, Latouche et al., 2013). Latouche et al. (2013) present an example involving patients aged at least 50 years old with acute

myeloblastic leukemia who received hematopoietic stem cell transplantation. The two treatment regimes investigated were reduced intensity conditioning (RIC) and myeloablative conditioning (MAC) and the two events of interest were relapse and treatment-related mortality (TRM). Treatment regime had similar sized effects on the two events, but in opposite directions. RIC was found to increase the cause-specific hazard of relapse but decrease the cause-specific hazard of TRM. These effects were mirrored in the corresponding results for the cumulative incidence of each event. Because the analyses based on the cause-specific hazard and cumulative incidence gave consistent results, the authors concluded that the effect of treatment regime on the cumulative incidence of relapse was an actual effect and not due to an indirect effect on the competing event TRM.

2.5.4 the covariate affects the cause-specific hazard of each event in the same direction (scenarios D1-D8)

Allignol et al. (2011) highlight that unidirectional covariate effects on cause-specific hazards constitute the most challenging scenario in terms of understanding the corresponding effects on the cumulative incidences. In particular, they outline the situation when having a specific covariate level reduces the cause-specific hazard of both the event of interest and the competing event, but the effect for the competing event is more pronounced (scenario D1 in Table 2.1). Allignol et al. (2011) and Wolbers et al. (2014) emphasis a difficulty in interpretation that may arise from this. Wolbers et al. (2014) discuss an example involving the effect of an intervention on CHD death and non-CHD death. The intervention reduces both the cause-specific hazard of CHD death and non-CHD death, but that decrease is more pronounced for non-CHD death. They highlight that the intervention may be associated with an increase in the cumulative incidence of CHD death, even though it was associated with a reduction in the cause-specific hazard of the event. This would likely be due to the more dramatic decrease in the hazard of non-CHD death meaning there would be more patients surviving to be at risk of a CHD death.

When the more pronounced reduction in cause-specific hazard is seen for the event of interest (scenario D2), the analysis may indicate an increase in the

cumulative incidence of the competing event associated with the specified level of the covariate.

Bakoyannis and Touloumi (2012) use a synthetic example to demonstrate that cumulative incidence curves can cross in scenario D3. They outline an investigation into the effect of gender on the event of interest treatment interruption (TI) and the competing event switching to a new regime (NH) among HIV seropositive patients. The authors fix the coefficient of gender (female versus male) at 0.531 for TI and vary the corresponding coefficient for NH between -0.018 and 1.75. It is shown that, as the coefficient of gender on NH increases, the gap between the curves narrows, and when the coefficient for NH becomes more and more pronounced (1.5 and 1.75) the curves actually cross. While the cumulative incidence of TI is initially higher for women, it is higher for men as time progresses. This more pronounced effect of the competing risk NH means more women experience NH – therefore precluding TI as the first event for women – leading to the crossing of the curves and the eventual higher cumulative incidence of TI in men.

The crossing of curves may also be apparent when a specified level of a covariate is associated with a more pronounced increase in the cause-specific hazard of a competing event (scenario D4).

In another illustration, Dignam and Kocherginsky (2008) describe a trial involving early stage breast cancer patients comparing the effect of tamoxifen versus placebo on the four events breast cancer recurrence, contralateral breast tumours, endometrial cancer and other events. It was found that the cumulative incidence of other events was lower in the tamoxifen arm initially, before becoming higher as time passed, and hence the curves crossed. The authors explain that this is likely due to more of those in the placebo arm having a recurrence or contralateral breast tumour first, whereas in the tamoxifen arm there are relatively more patients remaining at risk of experiencing an other event first. The number of susceptible patients in each group to each event, over time, is the likely reason for the eventual crossing of the curves.

When the effects of a covariate on the cause-specific hazard of each event are in the same direction and of similar size (scenarios D5-D8), then the effects on the cumulative incidences of each event are likely to be in this same direction. However, the effect on each cumulative incidence is likely to be attenuated compared to the corresponding effect on the cause-specific hazard. This is due to the covariate having an effect on the other event, and the stronger this association the more the effect on the cumulative incidence of the event of interest lessens.

This section has emphasised that, when competing risks are present, the effect of a covariate on the cumulative incidence for a particular event is influenced not only by the effect on the cause-specific hazard of the event, but also by the effect on the cause-specific hazard of any competing events. An aspect that is not always highlighted in the literature is that the baseline hazard of the event of interest, and any competing events, can also affect the cumulative incidence of an event. This is not immediately apparent from an analysis of data from a trial or study. However, many contributions show simulations where the baseline hazards are varied to illustrate the impact this can have on the cumulative incidence, e.g. Putter et al. (2007).

2.6 Derivation of the cumulative incidence using cause-specific hazard modelling

Section 2.4.4 described Fine and Gray's subdistribution proportional hazards model. In particular, it emphasised that the subdistribution hazard has a one-to-one correspondence with the cumulative incidence function. However, there is some confusion in the literature as to whether the cause-specific hazard approach to modelling competing risks can also be used to calculate the cumulative incidence. The previous section highlighted that, when the cause-specific hazard for a particular competing risk is considered in isolation, the effect of a covariate on that cause-specific hazard may not be reflected in the corresponding effect on the cumulative incidence for the same competing risk. However, as previously explained in sub-section 2.3.1, if the cause-specific hazards of each of the competing risks are considered, the cumulative incidence for a particular competing risk can be estimated. Even so, many of the

contributions fail to recognise that the cumulative incidence can be derived from cause-specific hazard modelling [e.g. Kim (2007), Latouche et al. (2013) and Austin et al. (2016)]. Kim (2007) stipulates that cause-specific hazards Cox modelling is not suitable for competing risks analysis due to the censoring of competing risks. It is implied that this is the case for the calculation of the cumulative incidence. However, the argument is not conveyed explicitly as relating to a "naïve" cause-specific hazard approach, i.e. when the cumulative incidence is inappropriately estimated using the cause-specific hazard of one event in isolation. The paper just categorically states that cause-specific hazard modelling is inadequate, rather than explaining how the cumulative incidence could be calculated using the modelling from each of the events. The article does not provide this information, even though it is illustrated earlier in the paper that the non-parametric estimation of the cumulative incidence involves the hazards of each of the competing risks. Other work (Latouche et al., 2013, Austin et al., 2016) does acknowledge that the "naïve" cause-specific approach is inappropriate for calculating the cumulative incidence. However the papers do not expand on this and outline how the cumulative incidence can be appropriately estimated from the cause-specific hazards.

As part of their contributions, other authors do recognise that the cumulative incidence can be derived from cause-specific hazard modelling [e.g. Putter et al. (2007), Lau et al. (2009), Varadhan et al. (2010), Allignol et al. (2011) and Andersen et al. (2012)]. Putter et al. (2007) point out that the confusion has arisen because analysts have become accustomed to the one-to-one correspondence between the hazard and survival in standard survival analysis. While this may no longer hold with competing risks, the authors maintain there is nothing fundamentally wrong with using the cause-specific hazards approach. It is just that the interpretation of the results requires more care.

Another contentious issue with cause-specific hazards modelling is whether it requires the competing risks to be independent of each other, in the sense that the occurrence of one does not affect the probability of others occurring. Pintilie (2007a) asserts that it does. However, Andersen et al. (2012) explain that the concept of independence in competing risks analysis is a throw-back to

the classical latent failure time approach which was discussed in sub-section 2.4.2. This approach is now widely thought to be outdated and has been superseded with approaches based on observable quantities such as cause-specific hazards and cumulative incidences.

Several authors criticise aspects of Pintilie's work, including that it is based around a latent failure time approach (Pintilie, 2007b, Pintilie, 2007c). In their letter to the editor Latouche et al. (Pintilie, 2007c) in particular object to Pintilie's use of the term "cause specific hazard" with the latent failure time approach. They highlight that the article adds confusion because cause-specific hazard is terminology that is normally reserved for transition intensities in the more accepted approaches based on observable quantities. Related to this, they take issue with Pintilie's statement that "When modelling the cause specific hazard, one performs the analysis under the assumption that competing risks do not exist". This assumes independent latent failure times, which is untestable.

On the contrary, Andersen et al. (2012) argue that inference in the conventional cause-specific hazard analysis makes no assumption about the independence of competing risks. Instead, the authors reiterate that the cumulative incidence can be estimated by taking into account the cause-specific hazard of each event. They then emphasise that the cause-specific hazard for a particular event can be calculated by censoring competing events. Furthermore, they stress that this censoring of competing events works because cause-specific hazards are time-local rates of occurrence of events that are mutually exclusive, and hence the likelihood factorises.

2.7 Variable selection strategies with competing risks

While there are other variable selection strategies, likelihood-based approaches such as likelihood-ratio tests and information criteria, e.g. Akaike information criterion (AIC) and Bayesian information criterion (BIC), are popular due to being intuitive and easy to use. They are widely used in standard survival analysis, just as they are in other common modelling techniques in statistics. However, variable selection is less straightforward with competing risks survival analysis. This section discusses variable selection in the context of two common

approaches to modelling competing risks. Sub-section 2.7.1 focuses on Fine and Gray's proportional subdistribution hazards model and sub-section 2.7.2 outlines considerations when using the cause-specific hazards approach.

2.7.1 Variable selection and Fine and Gray's proportional subdistribution hazard approach

Fine and Gray's proportional subdistribution hazards model was previously described in sub-section 2.4.4. Motivated by the importance of parsimonious and accurate models to predict risk in medical research, and the lack of methods for covariate selection with Fine and Gray's model, Kuk and Varadhan (2013) developed an approach to address this issue. The approach is that of an information criteria-based stepwise regression procedure for use with Fine and Gray's model. The authors developed a version of their technique based on the well-recognised AIC and BIC and a new criteria that they propose called the BICcr. The BICcr is similar to the BIC, and differs only in the term used for penalisation. The penalisation term in the BIC is the familiar $p + \log n$, where p is the number of parameters in the model and n is the number of observations. In contrast, the penalisation term in the newly proposed BICcr is $p + \log n^*$ where p is as before and n^* is the number with the event of interest. Therefore the penalty is more lenient with BICcr than it is with BIC.

Kuk and Varadhan's approach is used to decide which covariates to include in the Fine and Gray models presented later in this thesis in section 3.3. However, while their approach is one of stepwise variable selection it does not operate like a standard likelihood-ratio test. This is because it does not produce a statistic that follows a (asymptotic) chi-square distribution, or indeed that assumes any particular distribution. Therefore the standard likelihood-ratio test based approach for obtaining an overall p-value for a covariate with more than two levels is not applicable. Therefore, in the presentation in section 3.3, no overall p-values are shown for such covariates.

2.7.2 Methodological issues relating to variable selection with the cause-specific hazards approach

Kuk and Varadhan's approach outlined in the previous sub-section was made possible because of the one-to-one correspondence between the subdistribution hazard and the cumulative incidence. However, as mentioned previously in this thesis, such a correspondence may no longer exist between the hazard and probability/cumulative incidence in the presence of competing risks. Section 2.5 described in detail a consequence of this, namely that the effect of a covariate on the cause-specific hazard may not necessarily translate into the same effect on the cumulative incidence. This is due to the cumulative incidence for an event of interest not only being affected by the cause-specific hazard of the event, but also by the cause-specific hazard of any competing events. Therefore, I would argue that considering the cause-specific hazard of one event in isolation from those of the competing events is not to be recommended. Consequently, likelihood-ratio tests and information criteria focused solely on one cause-specific hazard should be treated with caution in the presence of competing risks. The results are liable to mis-interpretation and misleading conclusions. Considering one cause-specific hazard in isolation is analogous to naïve Kaplan-Meier/Cox predictions. Such naïve predictions introduce inflated probabilities of the event of interest due to not taking into account the competing risks. Therefore for the cause-specific hazard modelling approach to competing risks presented in Chapter 3, no formal variable selection method is used to choose between the models for prediction. Instead, the variables selected are based upon those chosen for the two approaches involved in the comparison, namely the composite outcome and Fine and Gray's proportional subdistribution hazards model.

2.8 Two schools of thought on the use of two common approaches to competing risks

This section outlines the two schools of thought in the literature on the use of cause-specific hazard modelling and Fine and Gray's proportional subdistribution hazard model. The first school of thought is that the research question should drive the modelling approach used. Alternative thinking is that the cause-

specific hazard and cumulative incidence of each competing risk should be considered together in order to gain a full understanding of the competing risk process. The idea behind each is discussed in this section.

2.8.1 the research question should drive the modelling approach used

Several contributions advocate that the research question should drive the approach to analysis, e.g. Dignam and Kocherginsky (2008), Lau et al. (2009), Dignam et al. (2012), Koller et al. (2012), Wolbers et al. (2014) and Austin et al. (2016). The premise of this argument is that the cause-specific hazard and subdistribution hazard/cumulative incidence estimate different quantities. The cause-specific hazard is the instantaneous rate of experiencing the event of interest among those who are still at risk of the event i.e. among those who have survived free of *any* of the events. However, as previously outlined in section 2.5 of this thesis, the effect of a covariate on the cause-specific hazard of a specific event may not translate into the same effect on its cumulative incidence. This is due to the one-to-one correspondence between the hazard and probability seen in standard survival analysis, not necessarily holding when competing risks are present. This problem with the cause-specific hazard motivated an approach that modelled subdistribution hazards, with such a quantity having a one-to-one link with the cumulative incidence. However, the subdistribution hazard for an event of interest involves a somewhat counterintuitive risk set. This is because those who experience a competing event are retained in the risk set, even though they cannot then possibly experience the event of interest as their first event.

Because the cause-specific hazard and subdistribution hazard/cumulative incidence illustrate different aspects of a competing risks process, it naturally follows that the specific research question of interest can motivate the approach to analysis. In particular, due to the nature of the cause-specific hazard described above, etiology-based questions are most appropriately addressed by modelling cause-specific hazards. In other words, the cause-specific hazard approach can be useful if interest is in investigating whether there is a relative effect of a covariate on event rates (hazards) of a specific event of interest. In

contrast, modelling the cumulative incidence facilitates the estimation on the absolute scale of the effect of a covariate over time. An effect on the cumulative incidence for an event of interest could be due to a direct effect of that covariate on its cause-specific hazard, and/or due to an indirect effect that acts on the cause-specific hazard of a competing event. Therefore, as highlighted by Wolbers et al. (2014), modelling of the cumulative incidence seems the more pertinent approach for clinical decision making because it refers to absolute risks in the real-world where competing events do occur. Consequently, modelling of the subdistribution hazard/cumulative incidence is deemed the most appropriate approach to address questions of evaluating treatment policy in populations, for example risk/benefit trade-off in a health economics context, (Dignam and Kocherginsky, 2008) and for clinical predictions and risk-scoring systems (Austin et al., 2016).

The contribution by Lau et al. (2009) provides a useful illustration with example research questions that can be posed, and suggestions of the most appropriate approach to take to address them. The authors' example is that of the use of the antiretroviral drug abacavir which has been found to be associated with an increased risk of myocardial infarction (MI). They outline two competing research questions that can be posed:

1) Is the use of abacavir directly associated with MI?

2) Regardless of the direct association, are individuals taking abacavir more likely to experience an MI?

The authors then continue by suggesting that the first question might be most appropriately answered by modelling the cause-specific hazards. They emphasise that the cause-specific hazard approach to modelling allows assessment of whether, at any given time, patients taking the drug have an increased instantaneous hazard rate for MI among all patients who have survived up to that point event-free. Conversely, for the second question, their article suggests that modelling subdistribution hazards would be more appropriate. They illustrate a related possible situation and emphasise that it may help decide policy. The situation is that when taking the drug has no effect on the cause-specific hazard

of an MI, but that the subdistribution hazard of MI for those taking the drug is increased. They go on to explain that this is due to a reduced hazard of the competing event death without MI in those taking the drug, which means more of such patients are available to experience an MI.

2.8.2 both the cause-specific hazard and subdistribution hazard/ cumulative incidence should be modelled for each competing risk

Another school of thought is that both the cause-specific hazard and subdistribution hazard/cumulative incidence should be modelled for each competing risk [e.g., Andersen et al. (2012), and Bakoyannis and Touloumi (2012), Latouche et al. (2013) and Austin et al. (2016)]. This is advocated as a way of gaining a fuller understanding of the competing risks process. This is because, as mentioned previously, the cause-specific hazard is a key driving force in analysis that takes into account competing risks. The calculation of the cumulative incidence for a specific event involves the cause-specific hazard of each event (i.e. that of interest and any that compete). Considering both the cause-specific hazard and cumulative incidence together for every event can aid interpretation in competing risks analysis. It can provide a better appreciation of the results and the insights they give because the two outcomes complement each other. They each relay a related but different aspect of the competing risks process and together give a comprehensive picture of it. As mentioned previously in this section, an effect on the cumulative incidence of an event of interest could be due to a direct effect of that covariate on its cause-specific hazard, and/or due to an indirect effect that acts on the cause-specific hazard of a competing event. Furthermore, the effect of a covariate on the cause-specific hazard of an event may not translate into the same effect on the corresponding cumulative incidence. This phenomenon was explored more fully in section 2.5 of this chapter. It will also be demonstrated empirically later in this thesis in chapter 3.

2.9 Composite outcomes vs decomposition into competing risks

Combining endpoints into composite outcomes is popular in randomised trials and other studies. This amalgamation of multiple events to maximise event rates has many advantages. These include the related concerns of greater precision in estimates of effects and increased statistical power, which both improve efficiency of the study. They can also be useful when it is unclear which specific outcome to focus on, and in particular when the combination of potentially important outcomes is thought to describe a disease process. However, as Allignol et al. (2011) emphasise, medical problems are typically too complex to be addressed by the analysis of one combined event.

Furthermore, it is well-recognised in the literature that analysis of composite outcomes may only be advantageous when a covariate affects each of the individual components in the same direction (e.g. Varadhan et al. (2010)). When this is not the case, analysis that splits the outcome into its component competing risks is generally considered the most appropriate approach to take. Mell and Jeong (2010) explain how, in a competing risks scenario, using a composite outcome may actually reduce the efficiency of a trial. They highlight that patients at high risk of a competing event may be less likely to benefit from the treatment. Furthermore, they use the term “deadweight” in relation to such patients because including them in a trial may reduce efficiency of a trial and make it more costly, if the treatment being studied in the trial does not reduce the incidence of the competing outcome.

Decomposing a composite outcome into its component parts may provide further insight into the effect of a covariate on the separate events. For instance, it may help identify which of the particular events a covariate affects and in which direction. Separating an outcome into several cause-specific ones may help to assess the effect of a covariate on each cause, and therefore not attribute an effect to the wrong cause as could be the case with a composite outcome. In a paper by Mell et al. (2012), the authors discuss the issue of decomposing all-cause mortality into cancer- and non-cancer mortality. They highlight that it can

be beneficial to analyse each outcome specifically in order to ascertain the effectiveness of a treatment against cancer.

Another potential insight of competing risks analysis becomes apparent when it is evident that an analysis of a composite outcome alone would mask the effects on its component parts. That is to say, a covariate might have opposing effects on two component parts of a composite outcome, and as a consequence there may be no effect on the composite outcome due to the opposing effects cancelling each other out. In other instances, using a composite outcome may mask an effect on a specific event of interest if it is in the opposite direction to the effect on the composite outcome. This is particularly likely to be the case when the incidence of a competing event is high and therefore dominates the composite outcome, but the effect on the competing outcome is in the opposite direction to the event of interest.

The empirical analysis presented in chapter 3 will demonstrate the extra insight that can be gained by decomposing a composite outcome into its competing risks components.

2.10 Need for better recognition of competing risks in the clinical community

This chapter so far has emphasised the need to use an analysis approach that takes into account competing risks, if such a scenario exists. It has also highlighted the key issues that need to be considered when faced with competing risks. Sub-section 2.10.1 will demonstrate that there has been an emergence of competing risks in the literature, particularly in the last 10-15 years. Sub-sections 2.10.2 and 2.10.3 will then focus on a review of the contributions in the specific area of stroke epidemiology, and the competing risks of recurrence and death. The purpose of this is to identify gaps and issues in the literature that the empirical analysis in the next chapter will try to address.

2.10.1 Review of recognition of the need for competing risks analyses, where appropriate

A search was carried out to find literature that reviewed the extent of use of competing risks analyses (Appendix I). Two contributions were found (Mell et al. (2012) and Koller et al. (2012)). With their reviews, these authors have raised awareness of the need to recognise competing risks and analyse them appropriately. It was previously mentioned in section 2.9 that Mell et al. (2012) emphasised the need to decompose a composite outcome, when the treatment had differing effects on the component parts of the outcome. The authors' review focused specifically on cancer trials, and whether effects on the component parts of cancer events and non-cancer related mortality were analysed separately. The authors found this only to be the case in 40% (47/118) of studies.

The work by Koller et al. (2012) was a useful contribution that provided a more comprehensive review into the use of competing risks analyses. Their research began by emphasising that aging populations in particular necessitate considering competing risks analyses. The authors explained that an aging population is susceptible to competing risks due to a high level of comorbidities. As life expectancy continues to improve, this is only likely to increase further. Furthermore, the article included a literature search carried out to see how frequently competing risks were mentioned in biostatistical, core clinical and general high impact clinical journals. Their search strategy is detailed in Appendix II. I reproduced their search based on the criteria specified in Appendix II, and updated it by extending the dates of publication to 31 December 2015. Figure 2.2(a), (b) and (c) summarise the findings by year for biostatistical, core clinical and general high impact clinical journals respectively. It should be noted that the vertical axis of each of the Figures has a different scale.

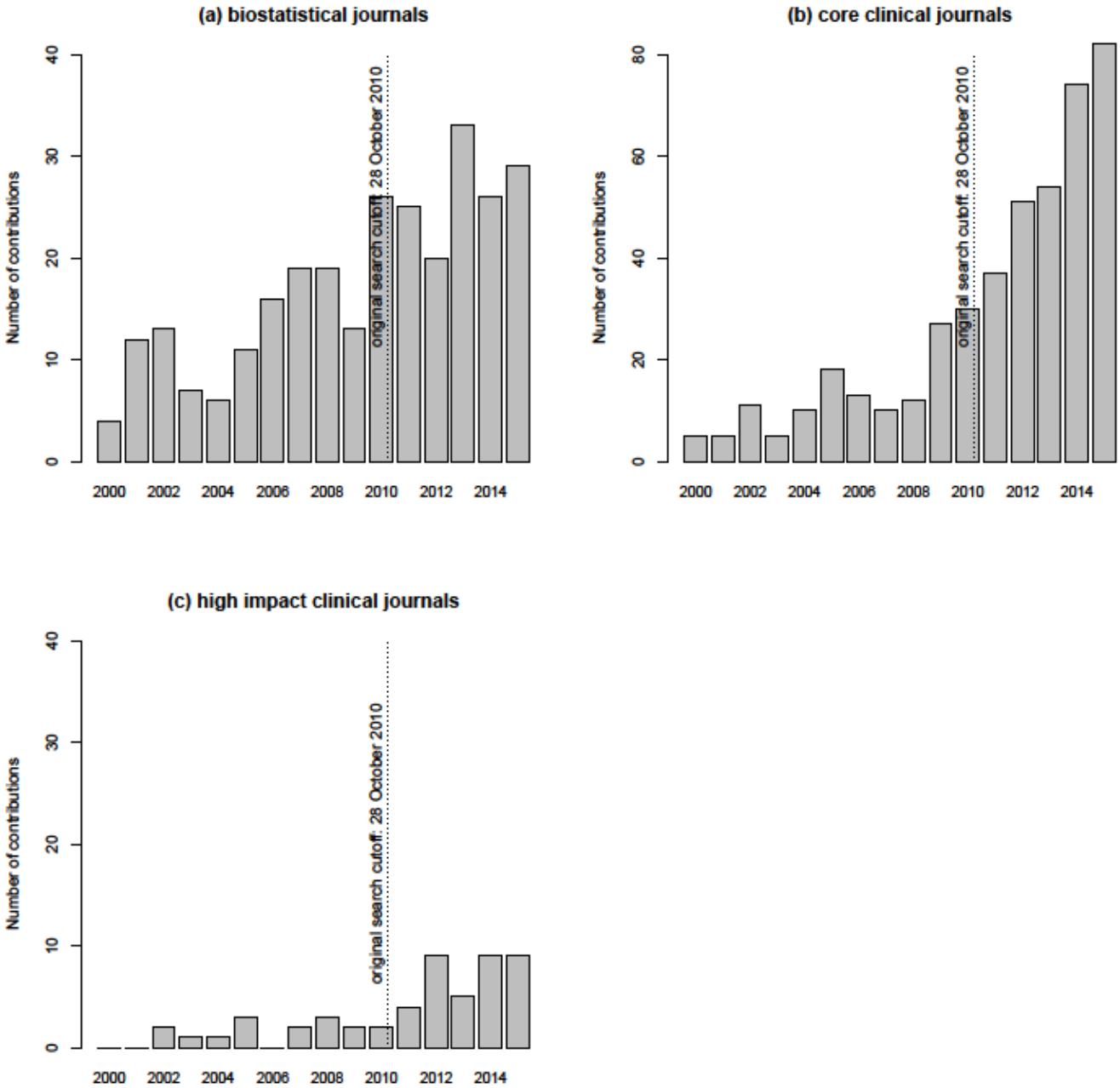


Figure 2.2 original and updated search of competing risks literature based on the criteria of Koller et al. (2012)

The original search found that, for the biostatistical and core clinical journals, there was generally a steady rise in the number of contributions over the decade starting 2000. In the updated search, this trend was continued in the core clinical journals while for the biostatistical journals there was a suggestion of levelling off. However, the appearance of competing risks in general high impact clinical journals was low. The original search only found between 0 and 3

articles published per year in such journals. The updated outlook also found small numbers, only achieving at most 10 articles per year.

As part of their original research, Koller et al. (2012) also explored how competing risks issues were treated in high impact clinical journals. The authors searched for articles focusing on diseases that are prevalent or typical in aging or multimorbid patients, and that would have follow-up that could potentially raise competing risks issues. They then concentrated on critically appraising 50 articles, paying particular attention to the following:

- the use of the naïve Kaplan-Meier approach to estimate cumulative incidence when competing risks were actually present

- neglecting competing risks either by not reporting competing events, or by reporting them but not analysing them

They found that at least one of the issues was apparent in 70% (35/50) of the studied articles.

2.10.2 Review of competing risks analyses of stroke recurrence and death in the literature

This sub-section describes a review of the literature that focused on the analysis of the competing risks recurrence and death in stroke. The purpose of this was to explore the existing use of, and in particular the approach used for, competing risks in this area. It provides the motivation for the empirical analysis of the stroke case study in the next chapter. A search strategy was developed to allow the review to take place. When this research began in 2011 an OvidSP search of titles, abstracts and keywords of contributions using relevant search terms was performed. However, this search has been an ongoing process. Email alerts of new articles were set up using Ovid and Web of Knowledge. A final OvidSP search was conducted to search contributions up to 31 August 2016. The entire search found 17 contributions and is summarised in Appendix III. Appendix IV contains an evidence table summarising what was found, which will be discussed in this sub-section.

In their study protocol, Wollenweber et al. (2014) state that they intend investigating post-stroke dementia as a primary endpoint. However stroke recurrence is also stipulated as a secondary endpoint. Unfortunately the paper does not provide any detail of a plan of analysis for stroke recurrence and only maintains that a very vague "competing risks analysis" will be carried out.

Some of the contributions have studies with endpoints of all-cause mortality and stroke recurrence (Sun et al., 2013, Arntz et al., 2014, Dhamoon et al., 2016a, Dhamoon et al., 2016b). These authors use a standard Kaplan-Meier and/or Cox regression to analyse all-cause mortality and a Fine and Gray proportional subdistribution hazards model to analyse stroke recurrence. In the latter, death is treated as a competing risk when modelling the subdistribution hazards of recurrence. However, these papers do not present a corresponding Fine and Gray model for death without recurrence, with recurrence treated as a competing risk. Therefore, the contributions have missed the opportunity to convey the extra insight that may have been gained from the analysis of the other competing risk. For example, in the research by Dhamoon et al. (2016a), modelling the other competing risk may have clarified and provided further explanation for the result that South Asians had a higher cumulative incidence of recurrence than non-South Asians. Given that ethnicity had no effect on recurrence-free survival, modelling of death without recurrence (treating recurrence as a competing risk) is likely to have shown that South Asians had a lower cumulative incidence of death without recurrence. This is not explicitly evident from the modelling of all-cause mortality because this shows death overall. The modelling of death without recurrence as a competing risk could have helped explain the higher cumulative incidence of recurrence in that less patients were dying in such a manner and so more were alive to be at risk of recurrence.

The analyses of all-cause mortality and those for just one competing risk in the four aforementioned contributions, fail to recognise that recurrence and death without recurrence act as competing risks and should be analysed separately and compared. Furthermore, the analysis of all-cause mortality does not distinguish between deaths with and without recurrence. Analysis involving the three events

recurrence, death after recurrence and death without recurrence simultaneously is the focus of Chapter 4. Instead of the analyses carried out, or ideally alongside them, it would have been preferable to also show the analysis of death without recurrence, with recurrence treated as a competing risk.

Only three of the contributions used Fine and Gray's model for both the outcomes recurrence treating death without recurrence as a competing risk and death without recurrence treating recurrence as a competing risk (Lewsey et al., 2010, Andersen et al., 2011, He et al., 2015). The research by Lewsey et al. (2010) was the only contribution found from this literature search to explicitly acknowledge that results should be shown for each competing risk.

Several of the contributions use a Cox (cause-specific hazards), rather than a Fine and Gray, model for their competing risks analysis (Ovbiagele, 2012, Stamplecoski et al., 2012, Ovbiagele, 2013, Castilloux et al., 2015, Choi et al., 2016). Choi et al. (2016) use the test devised by Gray (1988) to compare competing risks cumulative incidence curves at a univariable level and take a Cox approach to multivariable modelling. The methods section of the paper states that deaths are treated as a competing risk. However, the paper also states that Fine and Gray's proportional subdistribution hazards model was used, and then proceeds to show Cox regressions in the results. Therefore, this could cause confusion.

Ovbiagele (2013) uses backwards stepwise variable selection to decide which covariates to include in the competing risks Cox regression. However, likelihood ratio tests need to be treated with caution in competing risks scenarios, as explained in section 2.7.2, especially when only considering the results from one competing risk in isolation. Furthermore, the paper appears to show "naïve" Cox predictions, because they are described as cumulative incidence curves derived using the Cox regression previously presented.

The work by Rutten-Jacobs et al. (2013) uses the test devised by Gray (1988) to compare competing risks cumulative incidence curves for at most one covariate, and then Fine and Gray's model at a multivariable level. The authors present separate results for recurrent strokes and for other arterial events alongside

those for the composite of the two. It would have been preferable to display results separately for recurrent strokes, for arterial events and for deaths. This is especially the case as death was a strong competing risk with 172 (23.8%) of the 724 patients dying. Furthermore, a more appropriate composite outcome would have combined all three events rather than excluding the deaths.

2.10.3 Review of the possibility of decomposing recurrence-free survival after stroke in the literature that uses that outcome

The previous sub-section reviewed the approach used for competing risks analysis of stroke recurrence and death in contributions that included such terminology. This sub-section will describe a review of the literature that included the outcome recurrence-free survival. In particular, for each contribution, the number of stroke recurrence and death events will be identified. Then a discussion will follow on whether decomposition of the composite outcome recurrence-free survival could have been an option, if it was not carried out. A search strategy was developed to allow the review to take place. In a similar manner to the search described in sub-section 2.10.2, the search was an ongoing process after an initial search when this research first began in 2011. A final OvidSP search of titles, abstracts and keywords of contributions to 31 August 2016 using the search terms

(stroke AND recurrence-free survival).ab,ti,kw.

was carried out. The entire search found 20 contributions and is summarised in Appendix V. Appendix VI contains an evidence table summarising what was found.

2.10.3.1 Contributions that may have benefited from decomposition of the outcomes

In the study by Elneihoum et al. (1998), the Kaplan-Meier and Cox regression approaches were used for all-cause mortality. Furthermore, a Cox regression was also presented for non-fatal recurrences, but it was not clear whether death was treated as a competing risk in the modelling. The 959 (43.4%) deaths would have

been a strong competing risk for the 137 (6%) of recurrences. In addition, a plot of the probability of recurrence is shown. However, the paper does not mention any appropriate competing risks approach and therefore it was probably based on the naïve Kaplan-Meier method. Finally the label RR, for relative risk, was used in the Cox results table. This is technically incorrect for recurrence because it was the hazard which was modelled, not the risk (i.e. not the probability).

In the research by Ogasawara et al. (2002b), the authors use the Kaplan-Meier method for recurrence-free survival and a Cox regression for recurrence. However, in the latter the model included 10 covariates when there were only 11 recurrences. There were too few events to warrant this number of covariates given the standard rule of thumb of 10 events per covariate (Peduzzi et al., 1995). Furthermore, no details were provided of how deaths were treated in the Cox regression. The Kaplan-Meier plots presented suggest there were as many deaths as recurrences and therefore these would need to be treated as a competing risk. It is likely that the authors would have censored deaths when analysing recurrences. However clarification of this would have been beneficial due to the competing risks scenario.

The contribution by Hillen et al. (2003a) used Kaplan-Meier analysis for recurrence-free survival. However, it also used a Kaplan-Meier approach for stroke recurrence, which is inappropriate with the paper failing to recognise that a more suitable method that takes into account competing risks should be used. This was particularly applicable to this study because death without recurrence was a strong competing risk, at 48.7% by 5 years, for the 153 (16.6%) of recurrences by 5 years. However, the authors do specify that deaths are censored in their parametric proportional hazards model for recurrence, which is the appropriate approach. That said, the paper also states that likelihoods were used to decide between parametric distributions and likelihood ratio tests to select covariates, both of which can be dubious in a competing risk scenario.

The article by Hillen et al. (2003b) used a Cox regression for recurrence-free survival. However among the 561 patients in their study there were 66 recurrences and 146 deaths. This number of events warrants decomposing the

composite outcome into the two competing risks recurrence and death without recurrence. It could provide more insight in the study, especially as stroke severity may have had opposing effects on recurrence and death, which might have been cancelling each other out.

In their research Yokota et al. (2004) use Kaplan-Meier and Cox regression for recurrence-free survival. In a similar manner to Elneihoum et al. (1998) above, the label RR is used in the Cox results table, which is the wrong terminology in a competing risks context as previously explained. The patients in this study had 198 recurrences and 286 deaths within 3 years. Therefore, there were enough of each event to consider decomposition of recurrence-free survival, which may have provided more insight.

In the contribution by Lee et al. (2010), the methods section states that a Kaplan-Meier analysis and Cox regression was used for recurrence specifically. However, the Kaplan-Meier curve shown in the results is actually for recurrence-free survival. If the article wished to present the cumulative incidence of recurrence specifically an appropriate method to take into account competing risks should have been used. Furthermore, it was not clear how deaths were treated in the Cox regression, and the number of deaths was not mentioned. However the Kaplan-Meier curve of recurrence-free survival appeared to be at around 80% by 5 years, and given that 14(7.7%) of the 181 patients had a recurrence, this left around 12% of the patients that died. Therefore, deaths were a strong competing risk for recurrence and the approach to analysis should reflect this. It would have been preferable to have clarification that this was the case.

The work by Toschke et al. (2011) used a Cox regression for all-cause mortality and for recurrence. However, it was not clear whether death was treated as a competing risk in the modelling of the hazard of recurrence. The paper includes curves for the estimated probability of survival and for the estimated probability of stroke recurrence. It would appear from the survival curve, that among the 3690 patients who survived 90 days, the probability of death within 5 years was around 15-30% depending on use of antihypertensive treatment. From the

recurrence curve, it would appear that the probability of recurrence within 6 years was around 35-38% for the 3121 patients surviving 1 year. However, this recurrence curve presents naïve Cox predictions and as such is likely to have introduced inflation bias. It would have been more appropriate to also present Cox modelling of the hazards of the other competing risk, and then combine the hazards appropriately to create the predictions. Alternatively, the Fine and Gray approach could have been used to present the cumulative incidence of recurrence taking into account competing risks. Even with the inflation bias, there was still likely to have been enough patients with recurrence and death without recurrence to analyse these two competing risks appropriately. A further limitation of this study was that only recurrences that occurred after 1-year post stroke were considered, ensuring it was a recurrence that was captured rather than just a repeat recording of the index stroke, but this wasted valuable information on recurrences in the intervening period.

Chan et al. (2012)'s work used the Kaplan-Meier method for the two outcomes recurrence-free survival and major adverse cardiovascular events (MACE)-free survival. A Cox regression was used for the outcomes stroke recurrence and MACE. Among the patients in their study there were 10 recurrences, 12 MACEs and 8 deaths. Therefore deaths acted as a strong competing risk for recurrence and needed to be addressed appropriately. However, it was not clear whether competing risks were censored in their Cox models.

2.10.3.2 Contributions not likely to benefit from decomposition of the outcomes

Nadeau et al. (1992) used an unconventional way to present one of their outcomes. Overall survival was presented in the usual way by showing a Kaplan-Meier curve. However for the other outcome the "percentage of patients free of non-stroke death who were free of stroke recurrence" is presented, which is somewhat confusing. This would appear to be a round-about way of expressing recurrence-free survival. However the paper also states "In the recurrence studies, non-stroke deaths were treated in the same-way as withdrawals." This implies deaths were censored in the analysis of recurrence instead of included to

make a composite outcome. Regardless, there were too few events in the study to consider anything beyond a Kaplan-Meier analysis.

In their contribution, Yokota et al. (1998) use a Kaplan-Meier analysis for recurrence-free survival and a Cox regression for recurrence specifically. However the authors were limited in the number of covariates they could consider due to only having 13 recurrences and 11 deaths in their study.

The Kaplan-Meier and Cox regression approaches were used for recurrence-free survival in the research by Ogasawara et al. (2002a). Their study had 70 patients and 13 recurrences and 4 deaths. Given the small number of events, decomposition of recurrence-free survival was unlikely to provide any additional insight.

In other research, Marnane et al. (2010) also used a Kaplan-Meier to analyse recurrence-free survival. Cox regression was again mentioned but it was unclear whether this was for recurrence alone or the composite outcome, because the abstract contained contradictory statements. With only 10 recurrences, there were not enough events to benefit from decomposition of the outcome. In addition, the use of 9 covariates in the Cox regression with such a small number of events was inappropriate.

Kuwashiro et al. (2012) used a Kaplan-Meier approach for recurrence-free survival and logistic regression for the outcome recurrence by 12 months. The authors aimed to investigate factors associated with recurrence specifically. Their logistic regression approach was not unreasonable because there were not many deaths. Given the low number of deaths, obviously decomposition was not worth considering. However, the study only had 25 recurrences but 6 covariates in the model. The use of 6 covariates with 25 events was inappropriate given the standard rule of thumb of Peduzzi et al. (1996) that also applies to logistic regression.

The contribution by Kim et al. (2014) used a Kaplan-Meier approach for recurrence-free survival and a Cox regression for stroke recurrence. The paper emphasises throughout the objective is to assess the risk of recurrence. In doing

so it fails to recognise that a Cox regression assesses the hazard which is different from the risk. In addition, the article does not mention whether there were any deaths in the study. If there were no or very few deaths then a logistic regression would have been a more sensible approach given that their outcome was at 90 days, a fixed point in time. That said, a Cox and logistic regression are unlikely to provide results with substantive differences, provided there are no/very few deaths and there is (near) complete follow-up of all patients.

In another piece, Fujimoto et al. (2015) used Kaplan-Meier and a Cox regression for recurrence-free survival. The authors were unlikely to gain any additional insight by decomposing this outcome due to relatively few deaths.

2.10.4 Summary of the review

This section focused on an area where competing risks are not widely applied, specifically recurrence and death in stroke patients. The review in sub-section 2.10.2 of contributions that involved competing risks analyses found that either Fine and Gray's proportional subdistribution hazard model or cause-specific hazard Cox modelling was adopted. However, the presentation of analysis of both competing risks was very limited among the contributions using Fine and Gray's model. Furthermore, none of the authors that used the cause-specific hazard Cox approach showed analysis for each competing risk, opting instead to just show modelling of recurrence. However, earlier in this chapter it was emphasised that the cause-specific hazard is a key driving force in the calculation of the cumulative incidence. Furthermore it was stressed that, to fully understand the effects on the outcomes of interest, the cause-specific hazard and cumulative incidence for each of the competing risks should be analysed. Therefore, Chapter 3 of this thesis will add to the existing research and present such a fully comprehensive analysis of the competing risks stroke recurrence and death.

Sub-section 2.10.3 which focused on contributions that involved recurrence-free survival also found that, when authors used Cox regression for recurrence specifically, they neglected to present a corresponding analysis for death without recurrence. However, when modelling hazards of recurrence analysts

are likely to be censoring deaths, and therefore inadvertently taking into account competing risks appropriately even if they are unaware that such a scenario exists. What could cause a problem however is when analysts try to derive cumulative incidences/probabilities from this and present naïve Cox predictions. Chapter 3 will also include a section highlighting the bias that can occur by inappropriately using a naïve Kaplan-Meier or naïve Cox approach. Sub-section 2.10.3 also highlighted some instances when the opportunity was missed to decompose recurrence-free survival into the two competing risks recurrence and death without recurrence. Since some covariates are likely to have effects in opposite directions for the two competing risks, Chapter 3 will also compare analysis of the composite outcome with those of decomposition into the two competing risks.

Chapter 3 Competing risk analysis with the Edinburgh Stroke Study

The previous chapter described the background to the competing risks approach, highlighted methodological issues that need to be considered and reviewed the literature into the extent of its use. In particular, the review focused on stroke recurrence and death, an area where the technique has not been used to its full potential. All of this motivated this current chapter which demonstrates empirical competing risks analyses with a stroke case study. This chapter presents various different aspects of competing risks analysis to illustrate the insights that can be gained over using standard survival analysis.

The chapter begins by providing an overview of the case study, the Edinburgh Stroke Study, in section 3.1. Next, section 3.2 compares analyses of the composite outcome recurrence or death with those of the separate components recurrence and death without recurrence. The purpose of this is to illustrate the extra insight that can be gained by decomposing a composite outcome into its component competing risks. The focus of this section is Cox regression for the composite outcome and cause-specific hazards Cox regression for each of the two competing risks recurrence and death without recurrence. Then, section 3.3 presents the Fine and Gray subdistribution hazards modelling approach to competing risks. The results are presented alongside those using the cause-specific hazards approach demonstrated earlier in section 3.2. In particular, section 3.3 also includes cumulative hazard and cumulative incidence plots. The purpose of section 3.3 is to illustrate the greater understanding that can be gained of a competing risks process by considering both the hazard and cumulative incidence of each competing risk. Next, section 3.4 illustrates the consequences of not taking into account competing risks. Specifically, this section compares naïve Kaplan-Meier/Cox predictions to those that appropriately address that competing risks are present. Finally, section 3.5 summarises the analysis presented in this chapter and the extra insight gained over standard survival analysis.

3.1 Overview of the Edinburgh Stroke Study

All analysis in this chapter and chapter 5 was carried out using data from the Edinburgh Stroke Study (ESS), a case study used to demonstrate the potential in stroke epidemiology of the methods described in this thesis. ESS was a prospective, hospital-based cohort study of stroke patients followed-up with multiple overlapping methods for recurrent stroke, myocardial infarction, and death. The methods and process of data collection are described elsewhere (Jackson et al., 2009). 1237 patients who presented with a definite or probable stroke (excluding a subarachnoid haemorrhage) between the years 2002 and 2005 were included in the analysis. These patients were followed-up for between 1 and 4 years, and details of any stroke recurrence and/or death were obtained. Table 3.1 on the next page shows the patient characteristics used in the analysis.

The second column in Table 3.1 shows the distribution of covariates among the stroke patients. The remaining columns show, for each event of interest, the frequency of the event in patients with a particular covariate level, with the information with and without the characteristic shown for the binary covariates. While not the focus of this chapter, death after recurrence is considered in chapter 5 and is shown in Table 3.1 for completeness. To guide the reader, the third row shows that 621 of the patients were male. It also shows that 11.8% of males and 15.3% of females had recurrences, 18.7% of males and 23.4% of females died without recurrence and 42.5% of males and 34.0% of females died after a recurrence. Information on other rows of the table can be read similarly. The frequencies are shown as an indication of the (unadjusted) covariate effects on the events and should not be thought of as incidence in the usual sense due to censoring. However, they do show the effect of each of the covariates in absolute terms, ahead of the modelling of hazards later in the chapter which will be in relative terms.

Previous cerebrovascular event was defined as previous stroke or transient ischaemic attack. Previous other occlusive vascular disease consisted of previous ischaemic heart disease or peripheral arterial disease. Previous hypertension was

	n (%)	% experiencing event amongst those with(without) characteristic[2]		
		recurrence	death without recurrence	death after recurrence
Total [1]	**1237 (100%)**	**167 (13.5%)**	**260 (21.0%)**	**63 (37.7%)**
age at stroke - mean(s.d.)	71 (12)	72 (10)	78 (10)	76 (9)
male	621 (50.2%)	11.8% (15.3%)	18.7% (23.4%)	42.5% (34.0%)
independent before stroke	1147 (92.7%)	13.9% (8.9%)	18.2% (56.7%)	36.5% (62.5%)
previous cerebrovascular event	373 (30.2%)	16.4% (12.3%)	22.8% (20.3%)	37.7% (37.7%)
previous other occlusive vascular disease	864 (69.8%)	16.4% (12.2%)	27.8% (18.0%)	53.2% (28.6%)
previous hypertension	378 (30.6%)	15.1% (11.8%)	20.0% (22.1%)	37.8% (37.7%)
diabetes	150 (12.1%)	15.3% (13.2%)	22.7% (20.8%)	47.8% (36.1%)
non-smoker	436 (35.2%)	12.4% (14.1%)	25.0% (18.9%)	35.2% (38.9%)
lift both arms off bed	1041 (84.2%)	15.0% (5.6%)	14.2% (57.1%)	35.9% (63.6%)
walk independently	876 (70.8%)	14.4% (11.4%)	12.2% (42.4%)	29.4% (63.4%)
orientated speech	1038 (83.9%)	14.0% (11.1%)	14.6% (54.3%)	33.1% (68.2%)
stroke syndrome				
Cortical	655 (53.0%)	14.0%	26.6%	44.6%
Lacunar	324 (26.2%)	12.7%	11.4%	34.1%
Other	258 (20.9%)	13.2%	19.0%	23.5%
high blood pressure	444 (35.9%)	12.8% (13.9%)	21.4% (20.8%)	28.1% (42.7%)
delay in assessment				
0-1 days	282 (22.8%)	13.5%	34.4%	47.4%
2-6 days	300 (24.3%)	14.3%	25.3%	48.8%
7 days or more	655 (53.0%)	13.1%	13.3%	27.9%
symptomatic carotid stenosis				
<70% ICA	904 (73.1%)	12.7%	15.2%	31.3%
70-100% ICA	144 (11.6%)	22.2%	20.1%	50.0%
Unknown	189 (15.3%)	10.6%	49.7%	55.0%
in atrial fibrillation	257 (20.8%)	16.7% (12.7%)	36.2% (17.0%)	53.5% (32.3%)
visible infarct on scan	714 (57.7%)	13.7% (13.2%)	22.5% (18.9%)	43.9% (29.0%)
haemorrhage on scan	88 (7.1%)	17.0% (13.2%)	35.2% (19.9%)	53.3% (36.2%)

1 When symptomatic carotid disease was not recorded it was treated as a separate 'Unknown' category. This was because the information was not thought to be missing at random; instead the patients concerned were thought to be more likely to be too ill to have this assessed. When stroke syndrome was not recorded (55 patients) it was included in the category 'Other', which also consisted of patients with posterior circulation syndrome (POCS). For all other variables, patients without a value recorded were included in the category of the variable that was most common. The latter was the case for independent before stroke (6), previous cerebrovascular event (3), previous other occlusive vascular disease (2), previous hypertension (1), smoking status (24), lift both arms off bed (4), walk (7), orientated speech (6), high blood pressure (11), in atrial fibrillation (74), visible infarct on scan (26) and haemorrhage on scan (26) . There were no unrecorded values for age, gender, diabetes and delay in assessment.
2 with (without) characteristic shown for binary variables, and each level shown on separate rows for variables with 3 levels. mean (s.d.) are shown for the continuous variable age

Table 3.1 Patient characteristics involved in analyses of Edinburgh Stroke Study

defined as a history of treated hypertension. High blood pressure was defined as a systolic blood pressure in excess of 160mmHg or a diastolic blood pressure in excess of 80mmHg at the time of initial assessment. Symptomatic carotid disease was defined as the percentage stenosis in the internal carotid artery (ICA) on the same side as the brain lesion. If the lesion was not specific to one side, or the side was unknown, the largest stenosis was used. A patient was

considered to be in atrial fibrillation if they had a history of atrial fibrillation or it was present on ECG. Orientated speech was defined as able to talk and orientated in time, place and person. The date of initial assessment was chosen as the starting date rather than the date of stroke onset to exclude recurrences before baseline information could be collected. If patients had more than one stroke before initial assessment the stroke immediately before initial assessment was selected as the index stroke. This was the case for two patients. The illness-death modelling presented in Chapter 5 did not allow entry to different health states at the same time. Therefore patients with a recurrence on the same day as death were assumed to have died 0.5 days later than their recurrence. This was the case for four patients.

3.2 Analyses of the composite outcome versus decomposition of the outcome: cause-specific hazards approach to competing risks

It could be seen in the previous section that there were 167 recurrences and 260 deaths without recurrence among the 1237 patients. It is common in studies - and trials in particular - to study composite outcomes such as recurrence-free survival i.e. the composite event recurrence or death. However, decomposing a composite outcome into its component competing risks can often provide additional insight. This is particularly true when a covariate affects two different competing risks, but the effects are in opposing directions. In addition, it can be the case that one particular event is of interest but a competing event occurs more often. When this happens the analysis of the composite outcome may be dominated by factors associated with the competing event, and associations with the event of interest may become masked. In the Edinburgh Stroke Study, recurrences were of primary interest. Given that there were only 167 recurrences but 260 deaths without recurrence, there was potential to gain insight from an analysis that decomposed the composite outcome into these two competing risks. In addition, it was thought that there may be some covariates that would affect each competing risk differently. This provided the motivation for the analysis in this section that compares analyses of the composite outcome recurrence or death with those of the separate components recurrence and death without recurrence.

Table 3.2 shows the effects of the covariates on the hazard of each of the events at the univariable (unadjusted) level.

		Cox regression of recurrence or death		cause-specific hazard competing risks Cox modelling			
	Baseline	hazard ratio for recurrence or death without recurrence (95% CI)	p-value	hazard ratio for recurrence (95% CI)	p-value	hazard ratio for death without recurrence (95% CI)	p-value
age - centred on mean of 71		**1.05 (1.04, 1.06)**	**<0.001**	**1.02 (1.00, 1.03)**	**0.012**	**1.08 (1.06, 1.09)**	**<0.001**
male	Female	**0.74 (0.61, 0.90)**	**0.002**	**0.72 (0.53, 0.98)**	**0.038**	**0.76 (0.59, 0.97)**	**0.025**
independent before stroke	No	**0.36 (0.27, 0.48)**	**<0.001**	1.12 (0.55, 2.27)	0.763	**0.24 (0.18, 0.33)**	**<0.001**
previous cerebrovascular event	No	*1.22 (1.00, 1.49)*	*0.055*	*1.34 (0.98, 1.84)*	*0.066*	1.14 (0.88, 1.47)	0.329
previous other occlusive vascular disease	No	**1.65 (1.36, 2.01)**	**<0.001**	**1.53 (1.12, 2.09)**	**0.008**	**1.74 (1.36, 2.23)**	**<0.001**
previous hypertension	No	1.07 (0.88, 1.29)	0.504	*1.33 (0.98, 1.81)*	*0.069*	0.93 (0.73, 1.18)	0.543
diabetes	No	1.19 (0.90, 1.57)	0.223	1.23 (0.79, 1.91)	0.352	1.16 (0.81, 1.67)	0.417
current smoker and ex-smoker <12 months	Non-smoker	*0.85 (0.70, 1.03)*	*0.097*	1.09 (0.79, 1.51)	0.602	**0.73 (0.57, 0.93)**	**0.011**
lift both arms off bed	No	**0.30 (0.24, 0.37)**	**<0.001**	*1.70 (0.92, 3.14)*	*0.089*	**0.16 (0.13, 0.21)**	**<0.001**
walk independently	No	**0.37 (0.31, 0.45)**	**<0.001**	0.94 (0.66, 1.33)	0.714	**0.22 (0.17, 0.28)**	**<0.001**
orientated speech	No	**0.27 (0.22, 0.33)**	**<0.001**	0.77 (0.49, 1.20)	0.245	**0.17 (0.13, 0.22)**	**<0.001**
stroke syndrome			**<0.001**		0.448		**<0.001**
Lacunar	Cortical	**0.52 (0.40, 0.67)**	**<0.001**	0.79 (0.55, 1.15)	0.218	**0.38 (0.26, 0.54)**	**<0.001**
Other	Cortical	**0.75 (0.59, 0.96)**	**0.021**	0.89 (0.60, 1.31)	0.551	**0.68 (0.49, 0.93)**	**0.015**
high blood pressure	No	1.04 (0.85, 1.26)	0.728	0.98 (0.71, 1.34)	0.881	1.08 (0.84, 1.38)	0.573
delay in assessment			**<0.001**		0.267		**<0.001**
2 - 6 days	0-1 days	**0.70 (0.55, 0.90)**	**0.005**	0.90 (0.58, 1.39)	0.634	**0.63 (0.46, 0.85)**	**0.002**
7 days or more	0-1 days	**0.42 (0.34, 0.53)**	**<0.001**	0.74 (0.51, 1.09)	0.125	**0.30 (0.22, 0.40)**	**<0.001**
symptomatic carotid stenosis			**<0.001**		**0.013**		**<0.001**
70-100% ICA	<70% ICA	**1.63 (1.23, 2.15)**	**0.001**	**1.86 (1.26, 2.75)**	**0.002**	**1.43 (0.96, 2.13)**	**0.082**
Unknown	<70% ICA	**3.11 (2.49, 3.88)**	**<0.001**	**1.21 (0.75, 1.94)**	**0.435**	**4.68 (3.59, 6.09)**	**<0.001**
in atrial fibrillation	No	**2.20 (1.79, 2.70)**	**<0.001**	**1.65 (1.17, 2.34)**	**0.005**	**2.60 (2.01, 3.35)**	**<0.001**
visible infarct on scan	No	*1.20 (0.99, 1.46)*	*0.068*	1.10 (0.81, 1.50)	0.527	*1.26 (0.98, 1.62)*	*0.067*
haemorrhage on scan	No	**1.99 (1.46, 2.70)**	**<0.001**	*1.64 (0.96, 2.78)*	*0.069*	**2.21 (1.52, 3.22)**	**<0.001**

Table 3.2 Univariable Cox composite vs cause-specific hazard competing risks Cox modelling results

The first section of Table 3.2 shows the results of Cox regression modelling of the composite outcome regression or death without recurrence. The rest of

Table 3.2 displays the results of the Cox cause-specific hazard approach to modelling the competing risks recurrence and death without recurrence. Covariates found to be statistically significant at the 5% level are shown in bold. Those covariates found to be statistically significant at the 10% level (marginally significant) are shown in italics. Alternate shading of the rows is used to help distinguish between the rows. The usual appropriate checks for violations of proportional hazards were carried out, although they are not shown for reasons of brevity.

It can be seen in Table 3.2 that advancing age, being female, having other occlusive vascular disease, having high symptomatic carotid stenosis and having atrial fibrillation were each found to be associated with a higher hazard of the composite outcome recurrence or death and its components recurrence and death without recurrence. It can also be seen from Table 3.2 that there was a higher hazard of the composite outcome associated with not being independent before stroke, not being able to walk independently, not having orientated speech, having cortical syndrome, having the shortest delay in assessment and having a haemorrhage on scan. However, it could be seen from the competing risks analysis that these covariate levels were each all associated with a higher hazard of death without recurrence but were not found to have an association with recurrence (although having a haemorrhage on scan was associated with a higher hazard of recurrence at the 10% level). Therefore, these associations found with the composite outcome were primarily due to death without recurrence. There was also a higher hazard of the composite outcome associated with not being able to lift both arms off the bed and this too was primarily due to an increased hazard of death without recurrence. However, there was also a marginal association with the ability to lift both arms off the bed and recurrence but in the opposite direction. Importantly, this was only detectable from the competing risk analysis because associations with recurrence were masked in the composite outcome analysis, due to the dominance of death without recurrence.

In terms of recurrence specifically, advancing age, being female, having other occlusive vascular disease, having high symptomatic carotid stenosis and having atrial fibrillation were each found to be associated with a higher hazard of the

event. However, as previously mentioned above, each of these covariate levels were also associated with an increased hazard of the competing risk death without recurrence. Therefore, this was something to bear in mind as this could influence the predictions of recurrence, particularly as in many instances the effect of the covariate was more strongly associated with death without recurrence than recurrence.

Multivariable analysis was then carried out considering the variables together in a model, allowing the effect of each covariate to be adjusted by other relevant variables. Parsimony was used for ease of subsequent prediction. For the modelling of the composite outcome, likelihood-ratio tests using the 5% significance level were used to select variables. This was carried out using backward selection and then repeated using forward stepwise selection and then, for stability, the selections were checked to ensure they matched. A similar approach was not used for the competing risks cause-specific hazards modelling for the reasons raised in the section 2.7.2 of chapter 2. Instead, the cause-specific hazards modelling included the same variables that were included in the Fine and Gray subdistribution hazards approach presented in the following section of this chapter. The variable selection for the Fine and Gray model was based on the approach proposed by Kuk and Varadhan (2013). This provided a solution that was needed anyway to help aid the comparison of the two approaches in the next section. However, because the purpose of this section was to compare the effect of the covariates on each event, any covariate found to have a significant effect on any of the events was included in the models for each event. For each event, covariates found to be statistically significant at the 5% and 10% level, are shown in bold and italics, respectively. Consequently, any covariate without such highlighting was only included for comparison purposes. This variable selection approach produced a sensible result in the sense that it included all those covariates found to be significant at the 10% level for either of the two events in the corresponding unparsimonious multivariable models. Table 3.3 shows the final modelling results.

		Cox regression of recurrence or death		cause-specific hazard competing risks Cox modelling			
	Baseline	hazard ratio for recurrence or death without recurrence (95% CI)	p-value	hazard ratio for recurrence (95% CI)	p-value	hazard ratio for death without recurrence (95% CI)	p-value
age - centred on mean of 71		**1.04 (1.03, 1.05)**	**<0.001**	1.01 (0.99, 1.02)	0.240	**1.06 (1.05, 1.07)**	**<0.001**
male	Female	0.90 (0.74, 1.09)	0.274	**0.69 (0.50, 0.94)**	**0.021**	1.01 (0.79, 1.30)	0.909
independent before stroke	No	**0.70 (0.52, 0.94)**	**0.018**	1.22 (0.59, 2.51)	0.589	**0.66 (0.47, 0.91)**	**0.013**
previous other occlusive vascular disease	No	**1.30 (1.07, 1.59)**	**0.010**	**1.41 (1.02, 1.95)**	**0.037**	1.21 (0.93, 1.56)	0.154
lift both arms off bed	No	**0.55 (0.43, 0.70)**	**<0.001**	**2.37 (1.25, 4.49)**	**0.008**	**0.34 (0.26, 0.46)**	**<0.001**
orientated speech	No	**0.52 (0.41, 0.66)**	**<0.001**	0.79 (0.48, 1.28)	0.340	**0.43 (0.32, 0.58)**	**<0.001**
symptomatic carotid stenosis			**<0.001**		**0.004**		**<0.001**
70-100% ICA	<70% ICA	**1.50 (1.13, 2.00)**	**0.005**	**2.06 (1.38, 3.07)**	**<0.001**	1.26 (0.84, 1.89)	0.271
Unknown	<70% ICA	**2.15 (1.66, 2.79)**	**<0.001**	**0.99 (0.57, 1.71)**	**0.970**	**2.85 (2.08, 3.89)**	**<0.001**
in atrial fibrillation	No	**1.43 (1.15, 1.77)**	**0.001**	**1.67 (1.16, 2.40)**	**0.006**	*1.31 (1.00, 1.72)*	*0.052*
haemorrhage on scan	No	0.91 (0.64, 1.29)	0.603	**1.94 (1.05, 3.58)**	**0.036**	*0.68 (0.45, 1.04)*	*0.075*

Table 3.3 Multivariable Cox composite vs cause-specific hazard Cox modelling results

It can be seen from Table 3.3 that the ability to lift both arms off the bed was associated with the composite outcome and each of the two competing risks. However, while being able to lift both arms off the bed was associated with a reduced hazard of the composite outcome and of death without recurrence, it was associated with an increased hazard of recurrence. Therefore, the increased hazard of recurrence was undetectable from the composite outcome analysis alone due to the dominance of deaths without recurrence. Hence, extra insight was gained from the analysis of the competing risks. It has helped to reveal that those who were able to lift both arms off the bed were more likely to have a recurrence rather than die before doing so. It has hence identified a group of patients who may benefit from prevention strategies for recurrence.

Symptomatic carotid stenosis was also associated with the composite outcome and each of the two competing risks. Having 70-100% ICA stenosis, compared to <70% ICA, was associated with an increased hazard of the composite outcome and recurrence. Having unknown stenosis, compared to <70% ICA, was associated

with an increased hazard of the composite outcome and death without recurrence. Having atrial fibrillation was associated with an increased hazard of the composite outcome and recurrence, and such an association with death without recurrence only just failed to reach the 5% significance level. Advancing age, not being independent before stroke and not having orientated speech were found to be associated with an increased hazard of the composite outcome and death without recurrence. Having other occlusive vascular disease was also found to be associated with an increased hazard of the composite outcome and recurrence. Therefore, the decomposition of the composite outcome into the two competing risks highlighted which of the particular events had an association with the covariate. Consequently, the competing risks analyses provided added benefit over the analysis of the composite outcome alone.

Being female was only found to be associated with an increased hazard of recurrence, with sex having a very negligible effect on death without recurrence. Therefore, this was something that was not evident from the analysis of the composite outcome alone. Having a haemorrhage on scan was found to have an association with an increased hazard of recurrence and a (marginally) reduced hazard of death without recurrence. This resulted in a lack of association with the composite outcome, with the opposing effects on each of the competing risks effectively cancelling each other out. Therefore, another added benefit of the analysis of each of the two competing risks was demonstrated. It unmasked associations not evident from the composite outcome analysis, highlighting information could be lost if a composite outcome is comprised of two events on which there are opposing effects.

With regards to recurrence specifically, being female, having other occlusive vascular disease, being able to lift both arms off the bed, having high carotid stenosis, being in atrial fibrillation and having a haemorrhage on scan were associated with an increased hazard of recurrence. As mentioned above, the ability to lift both arms off the bed and symptomatic carotid stenosis were also associated with death without recurrence, although the effects were in the opposite direction. This emphasises again the analysis of recurrence, separate from that for death without recurrence and the composite outcome, provided

insights that were not evident from the other two analyses. It identified characteristics in patients that could help clinicians deciding on a strategy to improve the lives of the patients under their care.

3.3 Cause-specific hazards vs Fine and Gray subdistribution modelling approaches

In section 2.5 of chapter 2, it was emphasised that a greater understanding of a competing risks scenario can be gained when the hazards and cumulative incidence functions for each of the competing risks are investigated. In particular, Latouche et al. (2013) recommend considering both cumulative hazard plots and cumulative incidence plots to assess the effects of covariates. This section aims to demonstrate the extra insight that such a comprehensive analysis can provide to help interpretation in the presence of competing risks. This would appear to be the first study to demonstrate the potential of such analysis of recurrence and death without recurrence in stroke patients.

The section begins by showing, in Table 3.4 for each of the competing risks, the effects of the covariates on the cause-specific hazard alongside those on the subdistribution hazard. As indicated previously in chapter 2, the subdistribution hazard has a one-to-one correspondence with the cumulative incidence.

Broadly speaking, for each event, the two approaches produced similar results in terms of the covariate effects being in the same direction. The rest of this section will discuss the effect of the covariates on the hazard of each event and how this translated into the corresponding effect on the subdistribution hazard/cumulative incidence of each event. Covariates that had similar effects on the hazard in terms of direction are discussed together.

The section also includes Figures 3.1 – 3.5 that show, for each covariate in turn, cumulative hazard and cumulative incidence plots. Each of the Figures are based on the following reference patient: male, aged 71, who was independent before stroke, had previous other occlusive vascular disease, could lift both arms off

	Baseline	hazard ratio for recurrence (95% CI)	p-value	subdistribution hazard ratio for recurrence (95% CI)	p-value
age - centred on mean of 71		1.01 (0.99, 1.02)	0.240	1.00 (0.99, 1.02)	0.600
male	Female	**0.69 (0.50, 0.94)**	**0.021**	**0.68 (0.50, 0.93)**	**0.016**
independent before stroke	No	1.22 (0.59, 2.51)	0.589	1.43 (0.69, 2.96)	0.340
previous other occlusive vascular disease	No	**1.41 (1.02, 1.95)**	**0.037**	*1.35 (0.97, 1.89)*	*0.078*
lift both arms off bed	No	**2.37 (1.25, 4.49)**	**0.008**	**3.55 (1.71, 7.37)**	**0.001**
orientated speech	No	0.79 (0.48, 1.28)	0.340	0.94 (0.53, 1.65)	0.820
symptomatic carotid stenosis			**0.004**		
70-100% ICA	<70% ICA	**2.06 (1.38, 3.07)**	**<0.001**	**2.10 (1.40, 3.15)**	**<0.001**
Unknown	<70% ICA	**0.99 (0.57, 1.71)**	**0.970**	**0.81 (0.46, 1.42)**	**0.450**
in atrial fibrillation	No	**1.67 (1.16, 2.40)**	**0.006**	**1.62 (1.11, 2.36)**	**0.012**
haemorrhage on scan	No	**1.94 (1.05, 3.58)**	**0.036**	**2.13 (1.15, 3.95)**	**0.016**

	Baseline	hazard ratio for death without recurrence (95% CI)	p-value	subdistribution hazard ratio for death without recurrence (95% CI)	p-value
age - centred on mean of 71		**1.06 (1.05, 1.07)**	**<0.001**	**1.06 (1.04, 1.07)**	**<0.001**
male	Female	1.01 (0.79, 1.30)	0.909	1.04 (0.80, 1.36)	0.750
independent before stroke	No	**0.66 (0.47, 0.91)**	**0.013**	**0.67 (0.46, 0.98)**	**0.039**
previous other occlusive vascular disease	No	1.21 (0.93, 1.56)	0.154	1.16 (0.89, 1.53)	0.270
lift both arms off bed	No	**0.34 (0.26, 0.46)**	**<0.001**	**0.31 (0.23, 0.41)**	**<0.001**
orientated speech	No	**0.43 (0.32, 0.58)**	**<0.001**	**0.45 (0.33, 0.60)**	**<0.001**
symptomatic carotid stenosis			**<0.001**		
70-100% ICA	<70% ICA	**1.26 (0.84, 1.89)**	**0.271**	**1.10 (0.72, 1.67)**	**0.680**
Unknown	<70% ICA	**2.85 (2.08, 3.89)**	**<0.001**	**2.81 (2.07, 3.81)**	**<0.001**
in atrial fibrillation	No	*1.31 (1.00, 1.72)*	*0.052*	1.19 (0.89, 1.59)	0.230
haemorrhage on scan	No	*0.68 (0.45, 1.04)*	*0.075*	**0.62 (0.39, 1.00)**	**0.048**

Table 3.4 cause-specific hazards vs Fine and Gray's subdistribution hazards multivariable modelling

the bed, did not have orientated speech, had <70% ICA stenosis, did not have atrial fibrillation and did not have a haemorrhage on scan. For each covariate in turn, every level of that covariate is shown with the other covariates held at the reference level specified.

The cumulative hazard plots are based on the Cox cause-specific hazard modelling. The cumulative incidence curves are based on Fine and Gray's subdistribution hazard model. The latter were calculated using the default weighted equation approach in `cmprsk` from `R`.

A crucial point regarding the results presented in Table 3.4 is that the effects of the covariates are displayed as (subdistribution) hazard ratios, and as such are relative effects. In contrast, the results presented in Figures 3.1 – 3.5 show absolute difference in cumulative incidence between the levels of a given covariate. When the underlying prevalence of an event is low, then even a large relative effect will not impact greatly on the difference in cumulative incidences displayed on the absolute scale. This is particularly pertinent in this case study because the frequency of deaths without recurrence greatly exceeded the frequency of recurrences.

3.3.1 covariates that affect the hazard of each event but in opposite directions

It can be seen that the ability to lift both arms off the bed, having a haemorrhage on scan and being independent before stroke were each associated with a higher hazard of recurrence and a reduced hazard of death without recurrence (Table 3.4). Alternatively, those without those characteristics were at a reduced hazard of recurrence and an increased hazard of death without recurrence. There were also corresponding effects on the cumulative incidences (Figure 3.1 and Figure 3.2).

For each competing risk, the effect on the subdistribution hazard was more pronounced than the effect on the hazard (lift arms off bed and haemorrhage on scan). This would normally suggest that the hazard of the other event may be having an influence on the cumulative incidence of the event. However, when there are opposing effects on the hazards, as in this case, the influence of the competing hazard is not such an issue. This is because any influence of the competing hazard will not alter the conclusions from the effect from the hazard of the event of interest. For instance, being able to lift both arms off the bed was found to be associated with an increase in the hazard of recurrence and this contributed to a relative increase in the cumulative incidence of recurrence seen in Figure 3.1(b). That there was a decreased hazard of death without recurrence may mean there would be more available to be at risk of recurrence. However this would only result in an increase in the cumulative incidence of

recurrence and therefore not contrast with the conclusion from the effect on the hazard of recurrence.

In Figure 3.1(a) and (b) it can be seen, as expected from Table 3.4, that the cumulative hazard and incidence of recurrence was higher for those who could lift both arms off the bed compared to those who could not, and that for death without recurrence the opposite was true. However, the relative effects found in Table 3.4 were not always apparent in Figure 3.1(b) and 3.1(d) due to the dominance of the absolute effect of death without recurrence over the absolute effect of recurrence. For example, Figure 3.1(b) did not reflect that the relative effect of lift arms was more pronounced for the subdistribution/incidence of recurrence than it was for the subdistribution/incidence of death without recurrence. Nor did it reflect that the effect on the subdistribution/incidence of recurrence was more pronounced than the effect on the hazard of recurrence in relative terms. Therefore it is imperative to consider both the relative effects (Table 3.4) and the absolute effects (Figure 3.1 and Figure 3.2) to fully appreciate the effects of the covariates on both the hazard and incidence of each of the competing events. This is also true of the effects of covariates that follow.

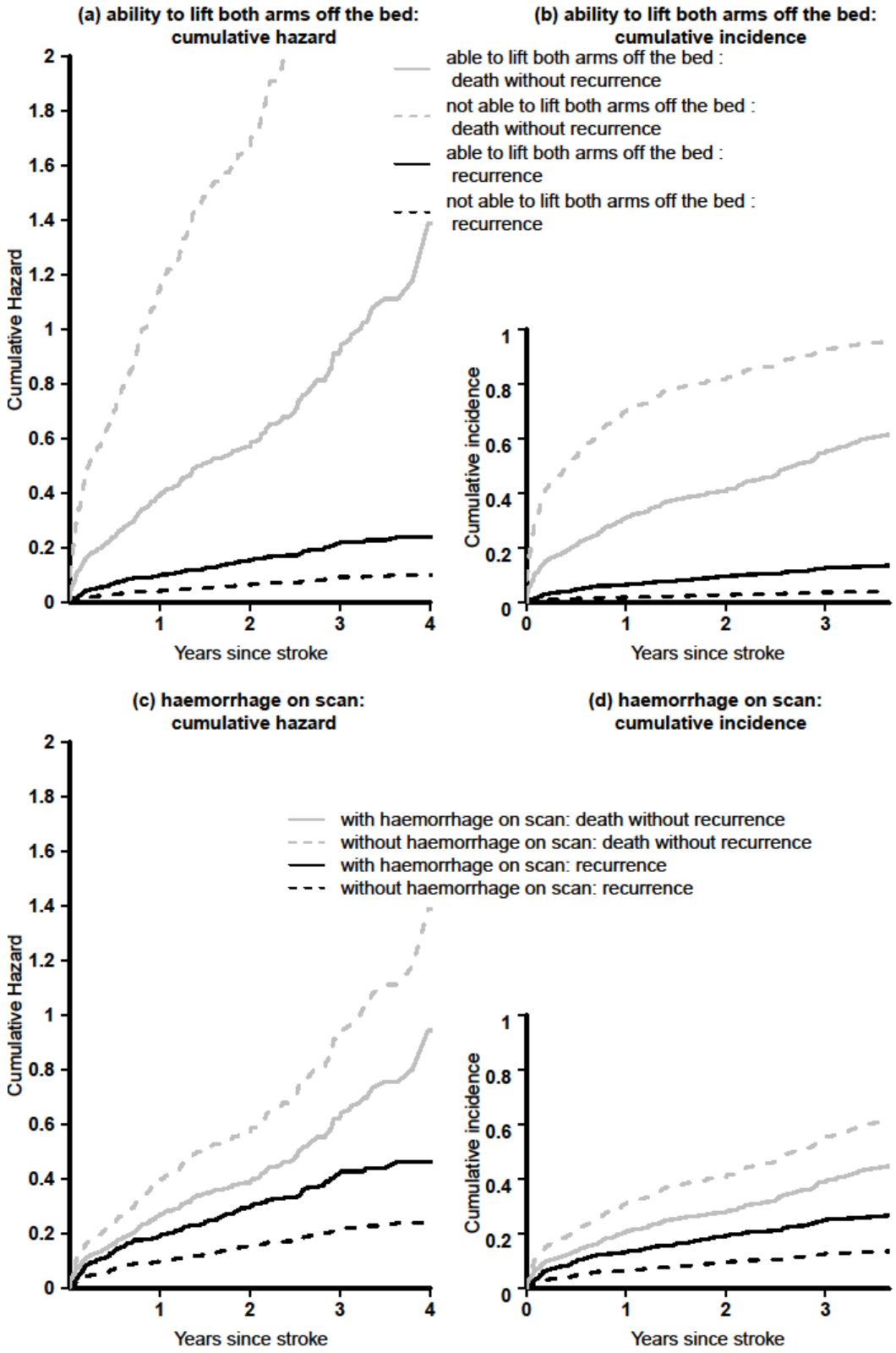


Figure 3.1 Cumulative hazard vs cumulative incidence: lift both arms and haemorrhage on scan

3.3.2 covariates that only affected the hazard of one event

It could be seen that age had no effect on the hazard of recurrence but that advancing age increased the hazard of death without recurrence (Table 3.4). The corresponding effects on the cumulative incidences were consistent with this (Figure 3.2(d)). Therefore, in particular, the increased hazard of the competing event death without recurrence was not strong enough to materially affect the cumulative incidence of recurrence. Specifically, the heightened hazard of death before recurrence could occur, as age increased, did not have any discernible reduction on the risk of recurrence, as age increased, even though there would be less patients alive to experience it.

It could also be seen that sex had no effect on the hazard of death without recurrence but being male decreased the hazard of recurrence (Table 3.4). Again, this was consistent with the corresponding effects on the cumulative incidences (Figure 3.3 (a) and (b)). This implies that the relative effect of being male on the cumulative incidence of recurrence was directly related to the hazard of recurrence without any indirect influence from the competing risk death without recurrence. Similarly, the cumulative incidence of death without recurrence was a direct effect which was not influenced by the effect of sex on recurrence.

Analogous conclusions could be drawn for those with unknown carotid stenosis compared to those with 70-100% ICA, except it was unknown stenosis that had no effect on the hazard of recurrence while it was associated with an increase in the hazard of death without recurrence (Figure 3.3 (c) and (d)).

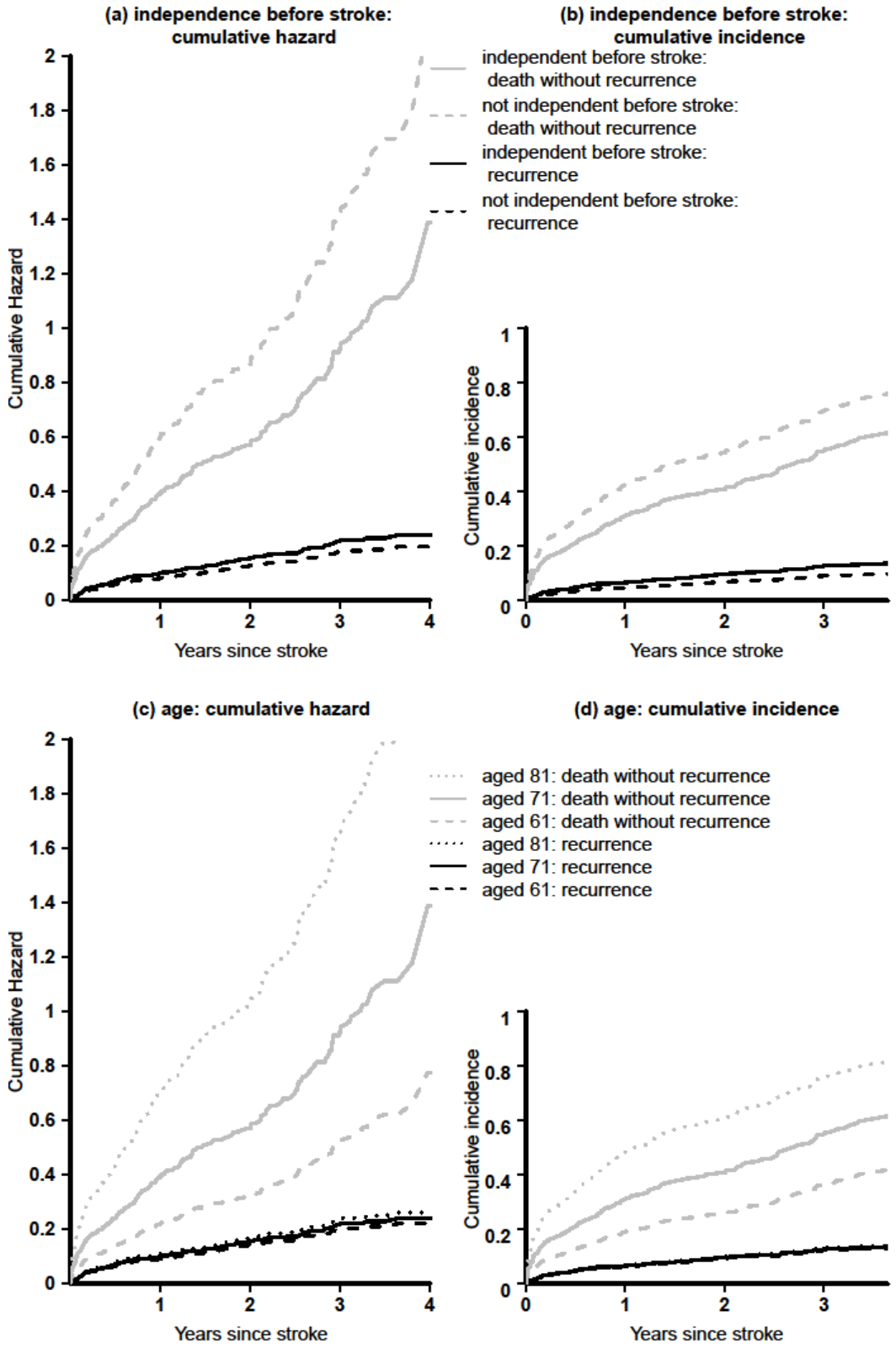


Figure 3.2 Cumulative hazard vs cumulative incidence: independence before stroke and age

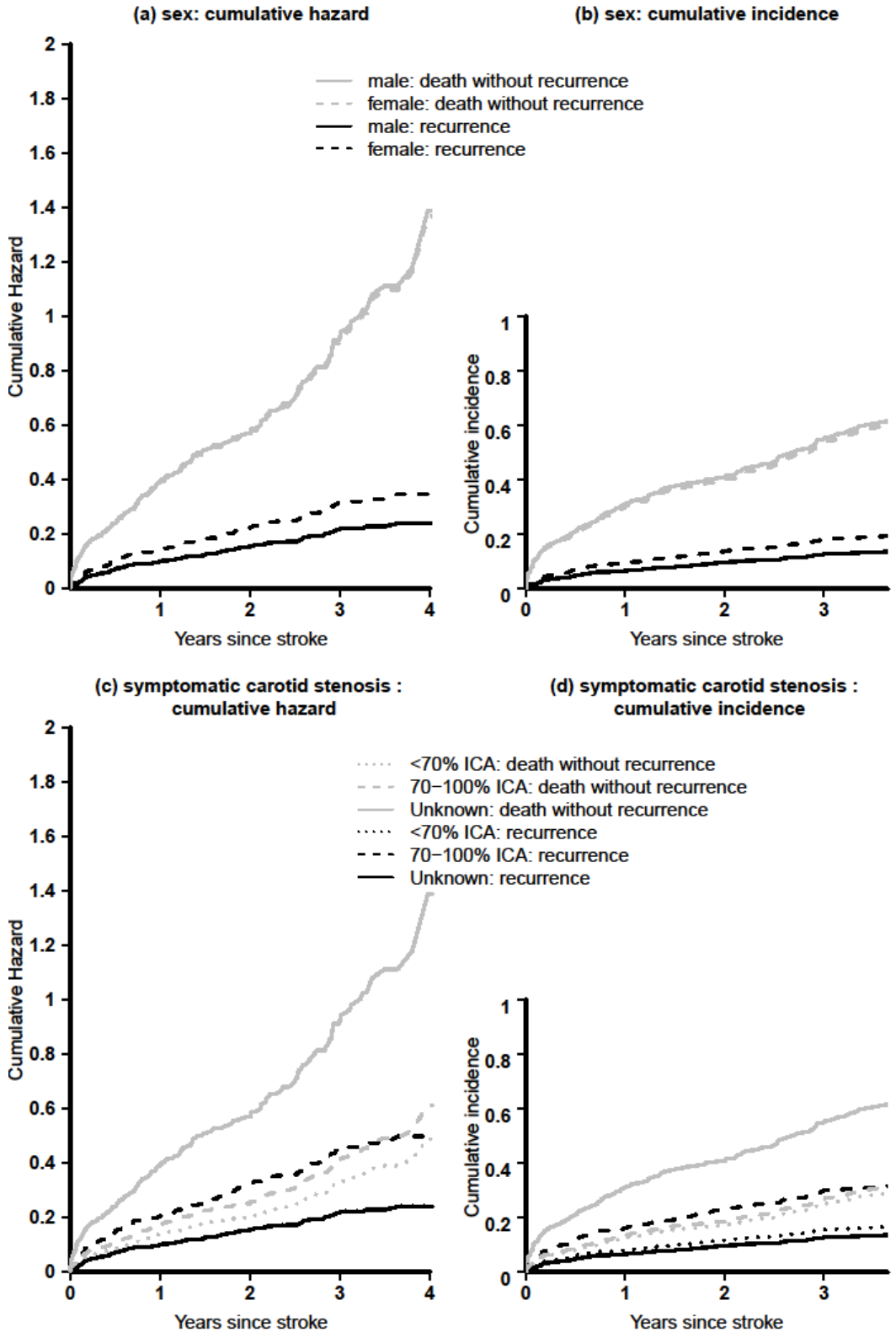


Figure 3.3 Cumulative hazard vs cumulative incidence: sex and stenosis

3.3.3 covariates that affected the hazards of each event in the same direction

It could be seen that having atrial fibrillation and other occlusive vascular disease each increased the hazard of both competing risks (Table 3.4 and Figure 3.4(a) and (c)). These effects were also seen in the cumulative incidences although the effects were slightly attenuated. This suggests that there may have been a degree of influence of the competing risk on each of the events. Specifically, the attenuation in the effect on the cumulative incidence of recurrence may have been due to the increased hazard of death without recurrence, meaning there were less at risk to be susceptible to recurrence. A similar argument may explain the attenuation in the cumulative incidence of death without recurrence.

Again, with reference to atrial fibrillation and other occlusive vascular disease, Table 3.4 shows that the relative effect in hazard/subdistribution hazard is actually higher for recurrence than for death without recurrence, but Figure 3.4 seems to suggest the opposite. For instance, for those with other occlusive vascular disease, the increase in relative hazard of recurrence of 41% only equated to a difference in cumulative incidences at 3 years of 3.1%; specifically 12.5% versus 9.4% respectively for those with and without other occlusive vascular disease. In contrast, the increase in relative hazard of death without recurrence of only 21% equated to a larger difference in cumulative incidence at 3 years of 5.3%; specifically 54.4% versus 49.1% respectively for those with and without other occlusive vascular disease. Therefore, it helps to use both Table 3.4 and the Figures to gauge the underlying prevalences that hazards are acting on to produce the cumulative incidences, especially when the occurrence of one event is so much higher than the other.

There was a relative reduction in hazard of each competing risk for those with orientated speech, with a more pronounced effect for death without recurrence (Table 3.4). However on the absolute scale, there was no effect of orientated speech on the cumulative incidence of recurrence, even though there was a trend towards a reduced hazard (Figure 3.5).

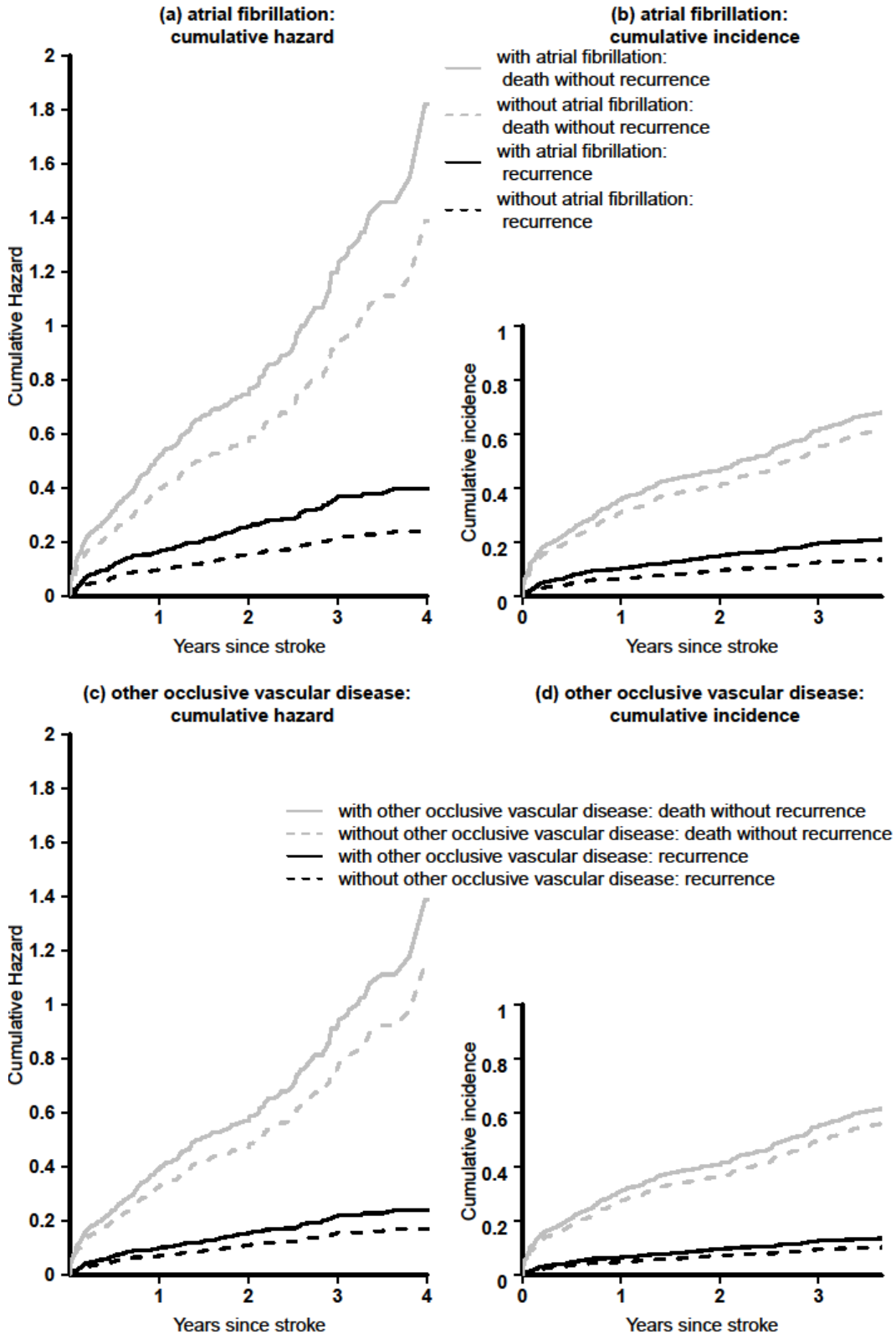


Figure 3.4 Cumulative hazard vs cumulative incidence: atrial fibrillation and other occlusive vascular disease

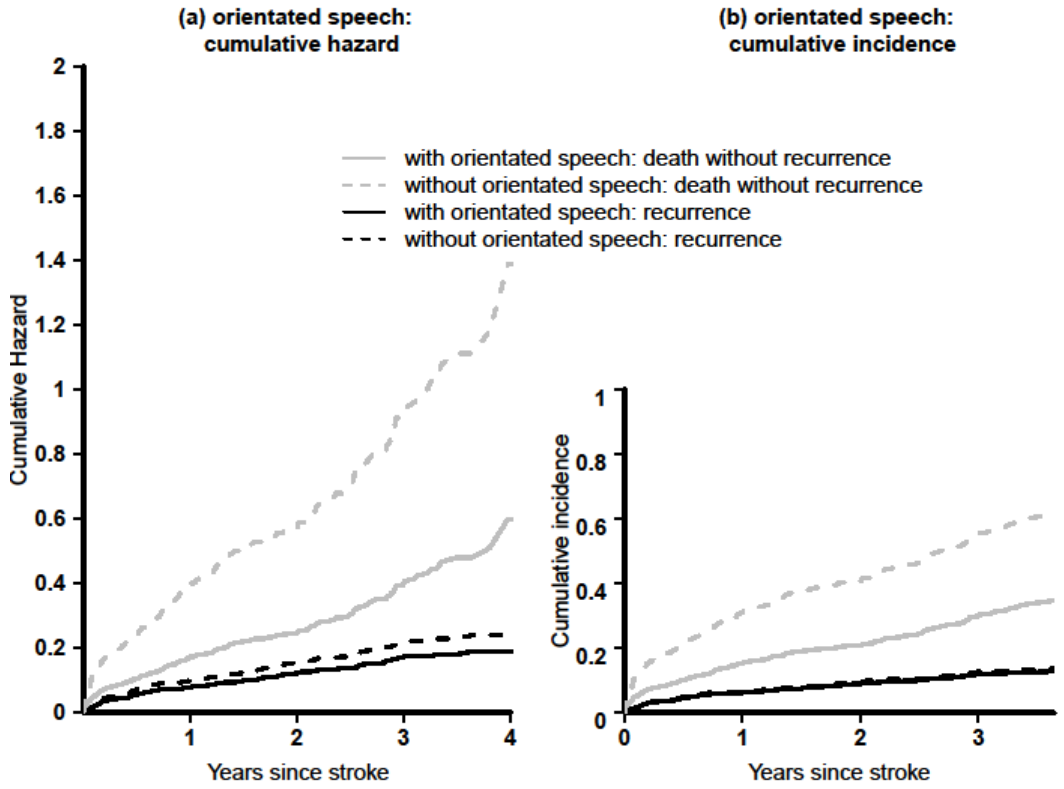


Figure 3.5 Cumulative hazard vs cumulative incidence: orientated speech

This may be explained by more of those with orientated speech having a recurrence than was perhaps expected, due to the decreased hazard of death without recurrence meaning there was more left to be at risk of recurrence. As such, the (non-significant) increase in the hazard of recurrence was masked from the cumulative incidence of recurrence, due to the more pronounced decrease in the hazard of death without recurrence.

Finally, there was a relative increase in the hazard of each competing risk for those with 70-100% ICA symptomatic carotid stenosis compared to those for whom it was unknown/unassessable, with a more pronounced effect for recurrence (Table 3.4). Even though there was a (trend towards a) relative increase in the hazard of death without recurrence, on the absolute scale, there was no effect on the corresponding cumulative incidence (Figure 3.3 (d)). An inspection of the plots helps interpret this finding. It can be seen in Figure 3.3 (c) and (d), in those with 70-100% symptomatic carotid stenosis, that the cumulative hazard and incidence of recurrence was higher than for death

without recurrence. The stronger increase in the hazard of recurrence could explain why the cumulative incidence of death without recurrence for those with 70-100% ICA stenosis rose more gradually, and therefore was more in line with that for <70% ICA stenosis.

3.4 Consequences of not taking into account competing risks

This section demonstrates some of the bias that can be introduced if an inappropriate approach that does not take into account competing risks is used. Two different illustrations are used. Firstly, the non-parametric naïve Kaplan-Meier approach is compared to the competing risks cumulative incidence approach. Secondly, the estimates are based on modelling that considers the effects of covariates. The naïve cause-specific hazard Cox approach to modelling is compared to the more appropriate calculation of the cumulative incidence, which is based on combining the cause-specific hazard of each of the competing risks. The effects of selected covariates on death without recurrence are used for illustration.

3.4.1 Non-parametric naïve Kaplan-Meier versus competing risks cumulative incidence

Figure 3.6 on the next page shows, for each of the competing risks, the non-parametric naive Kaplan-Meier estimate of cumulative incidence, alongside that of the more appropriate estimate that takes into account competing risks.

It can be seen in Figure 3.6 that, for death without recurrence (dashed lines), bias starts to be introduced from about 1 year onwards. The bias results in progressively more inflation of the cumulative incidence. By 3.5 years the biased cumulative incidence estimate is 29.2%, compared to the more appropriate estimate of 26.4%. Therefore, inflation of around 11% has resulted. With regards to recurrence (solid lines), the bias begins from as early as 6 months. By 4 years the biased estimate is 20.2%, compared to the more appropriate estimate of 17.1%, an inflation of 18%.

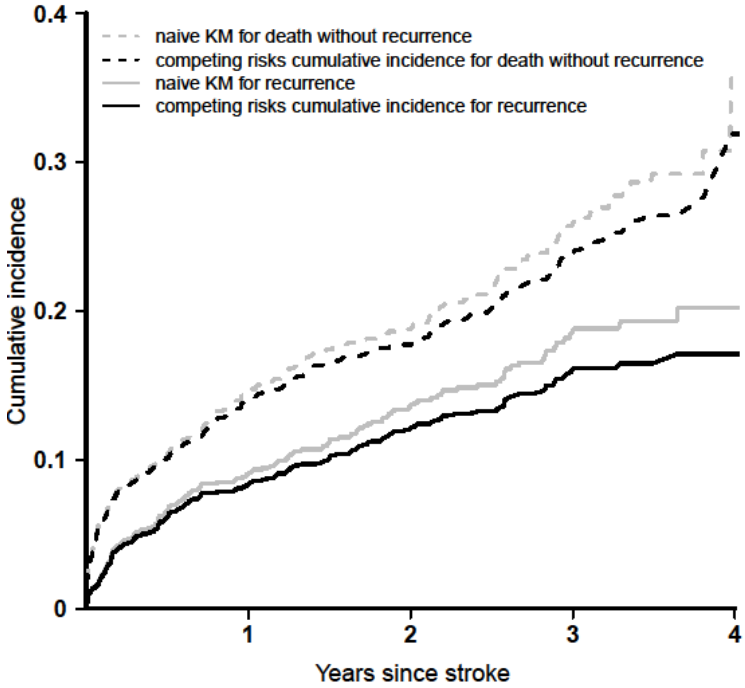


Figure 3.6 non-parametric naïve KM vs competing risks cumulative incidence

3.4.2 Naïve cause-specific hazard-based cumulative incidence versus that derived appropriately using the cause-specific hazards of each of the competing risks

Figure 3.7 on the next page shows the effect of four different covariates in turn on the cumulative incidence of death without recurrence. Cumulative incidences based on the naïve cause-specific hazard approach and those based on the appropriate derivation using all cause-specific hazards are shown for each level of the covariate. Each plot shows the outcomes for the reference patient described in section 3.3 on page 63, alongside that when the level of the specific covariate is changed.

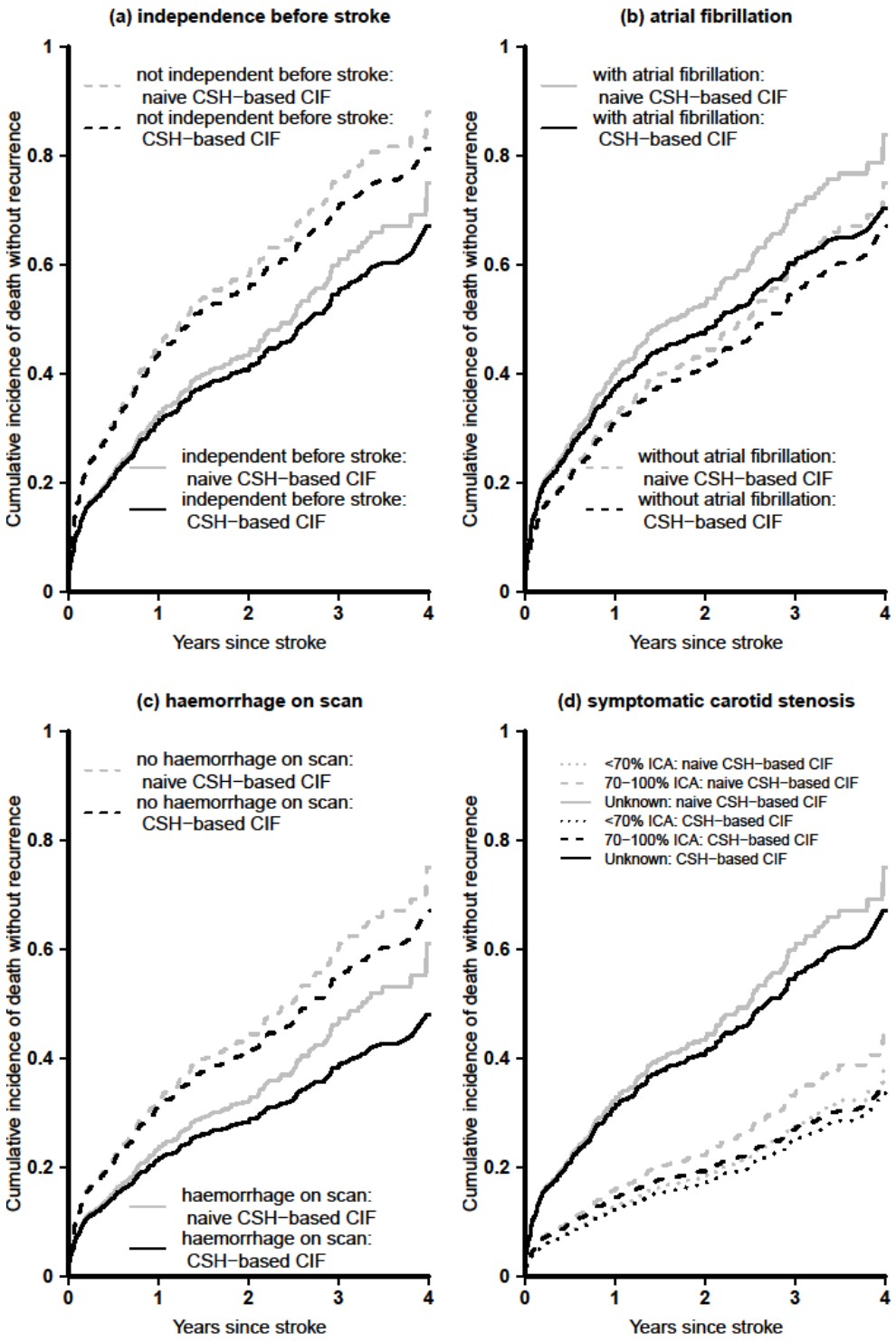


Figure 3.7 Cumulative incidences derived using the naïve cause-specific hazard approach versus those using the more appropriate cause-specific hazard approach. Abbreviations: CSH=cause-specific hazard, CIF=cumulative incidence function

The reference patient had a biased 3.5 year cumulative incidence of 67.1% and a more appropriate cumulative incidence of 60.3%, an inflation bias of 11%. Each

of the changed covariate levels will now by commented on in turn. Figure 3.7(a) shows, that for not independent before stroke, the biased cumulative incidence was 81.6% by 3.5 years compared to the corresponding more appropriate estimate of 75.6%. Therefore not taking into account the competing risk inflated the estimate by 8%. The corresponding information for those with atrial fibrillation (Figure 3.7(b)) was a biased cumulative incidence of 76.7% and an appropriate cumulative incidence of 65.1% resulting in an inflation of 18%. Furthermore, the three corresponding percentages of interest were 53.1%, 42.7% and 24% for having a haemorrhage on scan (Figure 3.7(c)). Figure 3.7(d) demonstrates the effect of not using an appropriate competing risks method with regards to symptomatic carotid stenosis. It can be seen in Figure 3.7(d) that the most dramatic bias was introduced for 70-100% ICA stenosis. By 3.5 years, the biased and more appropriate cumulative incidences were 38.8% and 30.4% respectively, inflation of 28%. There was also prominent bias by 3.5 years with <70% ICA stenosis, with biased and more appropriate cumulative incidences of 32.3% and 28.6% respectively, inflation of 13%.

3.5 Discussion/summary of results

This chapter presented competing risks analyses of the Edinburgh Stroke Study (ESS), with ESS described in section 3.1. Each of the remaining sections presented various different aspects of competing risks analysis to illustrate the insights that can be gained over standard survival analysis. Section 3.2 demonstrated the extra insight that can be gained from decomposing a composite outcome into its component competing risks. In particular, it was found that being able to lift both arms off the bed had opposing effects on the two competing risks. Those able to lift their arms off the bed were found to be at a reduced hazard of death without recurrence but at an increased hazard of recurrence. It may seem counterintuitive at first that the positive outlook of being able to lift arms could result in the negative outcome of recurrence. However, when considering this in the context of the competing risk death without recurrence, it becomes clear that recurrence, while still negative, is the least negative outcome. Those able to lift their arms off the bed were more susceptible to recurrence because they were more likely not to die before

having a recurrence. Not being able to lift both arms off the bed was found to be associated with the worse prognosis of death before even experiencing a recurrence. Importantly, the increased hazard of recurrence found to be associated with being able to lift both arms off the bed was undetectable from the analysis of the composite outcome alone, due to the dominance of death without recurrence. Therefore, the insightfulness of the competing risks analysis was demonstrated. It identified a group of patients who might benefit from targeted strategies to prevent recurrence.

Having a haemorrhage was also found to have opposing effects on the two competing risks. Specifically, having a haemorrhage was associated with a higher hazard of recurrence, and not having one indicated a reduced hazard of death without recurrence, although the latter failed to reach significance at the 5% level. The analysis of the composite outcome did not find any evidence of an effect and therefore masked both of these effects on the competing risks. A lack of effect resulted because the opposing effects on each of the two competing risks were cancelling each other out. Therefore, this again demonstrated the added benefit of the competing risks analysis over that of the composite outcome alone.

The effect of sex on the hazard of recurrence also highlighted the value of the competing risks analysis. The analysis of the composite outcome did not find any evidence of an effect of sex, and therefore masked that being female was associated with a higher hazard of recurrence.

Section 3.3 demonstrated that a greater understanding of a competing risks scenario can be gained by considering both the hazard and cumulative incidence of each competing risk. In particular, it illustrated the effects of covariates on the (subdistribution) hazard ratios of each competing risk alongside presenting cumulative hazard and cumulative incidence plots. Furthermore, it emphasised how these aided interpretation of the effects found in the competing risks scenario. It showed that the effect of a covariate on the cause-specific hazard of an outcome may not necessarily translate into the same effect on the cumulative incidence. It would appear to be the first time such a comprehensive

competing risks analysis had been carried out for stroke recurrence and death without recurrence.

For example, the cumulative hazard plot for orientated speech showed that the hazard of recurrence was lower for those with orientated speech, but the cumulative incidence plot showed no effect of orientated speech. This could be explained by the fact that those with orientated speech also had a lower hazard of death without recurrence, but that this reduced hazard was more pronounced than that for recurrence. This meant that there were more of those with orientated speech having a recurrence than was perhaps expected, due to there being more of them left to be at risk of recurrence owing to the decreased hazard of death without recurrence. Therefore, the cumulative incidence of recurrence for this group was in line with that for those without orientated speech, instead of being lower than it.

In another example, the cumulative hazard plot for symptomatic carotid stenosis showed there to be no difference between the hazard of recurrence between those with <70% ICA and unknown/unassessable stenosis. However the corresponding cumulative incidence plot showed that the risk of recurrence was lower in those with unknown/unassessable stenosis. This could be explained by the increased hazard of death without recurrence for those with unknown/unassessable stenosis compared to <70% ICA.

In another similar example, the cumulative hazard plot of symptomatic carotid stenosis showed that the hazard of death without recurrence was higher for those with 70-100% ICA than those with <70% ICA. However the corresponding cumulative incidence plot showed very little difference between the two groups. This could be explained by there being an even more pronounced effect of an increased hazard for recurrence than that for death without recurrence. This meant there were less of those with 70-100% ICA available to die without recurrence and therefore the corresponding cumulative incidence for this group was lower than expected and more in line with that for <70% ICA.

In this research, none of the cumulative incidence curves for the two competing risks for a particular covariate level crossed as more time passed. This might

have been seen, however, in a study where the occurrence of the competing risks was more evenly balanced. It would also be more likely in such a situation if the competing risks had a stronger competing effect on each other or the study had a longer follow-up period. Furthermore, because the prevalence of death without recurrence was so much higher than that for recurrence, it was especially beneficial to consider both the relative effects (hazards) and absolute effects (cumulative incidences) for each competing risk.

The chapter then proceeded to illustrate the consequences of not taking into account competing risks in section 3.4. To do this, naïve Kaplan-Meier/Cox predictions were compared to those that appropriately took competing risks into account. For instance, it was shown that inflation of 18% arose in the estimate of the 4-year cumulative incidence of recurrence when using the biased Kaplan-Meier method compared to the more appropriate competing risks approach (sub-section 3.4.1). Furthermore, when the effects of covariates were taken into consideration, there was inflation in the estimate of the 3.5-year cumulative incidence of death without recurrence of 11% for the reference patient (sub-section 3.4.2). In addition, inflation bias of 8%, 17% and 24% were evident when the values of independent before stroke, atrial fibrillation and haemorrhage on scan respectively were changed from that of the reference patient. Also, when the level of symptomatic carotid stenosis was changed to <70% and 70-100% ICA respectively, inflation biases of 27% and 13% were introduced. Therefore, it is imperative to use an approach that appropriately takes into account competing risks to avoid introducing such bias.

The research in this study found that being female, having previous other occlusive vascular disease, being able to lift both arms off the bed, having 70-100% ICA symptomatic carotid disease, having atrial fibrillation and having a haemorrhage on scan were each associated with a higher hazard of recurrence. It was also found that the above characteristics, with the exception of previous other occlusive vascular disease which was only significant at the 10% level, were associated with a higher *cumulative incidence* of recurrence.

However, the finding that females were more likely to have a recurrence is in contrast to the results from some of the studies summarised in Appendix III, the evidence table discussed in sub-section 2.10.2. While Rutten-Jacobs et al. (2013) (Netherlands) and Castilloux et al. (2015) (Canada) did not find any effect of gender on recurrence, Sun et al. (2013) (Singapore), Andersen et al. (2011) (Denmark) and Lewsey et al. (2010) (Scotland) did find being male to be associated with an increased risk of recurrence. This contrast with the study by Lewsey et al. (2010) in particular is worth exploring further because both studies used similar Scottish populations. In this current research, patients were followed-up from their index stroke and a limitation of this was that it was not possible to determine whether the stroke was a first-ever or a recurrent stroke. Therefore, for patients with recurrences, it was not known whether this was their first such event or a subsequent one. Time to recurrence and time to death are both likely to be very different between patients with a first-ever stroke and those who have already had a recurrence. However, the finding in the current research with regards to previous other occlusive vascular disease was similar to the results found by Castilloux et al. (2015) who found past medical conditions to be associated with a higher hazard of recurrence. In addition, the research in this chapter confirmed the results for stroke severity found by Andersen et al. (2011) and Rutten-Jacobs et al. (2013) and those for atrial fibrillation found by Lewsey et al. (2010).

In the case study used in this chapter, the incidence of death without recurrence far exceeded that of recurrence. In situations such as this when the number of events are not well balanced, or the strength of competition between the events differs, it is imperative to consider both relative and absolute effects because otherwise inferences from competing risks analyses can be problematic. For example, this research showed that, compared to those with <70% ICA symptomatic carotid stenosis, those with 70-100% ICA symptomatic carotid stenosis were 26% more likely to die without recurrence. However, there was only a marginal difference between the two groups in terms of the absolute probabilities of death without recurrence, with 3-year probabilities of 26.4% versus 24.4%. Therefore, looking at either the relative or absolute effect in

isolation could be misleading because it does not give the whole picture of the competing risk process.

On the other hand, it could be that the event that is not primarily of interest is so rare as to not affect the probability of an event of interest occurring. This would not constitute a competing risk process and therefore the rare event would not need to be included in the modelling.

However, if the event of interest is rare then an alternative to the competing risk analysis in this chapter would need to be applied. This is because the competing risks analysis is based on studying the course of the disease by prospectively following up a cohort to see which subsequent outcomes they experience. In common with other investigations of rare outcomes, there are other study designs which are more efficient such as retrospective-based designs.

Chapter 4 Background to multi-state modelling

4.1 Introduction

The competing risks survival analysis framework was the focus of the previous two chapters. Chapter 2 provided the background to the approach, highlighted methodological issues that need to be considered and reviewed the literature into the extent of its use. Motivated by this, Chapter 3 presented empirical competing risks analyses with a stroke case study, an area where the technique has not been used to its full potential. Competing risks analyses are concerned with investigating which of two or more mutually exclusive events occur first and in that sense the risks are competing with one another. However, with the exception of when all the competing risks are specific causes of death, interest can be in the subsequent events experienced. Competing risks analysis is a simple form of modelling under the more general multi-state modelling survival analysis framework. This chapter focuses on multi-state modelling, which extends the competing risks approach to also consider events of interest that can happen after the first event.

Multi-state modelling is of particular use when studying the course of a chronic and/or progressive disease. It allows each of the relevant stages (health states) across the disease pathway to be investigated. Therefore factors associated with an initial health state, a final health state (normally death) and all relevant states in between can be assessed. In particular, transitions involving different pathways can be compared. For example, in the stroke case study used in the previous chapter, the hazard of death in stroke patients could be compared between those who did and did not experience a recurrence. In addition, one of the main advantages of multi-state modelling over standard and competing risks survival analysis is the flexibility in predictions it can accommodate. Predictions of being in different health states over time can be estimated as patients enter the initial health state. Furthermore, dynamic predictions can be calculated, that is to say predictions as time progresses, especially as patients progress to other health states. Dynamic predictions have particular potential in an epidemiological/medical statistics context, but as yet that potential does not seem to be fully realised. In a health economics context, multi-state modelling

can be seen as an alternative form of state-transition modelling; a common approach in health economic modelling. Multi-state modelling has much to offer in that field, however as section 6.3 of chapter 6 will show it is not widely-applied in that area. The continuous-time multi-state modelling framework is the specific focus of this current chapter. Discrete time decision-analytic state-transition modelling is described in section 6.2 of chapter 6.

This current chapter describes the multi-state modelling method, providing the background to the remaining empirical chapters of this thesis. Section 4.2 introduces some important concepts relating to multi-state modelling, which is needed to fully understand the rest of the sections in this chapter. In particular, section 4.2 describes the illness-death model, a specific type of multi-state modelling that is later demonstrated empirically in Chapter 5 (stroke epidemiology) and Chapter 7 (health economics case study). Then, section 4.3 provides an overview of Markov multi-state models. These models assume the Markov "memoryless" property. This is the condition that the next state to be visited, and the time that that occurs, only depends on the present state and not on any of the previous states visited or the time spent in previous states. Non-parametric, semi-parametric and parametric approaches to fitting Markov multi-state models are outlined separately. Next, section 4.4 describes semi-Markov multi-state models. These models relax the Markov property by assuming the process depends not only on the present state but also on the time since entry into the present state. The focus of section 4.5 is the flexibility in predictions that multi-state modelling can accommodate. Specifically, overviews are given of predictions at the start of the study for different health states and dynamic predictions that update prognosis as time progresses, and especially as patients progress to other health states.

Contributions from the literature are used in the aforementioned sections of this chapter to help describe the multi-state modelling framework. Because these sections are discussing the methodological aspects of the approach, a robust search strategy was not developed to review all such contributions. However, for the final two sections of this chapter, section 4.6 and 4.7, search strategies were developed in order to review the extent of use of multi-state modelling.

Section 4.6 explores the use of multi-state modelling in a broad sense without concentrating on any particular area of medicine. Finally, section 4.7 discusses applied uses in the medical literature, with a particular focus on stroke epidemiology. This identifies gaps in existing research and motivates the empirical analysis of the stroke case study in the next chapter.

4.2 Preliminary concepts

The purpose of this section is to introduce important concepts related to multi-state modelling necessary to fully understand the rest of the sections in the chapter. There has been a rapid emergence of literature in the medical statistics/general medical fields describing multi-state modelling in the last 10-15 years. Contributions that provide succinct and useful introductions to multi-state modelling include those by Putter et al. (2007) Commenges (1999), Hougaard (1999), Andersen et al. (2002), Andersen and Keiding (2002), Andersen and Perme (2008), Meira-Machado et al. (2009). Much of this chapter is largely based on these contributions. Multi-state modelling is also being increasingly recognised in the health economics field. Section 6.3 of chapter 6 provides a review in this area.

Firstly, the illness-death model is introduced. The illness-death model, also known as the disability model, is one of the most straightforward and widely used models in multi-state modelling. The states and transitions involved in this model can be seen in Figure 4.1.

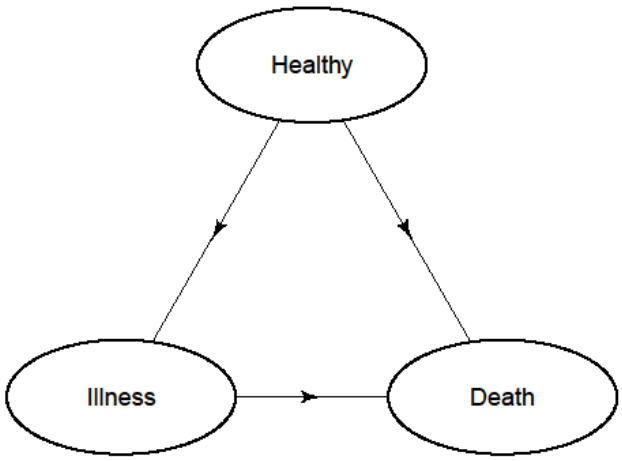


Figure 4.1 State transition diagram for an illness-death model

There is an initial healthy state, an intermediate illness state and an absorbing death state. The death state is absorbing in the sense that once it is reached, it is not possible to move from it. (Generally, any state is said to be absorbing if there is no interest in what happens after it is reached). The three transitions healthy → illness, healthy → death and illness → death are modelled. Therefore, patients in the healthy state can either move to the illness state, direct to the death state without entering the illness state or remain in the healthy state. Patients who reach the illness state can either move to the death state or remain in the illness state. One of the particular advantages of this modelling framework is that the risk of death with and without entering the intermediate state can be studied and compared.

The model depicted in Figure 4.1 is said to be an "irreversible" or "uni-directional" illness-death model. This is because once a patient reaches a new state, they cannot later revert to a previous one. However, it is also possible for an illness-death model to be "reversible" or "bi-directional" by adding a transition for illness → healthy. This is particularly applicable to conditions that have a cure/remission phase. This chapter will focus on irreversible illness-death models for ease of illustration. However all the concepts discussed can be generalised to more complex models with more states and transitions. It will also be assumed that the times of transition are known exactly. Considerations for interval-censored data (Commenges, 2002) with missing or misclassification of states (van den Hout et al., 2014) are described elsewhere. Topics closely related to multi-state modelling including frailty (Putter and van Houwelingen, 2015) and modelling of recurrences (Amorim and Cai, 2015) may also be of interest.

As described in Andersen and Perme (2008), the formal definition of a multi-state process is "a (continuous-time) stochastic process (X(t), t $\in \mathcal{T}$) with a finite state space $\mathcal{S}$= { 0, 1, …, p} with right-continuous sample paths: X(t+) = X(t)." In the process, time is defined on $\mathcal{T}$=[0, τ] or [0, τ) with τ ≤ + ∞. A multi-state process X(.) also generates a history $\mathcal{X}_t$ (an σ – algebra) which contains the history of the process in the interval [0, t]. This history consists of information relating to the previous states visited, and the time spent in previous states.

A fundamental concept in multi-state modelling is the transition probability. For a state h and a state j the probability of h → j, the transition from h to j, is defined as:

P_{hj} (s,t) = Prob (X(t) = j | X(s) = h, $\mathcal{X}_{s-}$) for h, j ∈ $\mathcal{S}$, s, t ∈ $\mathcal{T}$, s≤ t .

Transition probabilities can also be represented in a matrix. The transition probability matrix for the illness-death model is

$$\boldsymbol{P}(s,t) = \begin{pmatrix} p_{11}(s,t) & p_{12}(s,t) & p_{13}(s,t) \\ 0 & p_{22}(s,t) & p_{23}(s,t) \\ 0 & 0 & p_{33}(s,t) = 1 \end{pmatrix}$$

Because the entries are probabilities, each row of $\boldsymbol{P}(s,t)$ must sum to 1. $p_{33}(s,t)$ is always one because the third state death is absorbing.

Another related important concept is the transition intensity. This is defined as

$q_{hj}\,(t) = \lim_{\Delta t \to 0} \frac{P_{hj}(t\,, t+\Delta t)}{\Delta t}$ and as such is a derivative.

It is usually known as the instantaneous hazard (rate) of moving from state h to j in the small time interval (t, t + Δt).

The transition intensities are often expressed in a transition intensity (hazard) matrix. All relevant transitions in a model are represented by non-zero entries, with zero entries otherwise. For example, for the illness-death model described previously, the transition intensity matrix would be:

$$\boldsymbol{Q}(t) = \begin{pmatrix} q_{11}(t) & q_{12}(t) & q_{13}(t) \\ 0 & q_{22}(t) & q_{23}(t) \\ 0 & 0 & 0 \end{pmatrix}$$

The matrix has as many rows and columns as there are states i.e. three. The intensities (hazard rates) for transitions 1 → 2 (healthy → illness), 1 → 3

(healthy → death) and 2 → 3 (illness → death) are represented by $q_{12}(t)$, $q_{13}(t)$ and $q_{23}(t)$ respectively. Each of the rows in $\boldsymbol{Q}(t)$ must add up to zero. Therefore, $q_{11}(t) = -(q_{12}(t) + q_{13}(t))$ and $q_{22}(t) = -q_{23}(t)$. $q_{33}(t)$ = 0 because the third state death is absorbing. All other entries in $\boldsymbol{Q}(t)$ are zero because they do not represent a transition of interest in the model.

A further important and related quantity is the cumulative hazard for the transition h → j, which is defined as:

$$\mathrm{H}_{hj}(t) = \int_0^t q_{hj}(s)\ ds$$

These all lead to the state occupation probability which is the quantity most of interest in multi-state modelling. It involves combining the hazards and probabilities of relevant transitions and its estimation will be described in subsequent sections of this chapter. It is defined as

π_h(t) = Prob (X(t) = h) for h ∈ $\mathcal{S}$. The initial distribution, π_h(0), is defined as

π_h(0) = Prob (X(0) = h) for h ∈ $\mathcal{S}$ leading to π_h(t) = $\sum_{j\in\mathcal{S}}$ π_j(0) P_{jh}(0,t).

Another important consideration in multi-state modelling is the scale used to measure time. Two common approaches are “clock-forward” and “clock-reset”. With the clock-forward approach, time is measured from entry into the initial state, regardless of the state in the multi-state model. In contrast, with the clock-reset approach each time a patient enters a new state their clock is set back to zero. More details are available in the tutorial by Putter et al. (2007).

A related concept is the Markov property. The Markov property is the condition that the next state to be visited, and the time that that occurs, only depends on the present state, and not on any of the previous states visited or the time spent in previous states. Only clock-forward models can be Markov. The Markov property does not hold in clock-reset models because in that approach it is intrinsic that the time scale depends on the time spent in the previous state. However, if with the clock-reset approach, the sojourn time only depends on the

present state and the time since entry into that state, then the process is said to be semi-Markov. Markov and semi-Markov multi-state models are described in sections 4.3 and 4.4 respectively.

4.3 Markov multi-state models

This section describes estimation and prediction in Markov models. In Markov models, the Markov property described previously holds. Mathematically this is:

$$P_{hj}(s,t) = \text{Prob}\,(X(t) = j \mid X(s) = h, \mathcal{X}_{s-})$$

$$= \text{Prob}\,(X(t) = j \mid X(s) = h, X(t_{n-1})=i_{n-1}, \ldots, X(t_1)=i_1)$$

$$= \text{Prob}\,(X(t) = j \mid X(s) = h)$$

where $0 \le t_1 \le \ldots \le t_{n-1} \le s \le t \in \mathcal{T}$ is any non-decreasing sequence of n + 1 state occupation times and $i_1, \ldots, i_{n-1}, i, j \in \mathcal{S}$.

In other words, the state of the process at time t depends only on the most recent state occupied prior to time t.

Non – parametric, semi – parametric and parametric methods will now be described separately.

4.3.1 Non-parametric Markov multi-state models

This section describes estimation for non-parametric Markov multi-state models. The models are non-parametric in the sense there are no covariates to be modelled. The standard non-parametric estimator of the cumulative transition intensity (hazard) is the Nelson-Aalen estimator. This is defined, for the transition $h \rightarrow j$, as

$$\hat{A}_{hj}(t) = \sum_{t_i \leq t} \frac{dN_{hj}(t_i)}{Y_h(t_i)}$$

where t_i indicates the event times, $dN_{hj}(t_i)$ is the observed number of transitions from state h to state j at time t_i and $Y_h(t_i)$ is the number of subjects at risk of the h → j transition at time t_i .

At event times t_i , the Nelson-Aalen estimator, $\hat{A}_{hj}(t)$, makes jumps of magnitude $\Delta\hat{A}_{hj}(t_i)$.

The cumulative hazards can be expressed in a matrix $\boldsymbol{A}(t)$ with dimensions S × S, where S is the number of states. The Nelson-Aalen estimator can be used to estimate the off-diagonal entries of the matrix $A_{hj}(t_i)$ (h ≠ j). The diagonal entries $A_{hh}(t_i) = -\sum_{j \neq h} \hat{A}_{hj}(t)$. Therefore, each row of the cumulative hazard matrix $\boldsymbol{A}(t)$ sums to 0.

The transition probability matrix can also be estimated using the Aalen– Johansen estimator, which is a product integral approach.

The Aalen – Johansen estimator of the transition probabilities is defined as

$$\widehat{\boldsymbol{P}}(s,t) = \prod_{u \in (s,t]} \mathbf{I} + \Delta\, \widehat{\boldsymbol{A}}(u)$$

where u indicates the event time, $\mathbf{I}$ is the identity matrix and $\Delta\, \widehat{\boldsymbol{A}}(u)$can be estimated using the Nelson-Aalen estimator.

Alternative approaches for estimating transition probabilities include those that solve the Komogorov forward equation using either matrix exponentials or by calculating eigenvalues [Appendix A of Aalen et al. (2008)].

4.3.2 Semi-parametric Markov multi-state models

In the previous section, non-parametric estimators were described in the sense no covariates were involved. This section describes semi-parametric estimation

that allows for covariates to be included. The standard semi-parametric approach is Cox proportional hazard regression modelling for each transition. It is essentially the same as the cause-specific hazard approach to competing risks, with causes now replaced with the more broad transitions. The transition intensity (hazard) for transition h → j under a Cox model is defined as:

$$\alpha_{hj}(t) = \alpha_{hj,0}(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_{hj})$$

where $\mathbf{Z}$ is the vector of covariates at baseline
$\alpha_{h,j0}(t)$ is the baseline hazard for transition h → j
$\mathbf{Z}_{hj}$ is a vector of transition-specific covariates
and $\boldsymbol{\beta}$ is the vector of coefficients for the covariates.

It is semi-parametric in that the covariates introduce parameters but the distribution of the baseline hazard is unspecified.

The model specified above is the most general "full" model in that every transition has its own baseline hazard and each different transition has its own covariate effects. It is also possible for baseline hazards for relevant transitions to be related. For example, the baseline hazard for a transition that involves movement into the absorbing state death may have a proportional effect on the baseline hazard of another transition, that also involves movement into that state.

Just as in the previous sub-section, estimation of transition probabilities can be carried out using the Aalen-Johansen estimator. Strictly speaking, however, the Aalen-Johansen estimator does not involve any covariates, and therefore estimation in the semi-parametric approach is said to be by the Aalen-Johansen type estimator. However, in practice, estimation is for a particular patient with covariates evaluated for a specific combination of values. The cumulative hazard of each transition for the covariate combination of interest is all that is needed to calculate the transition probabilities. Estimation of a cumulative hazard for a particular covariate combination under Cox models is available in standard software packages. This can then be used as input in the Aalen-Johansen estimator.

4.3.3 Parametric Markov multi-state models

So far in this section, estimators have either been non-parametric or semi-parametric and as such no distributions have been specified for the (baseline) hazards. This part introduces parametric estimators whereby hazards follow a specified distribution. First, the time-homogeneous Markov model is outlined that uses constant hazards over time. Next, the piecewise constant model with hazards that are constant over fixed timed intervals is described. Finally, the concept of more general time-inhomogeneous models is introduced, where distributions are used that allow hazards to vary over time. Parametric models in particular facilitate extrapolation of survival, and other relevant outcomes, beyond the observation period of a study. Therefore, parametric models are also a particular focus of the extrapolation of survival section 6.5 of Chapter 6.

4.3.3.1 Time-homogeneous Markov model/constant hazard

It has just been mentioned that it can be beneficial to assume a specific parametric distribution for a transition hazard. The simplest and most commonly-applied parametric approach is the time-homogeneous Markov model where the hazard is assumed to be constant over time.

Mathematically, the transition intensity (hazard) for transition $h \rightarrow j$ is defined as:

$\alpha_{hj}(t) = \lambda \quad \forall\, t,\ t=[0, \tau]$
where λ is a constant and τ is the end of the observation period.

The time-homogeneous Markov process follows an exponential distribution with the hazard = λ, hazard rate= λt, and the transition probability $p_{hj}(t) = 1-\exp(-\lambda t)$.

4.3.3.2 Markov piecewise constant hazard model

It may not always be appropriate to assume that the hazard is constant (homogeneous) throughout the whole time-frame of interest. A widely-used approach that is used when the homogeneity assumption is not satisfied is the

piecewise constant hazard method. This is the simplest non-homogeneous approach. With this approach transition intensities (hazards) are defined by:

$\alpha_{hj}(t) = \alpha_{hj}^{l}$, $\theta_{l-1} < t \leq \theta_l$, $l = 1,2,3, \ldots, q$

where $\theta = (\theta_1, \ldots, \theta_{q-1})$ is a vector of cut-points for time intervals with

$0 = \theta_0 < \theta_1 < \theta_2 < \ldots < \theta_{q-1} < \theta_q = \infty$.

In other words, hazards are constant within fixed time intervals. It therefore offers the flexibility for hazards to be different across different time intervals.

4.3.3.3 More general Markov time-inhomogeneous parametric models

The assumption of constant hazards over the whole time period, or over time intervals, may be too stringent in practice. An alternative approach is to model the hazards with a parametric distribution that allows hazards to vary over time. For example, the Weibull and Gompertz distributions both have an additional parameter, compared to the exponential distribution, called a "shape" parameter. This parameter shapes the rate of change of the hazard over time. A Weibull distribution with a shape parameter of 1 and a Gompertz distribution with a shape parameter of 0 will each be equivalent to the exponential distribution. However, Weibull and Gompertz shape parameters >1 (<1) and >0 (<0) respectively indicate hazards that increase (decrease) over time. Distributions with the flexibility for varying hazards over time are detailed in the extrapolation of survival section, section 6.5, of Chapter 6.

4.3.3.4 Test of Markov property

There are several ways of overcoming violations, or relaxing, the Markov property. For example, extra states can be added to reflect the possible order of transitions (Putter et al., 2007). Another option is to use the state arrival extended (semi-) Markov approach. This involves including a covariate that depends on the time since entry into a previous state in a Markov model. As such, the state arrival extended approach could act as a useful tool in which to test the appropriateness of the Markov property. The significance, both

statistically and clinically, of the covariate could help aid the decision as to whether the Markov assumption was reasonable. However, as with any covariate included in a model, any relevant assumptions such as linearity or proportional hazards must be met. Alternative tests that are more appropriate when violations of such assumptions are evident have also been proposed e.g. Rodríguez-Girondo and Uña-Álvarez (2016).

4.4 Semi-Markov multi-state models

The Markov assumption can be relaxed by fitting a semi-Markov model, also known as embedded Markov or Markov renewal models. With the semi-Markov approach, the process depends not only on the present state but also the time since entry into the present state. However, the Aalen- Johansen estimator (and eigenvector/matrix exponential equivalent) presented in section 4.3 relies on the Markov property holding. This section describes a simulation-based approach to estimation of predictions that can be used with a semi-Markov model. It is the approach used in the `mssample` function in the `mstate R` package (de Wreede et al., 2010).

Given a cumulative hazard for each transition evaluated for a specified set of covariate values, i.e. for a specific patient, state occupancy probabilities can be estimated by simulation by repeatedly sampling complete paths through the multi-state model.

An algorithm to generate such paths is described in Fiocco et al. (2008) and repeated here. It is based on the ideas of Dabrowska (1995) for simulation in competing risks blocks. The algorithm makes use of the fact that a multi-state model can be separated into a series of (linked) competing risks blocks. It involves simulating transition times and states for each block. The algorithm is as follows:

Let i be the starting state and $t_i = s$ the starting time. Repeat the following steps:

1. Let $\mathcal{J}$ be the set of states that can be reached from i. If $\mathcal{J}$ =∅, stop. Otherwise, let, for j ∈ $\mathcal{J}$, $\Lambda_{ij}(t)$ be the cumulative hazard function for transition i → j .

2. Compute $\Lambda_i(t) = \sum_{j \in \mathcal{J}} \Lambda_{ij}(t)$

3. Sample $t^*(>t_i)$ from $\Lambda_i(t) - \Lambda_i(t_i)$. If $\Lambda_i(\infty)$is finite, $t^*=\infty$ may be sampled with positive probability.

4. If $t^*=\infty$, stop. Otherwise, select state j as the next state with probability $d\Lambda_{ij}(t^*)/\ d\Lambda_i(t^*)$.

5. Set i = j and $t_i = t^*$.

This process should then be replicated M times, where M is a large number, to generate complete paths through the multi-state model. The probability of a future event E given the patient's history $\mathbf{H}_s$, $P(E|\mathbf{H}_s)$, can then be estimated with $\mathbf{H}_s$ as specified by the starting state and starting time in the algorithm. An estimate of $P(E|\mathbf{H}_s)$ is given by the proportion of the M paths in which the event of interest E occurred.

Alternative kernel-based and landmark approaches are detailed in Spitoni et al. (2012) and Putter and Spitoni (2016) respectively.

As with Markov models, semi-Markov models can also be non-parametric, semi-parametric or parametric.

4.5 Flexibility of predictions at different times over the course of the disease

This section highlights one of the main advantages of multi-state modelling, that of the flexibility in predictions the approach can accommodate. Like standard survival analysis, multi-state modelling allows the calculation of predictions, at the start of a study, of having an event (health state) of interest at any given time over the period of observation of the study. Furthermore, because multi-

state modelling considers several different health states of interest simultaneously, it allows predictions of being in any given state at any given time, i.e. state occupancy probabilities. Just as above, the approach facilitates such predictions at the start of a study.

However, one of the additional benefits of multi-state modelling involving several health states of interest is that dynamic predictions can be carried out. That is to say, predictions can be updated as time elapses and especially as patients progress to other health states. This then allows, for example, predictions at the start of a study to be compared with those starting further along a disease process. In particular, predictions starting from a given health state at a specific time can be compared with those starting from a different health state at the same time.

In the multi-state modelling framework, the procedure used to calculate the state occupancy probabilities is depend on whether a Markov or semi-Markov model is built. In particular, obtaining dynamic predictions is more straightforward with a Markov than a semi-Markov model. In a paper by de Wreede et al. (2011) the authors explain that, as a patient moves to different states, the state occupancy probabilities obtained from a Markov model can be used to reflect these transitions. There are as many sets of state occupancy probabilities as there are states in the model, with each set corresponding to a different starting state. All these probabilities can be obtained automatically from just a single Markov model. An appropriate dynamic prediction can be obtained for a patient by changing the set of predictions that are used when entering a new state, with the prediction time updating to the time of transition. For example, say, an objective is to predict the probability of being in a health state representing the composite outcome relapse or death. Specifically, the prediction is for patient A who is in the initial state in the time interval [0, a), intermediate state from [a,b) and the relapse/death state from time b. Up until time a, the set of predictions with the initial state as a starting state would be used. However in [a,b) the predictions used would change to those with the intermediate state as the starting state.

The approach to (dynamic) prediction with Markov modelling relies heavily on the properties of the Aalen-Johansen estimator, which was outlined previously in section 4.3 on page 90. However this is not applicable to semi-Markov modelling, and therefore for such models an alternative approach to dynamic prediction is required. One such approach involves subsetting the data such that, for each starting state/time of prediction combination of interest, a dataset of patients who were in the given state at the given time is created. Modelling is then carried out with each different dataset to obtain the required dynamic predictions.

4.6 Broad review of multi-state models in the medical literature

Section 4.2 began by emphasising that there has been a rapid emergence of literature in medical journals providing a review of the multi-state modelling framework. This section investigates the extent to which the term multi-state modelling or related items has appeared in selected biostatistical, core clinical and general high impact clinical journals in a broad sense i.e. any mention of the terms from a methodology aspect or an application of the technique that is not specific to any particular area of medicine. The search strategy is detailed in Appendix VII. Figure 4.2(a) and (b) summarise the findings by year for the selected biostatistical and core clinical journals, respectively.

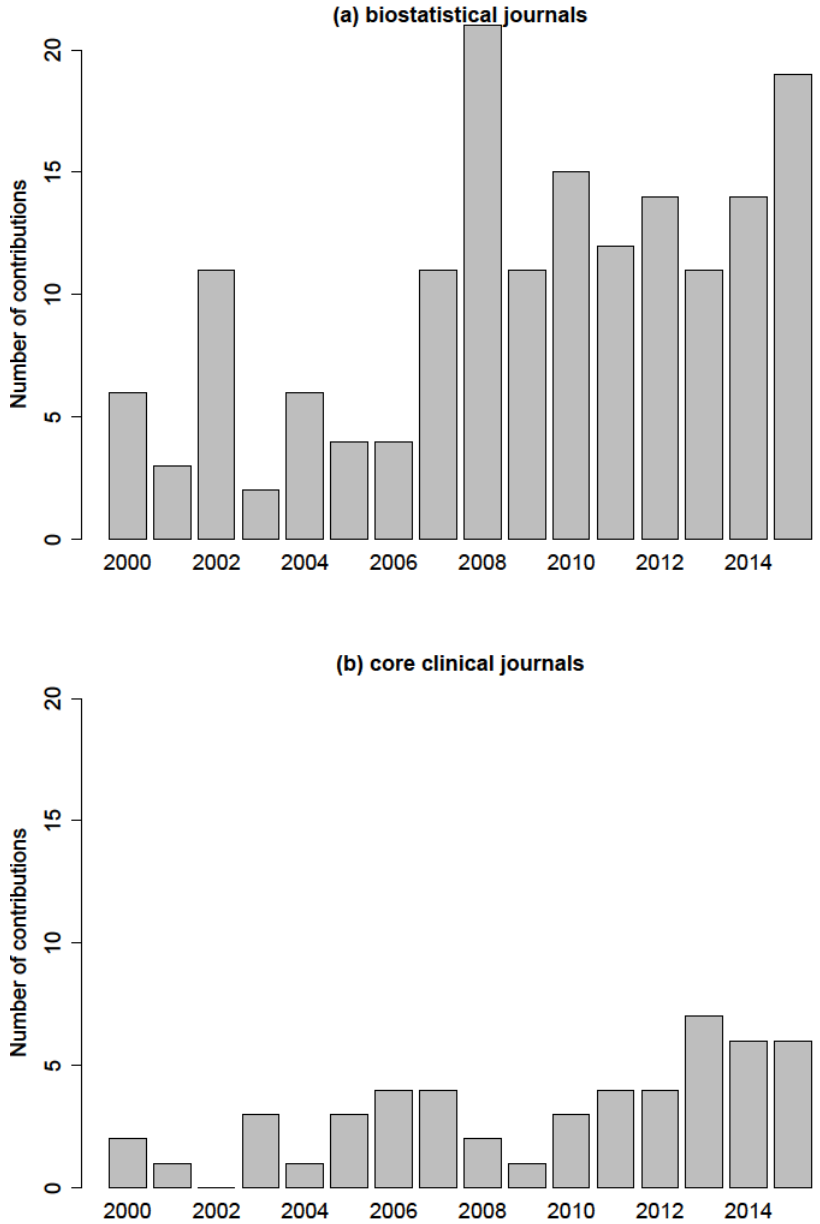


Figure 4.2 Broad search of multi-state modelling in the medical literature

The search found that, in the biostatistical journals, there was a steady emergence of between 11 and 21 articles in each year from 2007 onwards. However, the appearance of multi-state modelling in the core clinical journals was not as pronounced, only reaching 7 articles at most in any one year. In the general high impact clinical journals, there was only 3 articles in total over the 16-year period (one each in 2004, 2007 and 2015).

4.7 Review of illness-death modelling of stroke, recurrence and death in the literature

This section describes a review of the literature that focuses on illness-death modelling of stroke, recurrence and death. The purpose of this is to identify gaps in the research and provide the context for the empirical illness-death modelling of stroke, recurrence and death in the next chapter. An illness-death model is also known as a disability model and is a commonly-used multi-state model. A search strategy was developed to allow the review to take place. An OvidSP search of titles, abstracts and keywords of contributions to 29/09/2016 was performed using the following search terms:

("multi-state model" OR "multistate model" OR "multi state model" OR "illness-death model" OR "illness death model" OR "multi-state models" OR "multistate models" OR "multi state models" OR "illness-death models" OR "illness death models" OR "multi-state modelling" OR "multistate modelling" OR "multi state modelling" OR "illness-death modelling" OR "illness death modelling" OR "multi-state modeling" OR "multistate modeling" OR "multi state modeling" OR "illness-death modeling" OR "illness death modeling" OR "disability model*" OR ("multi-state" AND "Markov") OR ("multistate" AND "Markov") OR ("multi state" AND "Markov") OR ("illness-death" AND "Markov") OR ("illness death" AND "Markov") OR ("disability AND "Markov") OR ("multi-state" AND "semi-Markov") OR ("multistate" AND "semi-Markov") OR ("multi state" AND "semi-Markov") OR ("illness-death" AND "semi-Markov") OR ("illness death" AND "semi-Markov") OR ("disability AND "semi-Markov") OR ("multi-state" AND "model") OR ("multistate" AND "model") OR ("multi state" AND "model") OR ("illness-death" AND "model") OR ("illness death" AND "model") OR ("multi-state" AND "models") OR ("multistate" AND "models") OR ("multi state" AND "models") OR ("illness-death" AND "models") OR ("illness death" AND "models") OR ("multi-state" AND "modelling") OR ("multistate" AND "modelling") OR ("multi state" AND "modelling") OR ("illness-death" AND "modelling") OR ("illness death" AND "modelling") OR ("multi-state" AND "modeling") OR ("multistate" AND "modeling") OR ("multi state" AND "modeling") OR ("illness-death" AND "modeling") OR ("illness death" AND "modeling") AND (stroke OR ischemic OR ischaemic OR "intracerebral hemorrhage" OR

"intracerebral haemorrhage" OR "subarachnoid hemorrhage" OR "subarachnoid haemorrhage" OR ICH OR SAH) AND (recur* OR multiple)).ab,ti,kw.

The entire search found 15 contributions. After disregarding 8 duplicates, 3 that did not involve stroke patients, 1 that did not involve recurrence of stroke and 1 which did not involve multi-state modelling, this left only 2 contributions. Each of these contributions were published very recently. The first contribution was an abstract by Penn et al. (2016) for a conference held in February 2016. The second was a journal article by Wetmore et al. (2016) published in September 2016. The work by Penn et al. (2016) aimed to investigate whether referral to a transient ischaemic attack (TIA) unit reduced the risk of recurrent stroke. To do this, the authors used a multi-state model with the states symptom onset, referral, arrival at the unit, recurrent stroke and death. They found associations between age, ABCDD, gender and unit intervention and the transitions in the model. The main finding stated in the abstract was that referrals to TIA units reduced the risk of recurrent stroke within 90 days. The information was somewhat limited because only the abstract was available. However, this contribution did highlight, as well as being useful to study the clinical pathway of a disease (especially chronic or progressive), multi-state models can be used to see how patients flow through the health care system, what factors influence the transitions and possible critical points for intervention.

The article by Wetmore et al. (2016) reported on a study of hemodialysis patients who had experienced an ischemic stroke. The authors used a classic “illness-death” multi-state model to explore factors associated with stroke → recurrence, stroke → death and recurrence → death, with Cox regressions used to model each transition. However there was no clarification of whether a Markov or semi-Markov approach was used. The paper stated that age was treated as a time-dependent variable by assessing it at first stroke and then again at the recurrent stroke. This treatment of age suggests a “clock-reset” approach to time was adopted and a semi-Markov approach was used. However, the probability plots presented appear to be those that can be created after using the `probtrans` function in `mstate`, which is only applicable to Markov models. Updating the value of age on entering a new state does not seem

appropriate with a Markov model. This is because a Markov model measures time using the "clock-forward" approach which means time is measured from the initial state, even in states that are further along the clinical pathway of the disease.

Furthermore, on a number of occasions throughout the article, a parsimonious model is mentioned. However, while an AIC-based method has been devised by Kuk and Varadhan (2013) for the Fine and Gray proportional subdistribution hazards modelling approach to competing risks, no equivalent has yet been suggested for the cause-specific hazards modelling approach to competing risks used in multi-state modelling. Using AICs/likelihoods for variable selection could be inappropriate and needs to be treated with caution. This was discussed further in section 2.7.2 of chapter 2.

In addition, the explanation of transition probabilities is misguided (Wetmore et al. (2016)). The paper appears to be confusing state occupancy probabilities with transition probabilities. For instance, it uses the example of the transition probability for stroke → death but then outlines that this involves patients who could die directly after their first stroke, or die after experiencing a recurrent stroke. This latter explanation in fact relates to the state occupancy probability for the death state. In the multi-state modelling framework, the only probabilities it is appropriate to estimate is state occupancy probabilities. Because of the presence of competing risks, probabilities for a particular transition are not an appropriate quantity to estimate. This is analogous to estimating survival in a competing risks scenario, the inappropriateness of which is detailed in 2.3 of chapter 2.

In another statement that relates to the confusion between state occupancy and transition probabilities the paper states "the Cox proportional hazards framework facilitates calculates [sic] probabilities of transition from the current state to the next state within a given time interval" (Wetmore et al. (2016)). This is misleading because by using the term transition probability, rather than state occupancy probability, a naïve Cox approach to prediction is being advocated.

Wetmore et al. (2016) then proceed to show plots of predicted probabilities. The first plot shows the stacked predictions of being in a given state over time, i.e. the state occupancy probabilities. Nonetheless, the paper continues to cause confusion by describing the plot as a “stacked transition probability graph”. The predictions in this first plot were predictions from time zero and do not update to reflect any states patients may progress to as time elapses. In the second plot, the paper does attempt to plot such dynamic predictions. However, the approach to this is unconventional and misguided. The article aims to demonstrate “the impact on survival of a subsequent stroke”. It does this by showing what survival drops to should a patient experience a recurrent stroke at 1, 2 or 3 years. At the time of the recurrent stroke however, survival immediately drops 10-20% depending on when the recurrence occurred. This approach is not suitable because it fails to recognise that patients who have a recurrence are still alive at that point and so their survival is 100% at the time of recurrence. The article appears to have misinterpreted how to calculate dynamic predictions.

The review in this section has shown that there was very little in the way of literature that focused on illness-death modelling of stroke, recurrence and death. The two contributions that emerged were only published in 2016. One of these pieces was a conference abstract and as such was limited in the information it provided about the multi-state model. The other contribution was a journal article but misled readers on some aspects of multi-state modelling. Therefore, while it is encouraging to see the start of multi-state modelling involving stroke, recurrence and death in the clinical literature, there is much scope to raise awareness of the technique in this disease area. This illness-death model will be the focus of the next chapter of this thesis. The chapter will include Cox cause-specific hazard modelling of each transition to assess the effects of covariates. It will also present predictions both at the time of the index stroke and dynamically as patients enter new states. Furthermore, it will highlight the extra insight gained from the multi-state modelling approach over and above that of competing risks analysis.

Chapter 5 Multi-state modelling with the Edinburgh Stroke Study data

Chapter 3 presented competing risks analyses of the Edinburgh Stroke Study. These involved following patients to determine which one of the two events recurrence or death without recurrence they experienced first, if any. It did not consider any subsequent events experienced by those who had a recurrence. However, as explained in the previous chapter, the multi-state modelling approach can be used as an extension to competing risks to also consider events of interest that can happen after the first event. This chapter builds on the competing risks analysis of chapter 3 and presents a multi-state modelling approach.

Section 5.1 begins by describing the illness-death model used, a particular type of multi-state model, with the transitions stroke → recurrence, stroke → death and recurrence → death. The covariate effects on each of the transitions, at a univariable (unadjusted) and multivariable level are then presented. Predictions are then the focus of section 5.2, with both predictions at the time of index stroke and dynamic predictions illustrated. The dynamic predictions update the patient's prognosis taking into account the time elapsed and any subsequent events since the initial stroke. Next, section 5.3 highlights the extra insight gained from the multi-state modelling approach over and above the competing risk analysis in Chapter 3. Finally, section 5.4 summarises the analysis presented in this chapter.

5.1 Description of the multi-state model and results for covariate effects

This section begins by describing the specific multi-state model used in terms of the health states, and transitions between them, that were considered. The section then continues by demonstrating the effect of the covariates on each of the transitions.

Figure 5.1 displays the state-transition diagram for the illness-death model used throughout this chapter. The purpose of showing Figure 5.1 is to provide the context for all the analysis that follows. When introducing the model, particular emphasis is given to how it builds on from the previous competing risks analysis.

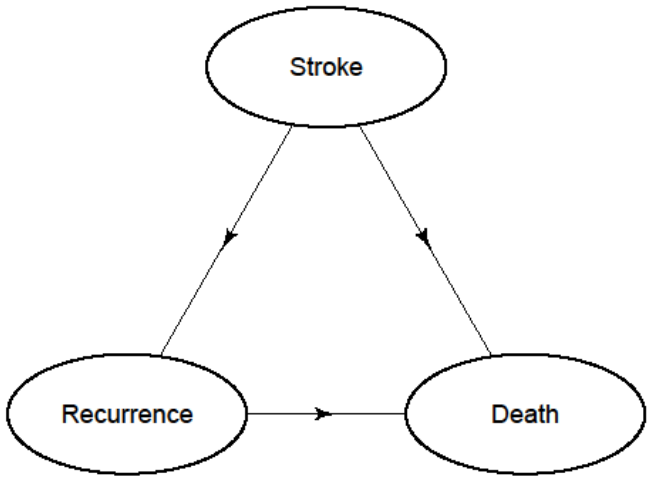


Figure 5.1 State-transition diagram for the illness-death model

It can be seen in Figure 5.1 that the three health states stroke, recurrence and death are of interest, together with the three transitions stroke → recurrence, stroke → death and recurrence → death.

While the competing risks analysis in chapter 3 also considered stroke → recurrence and stroke → death, the addition of recurrence → death in the multi-state modelling allows for different predictions. Now, recurrence acts as an intermediate state between the initial stroke state and final death stroke. Therefore, the multi-state modelling demonstrated in this chapter will build on the competing risks analysis and, in particular, allows comparison of the risk of death with and without recurrence.

The patient characteristics used in the Edinburgh Stroke Study were previously described, overall and by state in the model, in Table 3.1 within section 3.1 on page 56.

The effects of the characteristics, or covariates, on each transition were then assessed more formally using Cox proportional hazard regression models. All the modelling in this chapter was Cox proportional hazards regression. As such, the usual appropriate checks for violations of proportional hazards were carried out, although they are not shown for reasons of brevity.

Table 5.1 shows the effects of the covariates on the hazard of each of the transitions at the univariable (unadjusted) level.

The results for stroke → recurrence and stroke → death were previously presented in the competing risk analysis in section 3.2 of chapter 3. They are repeated in this chapter to allow comparison of the results between all three transitions in the multi-state model. However focus will be primarily on the effects of the covariates on recurrence → death, as the addition of this transition is what distinguishes the multi-state modelling from the previous competing risks analysis. Covariates found to be statistically significant at the 5% level are shown in bold and those found to be statistically significant at the 10% level are shown in italics.

Before carrying out the analysis for the recurrence → death transition, a Markov model was built to investigate the effect of time in the stroke state on post-recurrence death. The covariate for time in the stroke state was found to have a statistically significant effect on post-recurrence death (p-value <0.001). In addition, its coefficient (s.e) of 1.709 (0.314) was likely to be of a size of practical importance. That positive coefficient indicated that the longer patients spent in the recurrence-free state before reaching recurrence, the more likely they were to die following recurrence. It is unclear why this was the case.

Because the stroke state was a previous state, its statistical and practical significance indicates that the future did depend on patient history and therefore that the Markov property did not hold. Therefore a semi-Markov, rather than Markov, approach was taken for the modelling of the recurrence → death transition.

	Baseline	stroke -> recurrence hazard ratio (95% CI)	stroke -> recurrence p-value	stroke -> death hazard ratio (95% CI)	stroke -> death p-value	recurrence-> death hazard ratio (95% CI)	recurrence-> death p-value
age - centred on mean of 71		**1.02 (1.00, 1.03)**	**0.012**	**1.08 (1.06, 1.09)**	**<0.001**	**1.06 (1.03, 1.09)**	**<0.001**
male	Female	**0.72 (0.53, 0.98)**	**0.038**	**0.76 (0.59, 0.97)**	**0.025**	1.26 (0.77, 2.06)	0.360
independent before stroke	No	1.12 (0.55, 2.27)	0.763	**0.24 (0.18, 0.33)**	**<0.001**	0.59 (0.24, 1.49)	0.267
previous cerebrovascular event	No	*1.34 (0.98, 1.84)*	*0.066*	1.14 (0.88, 1.47)	0.329	1.09 (0.65, 1.82)	0.754
previous other occlusive vascular disease	No	**1.53 (1.12, 2.09)**	**0.008**	**1.74 (1.36, 2.23)**	**<0.001**	**2.03 (1.24, 3.34)**	**0.005**
previous hypertension	No	*1.33 (0.98, 1.81)*	*0.069*	0.93 (0.73, 1.18)	0.543	0.99 (0.60, 1.64)	0.972
diabetes	No	1.23 (0.79, 1.91)	0.352	1.16 (0.81, 1.67)	0.417	1.53 (0.80, 2.93)	0.202
current smoker and ex-smoker <12 months	Non-smoker	1.09 (0.79, 1.51)	0.602	**0.73 (0.57, 0.93)**	**0.011**	1.16 (0.68, 1.99)	0.588
lift both arms off bed	No	*1.70 (0.92, 3.14)*	*0.089*	**0.16 (0.13, 0.21)**	**<0.001**	**0.44 (0.20, 0.97)**	**0.043**
walk independently	No	0.94 (0.66, 1.33)	0.714	**0.22 (0.17, 0.28)**	**<0.001**	**0.40 (0.24, 0.66)**	**<0.001**
orientated speech	No	0.77 (0.49, 1.20)	0.245	**0.17 (0.13, 0.22)**	**<0.001**	**0.36 (0.20, 0.64)**	**0.001**
stroke syndrome			0.448		**<0.001**		**0.042**
Lacunar	Cortical	0.79 (0.55, 1.15)	0.218	**0.38 (0.26, 0.54)**	**<0.001**	**0.64 (0.35, 1.18)**	**0.151**
Other	Cortical	0.89 (0.60, 1.31)	0.551	**0.68 (0.49, 0.93)**	**0.015**	**0.43 (0.20, 0.92)**	**0.030**
high blood pressure	No	0.98 (0.71, 1.34)	0.881	1.08 (0.84, 1.38)	0.573	*0.57 (0.32, 1.00)*	*0.051*
delay in assessment			0.267		**<0.001**		**0.018**
2-6 days	0-1 days	0.90 (0.58, 1.39)	0.634	**0.63 (0.46, 0.85)**	**0.002**	**1.00 (0.53, 1.88)**	**1.000**
7 days or more	0-1 days	0.74 (0.51, 1.09)	0.125	**0.30 (0.22, 0.40)**	**<0.001**	**0.49 (0.26, 0.90)**	**0.021**
symptomatic carotid stenosis			**0.013**		**<0.001**		**0.026**
70-100% ICA	<70% ICA	**1.86 (1.26, 2.75)**	**0.002**	**1.43 (0.96, 2.13)**	**0.082**	**1.79 (0.99, 3.23)**	**0.053**
Unknown	<70% ICA	**1.21 (0.75, 1.94)**	**0.435**	**4.68 (3.59, 6.09)**	**<0.001**	**2.35 (1.20, 4.64)**	**0.013**
in atrial fibrillation	No	**1.65 (1.17, 2.34)**	**0.005**	**2.60 (2.01, 3.35)**	**<0.001**	**1.99 (1.19, 3.33)**	**0.008**
visible infarct on scan	No	1.10 (0.81, 1.50)	0.527	*1.26 (0.98, 1.62)*	*0.067*	*1.58 (0.93, 2.70)*	*0.090*
haemorrhage on scan	No	*1.64 (0.96, 2.78)*	*0.069*	**2.21 (1.52, 3.22)**	**<0.001**	1.83 (0.87, 3.85)	0.110

Table 5.1 Univariable Cox Regression for each of the transitions in the ESS multi-state model

It can be seen in Table 5.1 that aging, having previous other occlusive vascular disease, not being able to walk independently, not having orientated speech, having a cortical stroke, having the shortest delay in assessment, having high symptomatic carotid stenosis and having atrial fibrillation were each associated with a higher hazard of death after recurrence. For each of these effects, there were corresponding effects in the same direction for stroke → recurrence and

stroke → death. Not being able to lift both arms off the bed was also associated with a higher hazard of death after recurrence. However, it was associated with a reduced hazard of recurrence, albeit at the 10% significance level, with such patients having a higher hazard of death without recurrence. Furthermore, there was a reduced hazard of death after recurrence, albeit at the 10% significance level, for those with high blood pressure. This probably reflects that those with known high blood pressure were more likely to be receiving anti-hypertensive treatment. In addition, there was a higher hazard of death after recurrence, albeit at the 10% significance level, for those with a visible infarct on scanning.

An unparsimonious multivariable analysis was then carried out considering the variables together in a model, allowing the effect of each covariate to be adjusted by other relevant variables (confounders). Table 5.2 on the next page shows the resultant model.

It can be seen in Table 5.2 that aging, having previous other occlusive vascular disease, having the shortest delay in assessment and having high symptomatic carotid stenosis were still associated with a higher hazard of death after recurrence. In addition, having a visible infarct on scan was also now found to be associated with a higher hazard of death after recurrence.

A parsimonious multivariable model was then built for ease of subsequent prediction (Table 5.3). For the recurrence → death transition, backward selection using likelihood ratio tests were used to select variables for inclusion, with significance at the 5% level used as the criteria. The resultant model contained age, previous other occlusive vascular disease, symptomatic carotid stenosis and visible infarct on scan. The directions of the associations of each of these covariates with death after recurrence were as previously described for Table 5.2.

		stroke -> recurrence		stroke -> death		recurrence-> death	
	Baseline	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
age - centred on mean of 71		1.01 (0.99, 1.02)	0.233	**1.06 (1.05, 1.07)**	**<0.001**	**1.07 (1.03, 1.11)**	**0.001**
male	Female	**0.67 (0.49, 0.92)**	**0.014**	1.01 (0.78, 1.30)	0.950	*1.71 (0.95, 3.07)*	*0.071*
independent before stroke	No	1.19 (0.57, 2.46)	0.648	**0.66 (0.47, 0.93)**	**0.019**	2.58 (0.73, 9.18)	0.143
previous cerebrovascular event	No	1.17 (0.84, 1.63)	0.344	0.93 (0.71, 1.23)	0.624	1.44 (0.80, 2.61)	0.225
previous other occlusive vascular disease	No	*1.33 (0.95, 1.85)*	*0.094*	1.19 (0.91, 1.54)	0.203	**2.66 (1.50, 4.72)**	**0.001**
previous hypertension	No	1.12 (0.81, 1.56)	0.480	0.86 (0.66, 1.12)	0.263	0.95 (0.54, 1.68)	0.859
diabetes	No	1.16 (0.74, 1.82)	0.525	1.20 (0.82, 1.74)	0.353	1.83 (0.85, 3.93)	0.121
current smoker and ex-smoker <12 months	Non-smoker	1.14 (0.82, 1.60)	0.434	1.09 (0.84, 1.42)	0.506	0.90 (0.48, 1.68)	0.742
lift both arms off bed	No	**2.56 (1.27, 5.13)**	**0.008**	**0.37 (0.26, 0.53)**	**<0.001**	0.79 (0.28, 2.19)	0.647
walk independently	No	1.03 (0.66, 1.62)	0.887	0.80 (0.55, 1.15)	0.232	0.92 (0.41, 2.06)	0.833
orientated speech	No	0.88 (0.53, 1.46)	0.616	**0.46 (0.33, 0.64)**	**<0.001**	0.58 (0.25, 1.33)	0.197
stroke syndrome			0.981		0.240		*0.074*
Lacunar	Cortical	0.98 (0.66, 1.45)	0.925	0.73 (0.49, 1.07)	0.104	*0.99 (0.48, 2.03)*	*0.973*
Other	Cortical	1.03 (0.68, 1.56)	0.897	0.97 (0.69, 1.37)	0.868	*0.39 (0.16, 0.95)*	*0.037*
high blood pressure	No	1.06 (0.76, 1.48)	0.726	0.81 (0.61, 1.07)	0.137	0.73 (0.39, 1.36)	0.323
delay in assessment			0.307		0.718		**0.016**
2-6 days	0-1 days	0.80 (0.51, 1.26)	0.342	0.88 (0.64, 1.21)	0.424	**2.00 (0.89, 4.49)**	**0.092**
7 days or more	0-1 days	0.70 (0.45, 1.09)	0.118	0.95 (0.66, 1.37)	0.773	**0.68 (0.29, 1.61)**	**0.381**
symptomatic carotid stenosis			**0.012**		**<0.001**		0.107
70-100% ICA	<70% ICA	**1.94 (1.28, 2.96)**	**0.002**	**1.15 (0.76, 1.75)**	**0.499**	1.43 (0.72, 2.85)	0.303
Unknown	<70% ICA	**0.96 (0.55, 1.67)**	**0.887**	**2.67 (1.93, 3.70)**	**<0.001**	2.82 (1.06, 7.45)	0.037
in atrial fibrillation	No	**1.60 (1.10, 2.33)**	**0.014**	*1.27 (0.96, 1.68)*	*0.093*	1.11 (0.58, 2.13)	0.752
visible infarct on scan	No	1.13 (0.81, 1.57)	0.484	0.98 (0.74, 1.32)	0.915	**3.01 (1.48, 6.12)**	**0.002**
haemorrhage on scan	No	**2.06 (1.09, 3.91)**	**0.027**	0.69 (0.43, 1.11)	0.130	1.94 (0.64, 5.85)	0.239

Table 5.2 Unparsimonious multivariable Cox regression for each of the transitions in the ESS multi-state model

		stroke -> recurrence		stroke -> death		recurrence-> death	
	Baseline	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
age - centred on mean of 71		1.01 (0.99, 1.02)	0.240	1.06 (1.05, 1.07)	<0.001	1.08 (1.05, 1.11)	<0.001
male	Female	0.69 (0.50, 0.94)	0.021	1.01 (0.79, 1.30)	0.909		
independent before stroke	No	1.22 (0.59, 2.51)	0.589	0.66 (0.47, 0.91)	0.013		
previous other occlusive vascular disease	No	1.41 (1.02, 1.95)	0.037	1.21 (0.93, 1.56)	0.154	2.42 (1.46, 4.03)	0.001
lift both arms off bed	No	2.37 (1.25, 4.49)	0.008	0.34 (0.26, 0.46)	<0.001		
orientated speech	No	0.79 (0.48, 1.28)	0.340	0.43 (0.32, 0.58)	<0.001		
symptomatic carotid stenosis			0.004		<0.001		0.003
70-100% ICA	<70% ICA	2.06 (1.38, 3.07)	<0.001	1.26 (0.84, 1.89)	0.271	1.91 (1.06, 3.47)	0.032
Unknown	<70% ICA	0.99 (0.57, 1.71)	0.970	2.85 (2.08, 3.89)	<0.001	3.33 (1.65, 6.71)	0.001
in atrial fibrillation	No	1.67 (1.16, 2.40)	0.006	1.31 (1.00, 1.72)	0.052		
visible infarct on scan	No					2.11 (1.21, 3.66)	0.008
haemorrhage on scan	No	1.94 (1.05, 3.58)	0.036	0.68 (0.45, 1.04)	0.075		

Table 5.3 Parsimonious multivariable Cox Regression for each of the transitions in the ESS multi-state model

The modelling shown in Table 5.3 for each of the three transitions formed the basis of the predictions at the time of the index stroke displayed in the next section. Modelling was also carried out in preparation for dynamic predictions.

Such predictions took place after a specified time point and accounted for the time that had elapsed and any transitions experienced by the patients up to that time point. Time points of 6 months and 1 year on from the index stroke were used. Therefore modelling was undertaken using just those patients who had managed to survive for 6 months after their stroke, and then again with those who were still alive 1 year on from their stroke. Separate modelling was carried out for those who managed to stay in the initial stroke state to the specified time and for those who had experienced a recurrence by that time. For the modelling for those still in the initial stroke state, multi-state modelling of the three transitions was carried out. A similar approach was used for variable selection as for the predictions at the time of the index stroke. For those who experienced a recurrence, modelling of death after recurrence was carried out.

Table 5.4 shows the results of modelling for those who were still in the initial stroke state 6 months on from their stroke. This involved 1037 patients amongst whom there were subsequently 92 recurrences, 135 deaths without recurrence and 32 deaths after recurrence. Due to the modelling using a different cohort of patients – smaller in number but having survived 6 months free of recurrence – than at the time of the index stroke, the covariates included were slightly different. However, the effects of the covariates that were included were generally similar and in the same direction as in the previous modelling.

		stroke -> recurrence		stroke -> death		recurrence-> death	
	Baseline	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
age - centred on mean of 71		1.02 (1.00, 1.04)	0.029	1.07 (1.05, 1.09)	<0.001		
male	Female	0.68 (0.44, 1.04)	0.073	0.95 (0.67, 1.35)	0.785		
independent before stroke	No	1.29 (0.47, 3.55)	0.624	0.51 (0.31, 0.84)	0.008		
previous other occlusive vascular disease	No					4.09 (1.91, 8.79)	<0.001
lift both arms off bed	No	1.76 (0.79, 3.90)	0.167	0.66 (0.42, 1.05)	0.081		
orientated speech	No	0.73 (0.38, 1.41)	0.353	0.55 (0.35, 0.87)	0.010		
symptomatic carotid stenosis			0.059		0.004		0.008
70-100% ICA	<70% ICA	2.03 (1.18, 3.50)	0.011	1.55 (0.94, 2.56)	0.086	2.85 (1.20, 6.78)	0.018
Unknown	<70% ICA	1.07 (0.52, 2.19)	0.851	2.24 (1.39, 3.61)	0.001	3.87 (1.48, 10.17)	0.006
in atrial fibrillation	No	1.28 (0.76, 2.16)	0.361	1.56 (1.06, 2.29)	0.024	3.53 (1.58, 7.89)	0.002
haemorrhage on scan	No	2.12 (0.98, 4.57)	0.055	0.51 (0.23, 1.11)	0.088		

Table 5.4 Parsimonious multivariable Cox Regression for each of the transitions in the ESS multi-state model: in initial stroke state 6 months on from stroke

Table 5.5 shows the corresponding results for those who were still in the initial stroke state 1 year on from their stroke. This involved 960 patients amongst whom there were subsequently 64 recurrences, 88 deaths without recurrence and 23 deaths after recurrence.

		stroke -> recurrence		stroke -> death		recurrence-> death	
	Baseline	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value	hazard ratio (95% CI)	p-value
age		1.03 (1.00, 1.05)	0.031	1.06 (1.03, 1.08)	<0.001		
independent before stroke	No	1.61 (0.39, 6.64)	0.511	0.43 (0.23, 0.78)	0.006		
previous other occlusive vascular disease	No	1.04 (0.61, 1.79)	0.873	1.44 (0.94, 2.23)	0.097	4.01 (1.59, 10.10)	0.003
lift both arms off bed	No	1.49 (0.59, 3.78)	0.400	0.52 (0.30, 0.91)	0.021		
stroke syndrome			0.346		0.111		
Lacunar	Cortical	0.74 (0.40, 1.36)	0.325	0.56 (0.32, 0.99)	0.047		
Other	Cortical	0.63 (0.30, 1.29)	0.205	0.77 (0.43, 1.39)	0.388		
symptomatic carotid stenosis			0.041		0.230		0.010
70-100% ICA	<70% ICA	2.24 (1.20, 4.17)	0.011	1.37 (0.74, 2.54)	0.311	3.66 (1.38, 9.69)	0.009
Unknown	<70% ICA	1.62 (0.74, 3.53)	0.225	1.64 (0.90, 3.01)	0.108	5.09 (1.39, 18.66)	0.014
in atrial fibrillation	No	0.94 (0.47, 1.87)	0.853	1.68 (1.05, 2.70)	0.031		

Table 5.5 Parsimonious multivariable Cox Regression for each of the transitions in the ESS multi-state model: in initial stroke state 1 year on from stroke

Age was the only predictor included in the modelling of death after recurrence for the 56 patients (with 12 subsequent deaths) who were in the recurrence state 6 months on from their stroke. Age, centred on the mean of 71, had a coefficient (s.e) of 0.0533 (0.0265) and p-value of 0.044 in the resultant Cox model. The null model was used to model death after recurrence in the 70 patients who were in the recurrence state 1 year on from their stroke, 7 of whom experienced death.

5.2 Results of multi-state modelling: predictions

This section highlights some of the possibilities in prediction with multi-state modelling. Two different prediction procedures are illustrated: prediction from the start of the study and dynamic prediction. The dynamic aspect of the latter is two-fold, taking into account changes in prediction as time elapses and also as

patients move to different health states. For each type of prediction, four patient profiles of differing risk are used for illustration. The patient profiles used can be seen in Table 5.6.

	(a) high risk	(b) low risk	(c) mid risk 1	(d) mid risk 2
age at stroke	61	61	71	71
sex	female	male	male	male
independent before stroke	×	✓	✓	✓
previous cardiovascular or peripheral vascular disease	✓	×	×	×
lift both arms off bed	×	✓	×	✓
orientated speech	×	✓	✓	✓
symptomatic carotid stenosis	Unknown	<70% ICA	70-100% ICA	70-100% ICA
in atrial fibrillation	✓	×	×	×
visible infarct on scan	✓	×	×	×
haemorrhage on scan	×	✓	×	×
stroke syndrome	cortical	lacunar	cortical	cortical

Table 5.6 Risk profiles used for prediction

5.2.1 Predictions from the start of the study

Figure 5.2 on the next page shows, for each of the different risk profiles, predictions of being in each of the health states in the model with death divided into pre- and post- recurrence deaths. Predictions were at the time of the index stroke.

It can be seen in Figure 5.2(a) that the probability of high-risk patients remaining in the initial stroke state (Alive, no recurrence) was essentially zero by the end of the 4 years of observation. This was primarily due to the very high risk of death without recurrence at 97% by 4 years. It could also be seen that the risk of death without recurrence was particularly severe within the first year post stroke. The risk of such a death was already 50% by 0.1 years post-stroke, rising to 88% by 1 year. In addition, the percentage of patients with recurrence at any one time did not exceed 1.1%. However this risk was only prevented from increasing by the risk of post-recurrence deaths, which was 3.2% by 4 years.

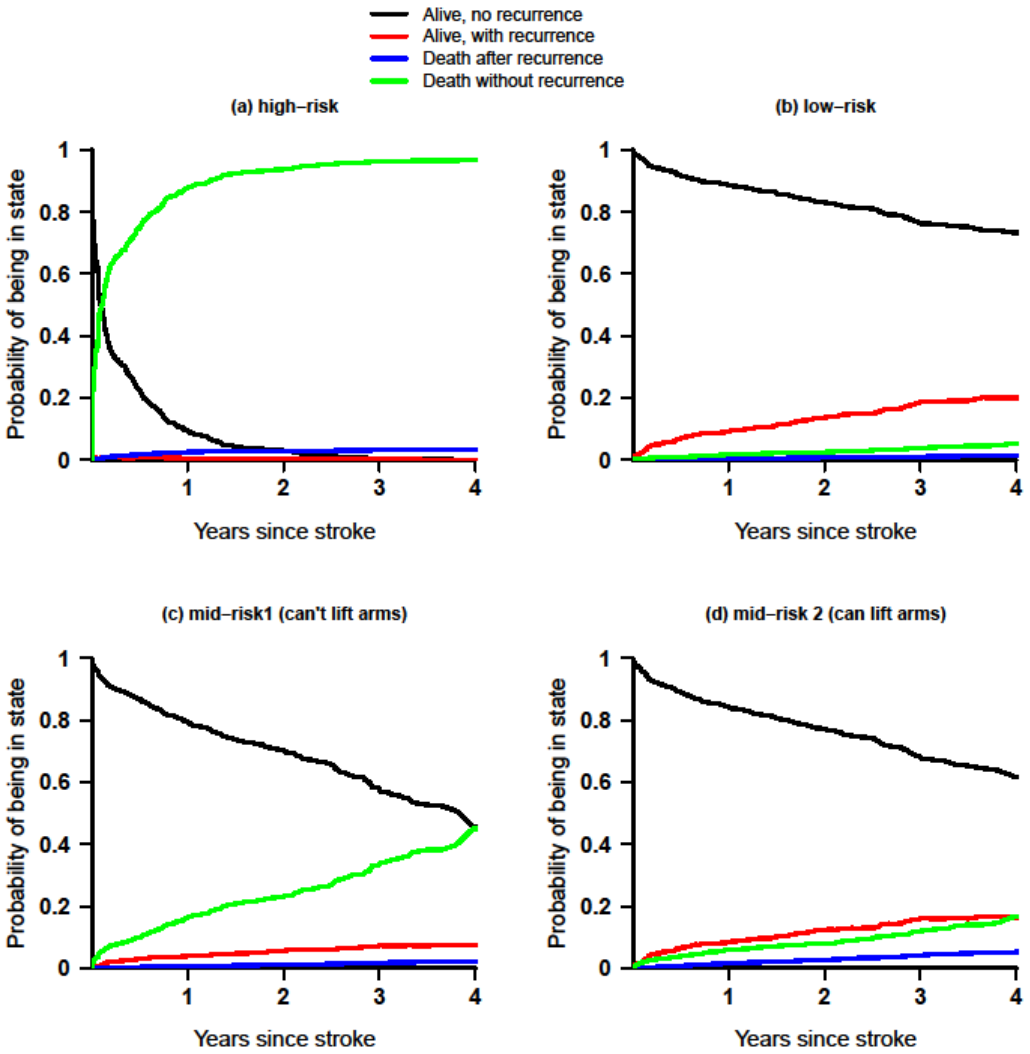


Figure 5.2 Predictions, at the time of the index stroke, of being in each health state over time

In contrast, 73% of those with the specified low-risk profile were predicted to still be in the initial stroke state 4 years post stroke (Figure 5.2(b)). Patients who did move from the stroke state were most likely to be alive with a recurrence. The percentage of patients with a recurrence had generally rose over the 4 years to 20% by 4 years. This was because there was a minimal risk of post-recurrence deaths. The risk of such deaths was only at 1% by 4 years. Even the risk of pre-recurrence deaths, prominent in the other risk groups, was only 5% by 4 years.

For those with the mid-risk1 profile (Figure 5.2(c)), the probability of remaining alive without a recurrence, i.e. in the initial stroke state, was 46% 4 years post stroke. By far the most likely reason to move from the initial stroke state was pre-recurrence death with the percentage of patients doing so having reached 45% by 4 years. The risk of being in recurrence at any one time and of post-recurrence death was minimal at 7% and 2% respectively at 4 years.

Figure 5.2(d), (mid-risk2 profile), is shown to demonstrate the effect of being able to lift both arms off the bed. The only difference between patients with mid-risk1 (Figure 5.2(c)) and mid-risk2 (Figure 5.2(d)) was that the latter were able to lift both arms off the bed and the former were not. Comparing Figure 5.2(c) and Figure 5.2(d) it can be seen that those who were able to lift both arms off the bed were more likely to be alive with no recurrence, 16% versus 7% of patients, respectively, 4 years post stroke. In addition, the risk of death without recurrence was much lower in those who could lift both arms off the bed, compared to those who could not, at 17% versus 45% 4 years post stroke. A higher percentage of those who could lift both arms off the bed were predicted to die after recurrence at 5% versus 2% in those who could not lift both arms off the bed. This is expected as there was a higher percentage of recurrences amongst those who could lift both arms off the bed compared to those who could not do so.

5.2.2 Dynamic predictions

The remainder of this section now focuses on dynamic predictions. These predictions provide an update to the predictions from the start of the study, that takes into account the time that has elapsed and whether any states of interest in the model have been experienced.

Figure 5.3 shows, for patients who were still in the stroke state at the time of prediction, dynamic predictions of being in each of the health states in the model with death divided into pre- and post-recurrence deaths. Each of the four plots represent a different risk profile. Different health states are distinguished by colour and different starting times of prediction are distinguished by line style.

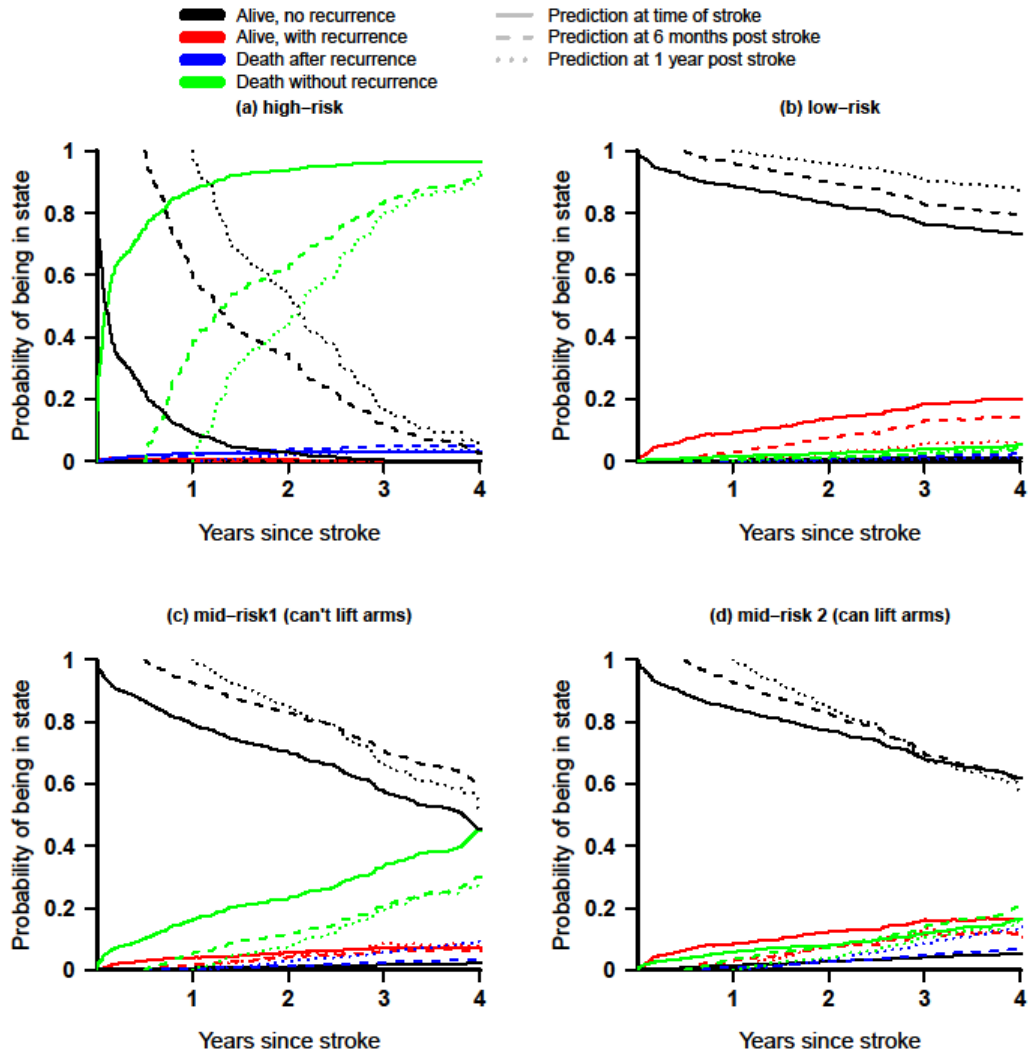


Figure 5.3 Dynamic predictions of each event

Figure 5.3(a) demonstrates, for those at high risk, the predictions of being alive with no recurrence from the later starting times showed substantial improvement, albeit with convergence resulting in similar predictions by 4 years. For instance, the 1-year prediction of being alive with no recurrence at the time of stroke of 9% improved to 61% for those who managed to stay in that state 6 months on from the index stroke. In addition, the corresponding 2-year prediction improved from the 3% at the time of the stroke to 34% and 55%, respectively, 6 months and 1 year on from the stroke. However, there was a lot less discrepancy in the 4-year predictions of 0%, 2% and 3% at the time of stroke, 6 months on and 1 year on respectively. It can also be seen in Figure 5.3(a) that,

for predictions from each of the starting points, the probability of being in recurrence was virtually zero over time. This was due to the post-recurrence deaths which prevented this probability from increasing. The percentage of post-recurrence deaths was also relatively low, and therefore the dramatic drops seen in the percentage of being alive with no recurrence were predominately due to the sharp increases in the percentage of pre-recurrence deaths.

For those at low risk, the 4-year predictions of being alive with no recurrence improved from the 73% at the time of the index stroke to 80% starting 6 months on, and improved further still to 88% when the prediction started 1 year on (Figure 5.3(b)). Figure 5.3(b) also exhibits a reduction in the 4-year probability of being in recurrence, at 0.20, 0.14 and 0.06 for the predictions starting at the index stroke, 6 months on and 1 year on respectively.

It can be seen in Figure 5.3(c) that, for those at mid-risk, the 4-year prediction of staying alive with no recurrence 6 months on from the index stroke was 60%. This was a marked improvement on the corresponding prediction at the time of the index stroke of 46%. It can be clearly seen that this was due to a corresponding reduction in pre-recurrence deaths. It can also be seen in Figure 5.3(c) that the probability of being alive with no recurrence at 4 years was lower when the prediction started 1 year on, compared to 6 months on, from the index stroke. It was evident that this was due to a higher percentage of post-recurrence deaths by 4 years when prediction started 1 year on, compared to 6 months on, from the index stroke.

The most marked difference between those at mid-risk who could not lift both arms off the bed (Figure 5.3(c)) and those who could (Figure 5.3(d)) was in the percentage of pre-recurrence deaths. In the latter, the percentage of pre-recurrence deaths at 4 years was only 17% when predicting from the time of the index stroke, with similar predictions when starting 6 months on and 1 year on from the index stroke. Consequently, Figure 5.3(d) did not demonstrate the improvement in the prediction of being alive with no recurrence seen in Figure 5.3(c) between the predictions that started at the index stroke and 6 months on from then. Instead, in Figure 5.3(d), the predictions of being alive with no

recurrence starting from progressively later times showed initial improvement before converging to a similar point by 4 years. The convergence of the predictions at the time of the index stroke and 6 months on was due to a combination of, in the predictions starting 6 months on, a steady relatively smaller frequency of recurrences over time together with the percentage of pre-recurrence deaths rising more sharply. In addition to this, a rise in the percentage of post-recurrence deaths also contributed to the convergence of the predictions that started at 1 year.

5.3 Comparison with competing risks analysis

This section demonstrates the differences in predictions that can occur when using multi-state modelling compared to competing risks analysis. The risk profile mid-risk 1 previously outlined in Table 5.6 on page 112 is used for illustration.

Figure 5.4 shows the competing risks cumulative incidence of recurrence alongside the probability of being in recurrence from the multi-state model. The probability of death after recurrence from the multi-state model is also shown because it aids interpretation of the difference in predictions as explained below.

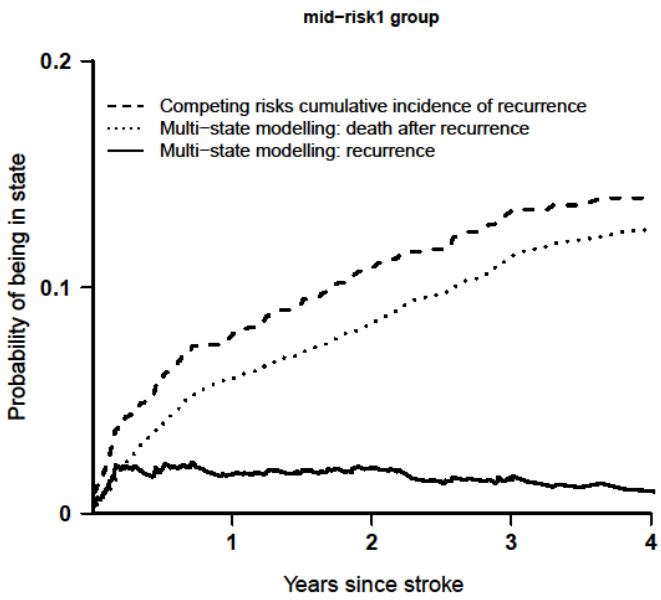


Figure 5.4 comparison of predictions of recurrence: multi-state modelling vs competing risks

It can be seen in Figure 5.4 that there was a marked discrepancy between the competing risks cumulative incidence of recurrence (dashed line) and the probability of being in recurrence from the multi-state model (solid line). Due to the estimate from the competing risks analysis being a cumulative incidence, it increased monotonically. It reflected the proportion of patients who had experienced a recurrence by a certain point in time. However, it did not take into account that patients will not be in the recurrence state indefinitely; patients will all inevitably die and therefore leave the recurrence state and enter the death state. The inclusion on the plot of the probability of death after recurrence from the multi-state model (dotted line) helps emphasise this. The competing risks analysis considered the flow from the stroke state into the recurrence state. However the predictions from the multi-state modelling gave a fuller picture by reflecting the flow both in and out of the recurrence state.

5.4 Discussion

This chapter presented analysis demonstrating the effects of covariates on each of the hazards for each relevant transition in a multi-state model of stroke, recurrence and death. Furthermore, it illustrated how these effects translated into state occupancy probabilities, with predictions at the time of the index stroke, and dynamically as time elapsed with any transitions experienced taken into account. In addition, it demonstrated the extra insight that can be gained from predictions from a multi-state model over and above that from a competing risks analysis. In particular, it emphasised that because cumulative incidences are the way predictions are presented from a competing risk analysis, the effect is one of monotonic increases. However, predictions in multi-state modelling consider the flow in and out of a health state and therefore provide additional insight to inform clinical decision making. That is to say, for a patient with a given set of risk factors, the modelling can give a prediction of being alive with recurrence over time, rather than just a prediction of the time that a recurrence is likely to happen. In doing so, it can help in identifying those likely to benefit from interventions to prolong their life after recurrence, or perhaps a group for inclusion in a trial investigating strategies to prevent subsequent recurrences. From another perspective, it may help with resource allocation by

differentiating those patients unlikely to benefit from such strategies because they have been severely debilitated and show no sign of improvement, and it may be more appropriate not to prolong their suffering.

Multi-state modelling could provide a useful tool for clinicians to predict outcomes for their patients and to aid discussions with them. In particular, dynamic predictions could be used to communicate any improvements in prognosis as time has elapsed. For example, the concept of overcoming a "critical risk period" could be emphasised as positive news for a patient. That is to say, explaining that now they have survived through a particular period when the risk of death (or recurrence) is high, highlight the improvement in prognosis compared to earlier. Another use for the predictions that multi-state modelling can accommodate could be to encourage patients to change an aspect of an unhealthy lifestyle to improve their health. This is discussed further in Chapter 8.

A consideration with dynamic prediction can be whether the values for covariates, where appropriate, should be updated to reflect the transitions experienced. This is particularly applicable to models involving recurrence, as often updated information on covariates is available at the time of recurrence, such as severity of disease. Even when predicting at the time of the initial state, it can be tempting to use covariates measured at the time of recurrence in the modelling of the recurrence → death transition. However, this would not be of any practical use for prediction for those in the initial state, at time 0 or dynamically, because this information would not be known at the time of prediction. Furthermore, in particular, all covariates included in any transition in a Markov model would need to start at the initial state value, as this is the way time is measured in Markov modelling.

However, the landmarking approach does offer more flexibility when the predictions are for patients in the recurrence state at the time of prediction. Therefore, covariates measured at the time of recurrence can be considered. The data used for the analysis in this chapter was originally extracted for a competing risks analysis and did not contain covariate information at the time of

recurrence. However, because the study formed a register of stroke patients, all covariate information collected at the time of stroke would have also been available at the time of recurrence of strokes, because recurrences would have been included as separate strokes. That the analysis did not include covariates measured at the time of recurrence was also a limitation due to the omission of potential confounders. It is quite possible/likely that age at recurrence and severity of recurrence account for some of the association found between having a recurrence and subsequent mortality after recurrence. Furthermore, the recording of such information at post-stroke assessments could help to inform dynamic predictions, for both those still in the initial stroke state and those who had experienced a recurrence by the time of prediction. Therefore, it is worth bearing in mind the flexibility offered by the approach to accommodate updated covariate information and to exploit this in future research. Doing so would enhance the insights such as those outlined at the end of the first paragraph of this discussion section.

Chapter 6 Health economic modelling:

two common approaches, review of the use of multi-state modelling, the assessment of model fit and extrapolation

This chapter begins by summarising two common approaches in health economic modelling: partitioned survival (section 6.1) and Markov decision-analytic modelling (section 6.2). The purpose of this is to provide background and context for the next chapter which presents empirical analysis that compares these two methods with multi-state modelling. In order to put into context the use of multi-state modelling for health economic modelling, section 6.3 contains a review of multi-state modelling in the health economics literature. The chapter then discusses the assessment of fit of models over the observed period of a study in section 6.4. Finally, section 6.5 focuses on extrapolation of outcomes beyond the observed period, which is a particularly important issue and has been a prominent topic in the health economics literature in recent years. After some initial background to the need for extrapolation, section 6.5 describes the conventional approach to extrapolation of parametric regression and some alternative approaches. The topic of extrapolation is then revisited at the end of this thesis in Chapter 8, where further developments in contributions to the literature in this fast-moving area are discussed.

This thesis does not describe the following standard concepts in health economic modelling for which references are provided for the interested reader: calculating cost-effectiveness (Drummond et al., 2005), assessing robustness of results with one-way and probabilistic sensitivity analyses [Drummond et al. (2005) and Briggs et al. (2006)] and discounting of costs and benefits (Drummond et al., 2005).

These concepts are not specific to any one approach to modelling. In particular, they are applicable when using the two approaches described in this chapter and the multi-state modelling described in chapter 4.

6.1 Partitioned survival

One common approach in health economic modelling is partitioned survival. Partitioned survival involves splitting overall survival into partitions that represent different health states experienced while alive. Typically, each health state differs in terms of the quality of life it provides. The method was proposed by Glasziou et al. (1990) and is often called the Q-TWiST (Quality-adjusted Time Without Symptoms of disease and Toxicity of treatment) method. It is also described in Gelber et al. (1991) and Gelber et al. (1995), with the latter forming the basis for the description presented in this section.

There are three steps to the partitioned survival method:

1. define relevant outcomes that will determine the health states of interest
2. partition the overall survival
3. compare the treatments

Step 1: Define the relevant outcomes/health states of interest

Illustration of The Q-TWiST method normally involves the outcomes toxicity, disease-free survival and overall survival. It gets its name because time without either symptoms of disease or toxicity of treatment is one of the health states of interest. This description of the method, however, uses the outcomes progression-free survival and overall survival. This is a common approach in the oncology setting and chapter 7 of this thesis, the empirical chapter that includes an application of this method, is based on a clinical trial with these outcomes.

Given that the relevant outcomes are progression-free survival and overall survival, the health states while alive that can be considered are progression-free and progression. Figure 6.1 illustrates these different health states across time to eventual death for an individual patient. Utility weights are often assigned to health states to represent the quality of life experienced while in them. For illustration, Figure 6.1 assumes utilities of 0.8 and 0.6 for the progression-free and progression states respectively.

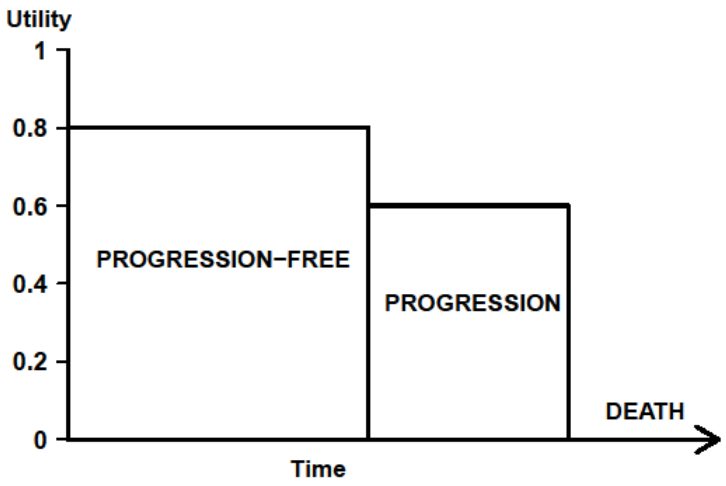


Figure 6.1 Illustration of quality-adjusted survival

The two extremes for utility weights are 1 representing perfect health and 0 for death or states as bad as death. The utilities of 0.6 for progression and 0.8 for progression-free reflect the quality of life experienced relative to these two extremes. An estimate of quality-adjusted survival can then be obtained, for an individual patient, from:

$$\text{Q-SURV} = U_{PROGFREE} \times \text{PROGFREE} + U_{PROG} \times \text{PROG}$$

where PROGFREE = time spent progression-free,

PROG = time spent in the progression state,

$U_{PROGFREE}$ = utility while progression free,

U_{PROG} = utility while in the progression state.

However, interest is in comparing treatment groups rather than individual patients. For this, PROGFREE and PROG need to represent the mean time spent in these states in a particular treatment group. Their derivation is explained in Steps 2 and 3 below.

Step 2: Partition overall survival

The next step is to partition the overall survival Kaplan-Meier curve into the health states defined in Step 1. Partitioning overall survival into progression-free and progression requires the progression-free survival Kaplan-Meier curve. Each

treatment group should be considered separately. Figure 6.2 shows an example of the partitioning of overall survival (OS) using progression-free survival (PFS).

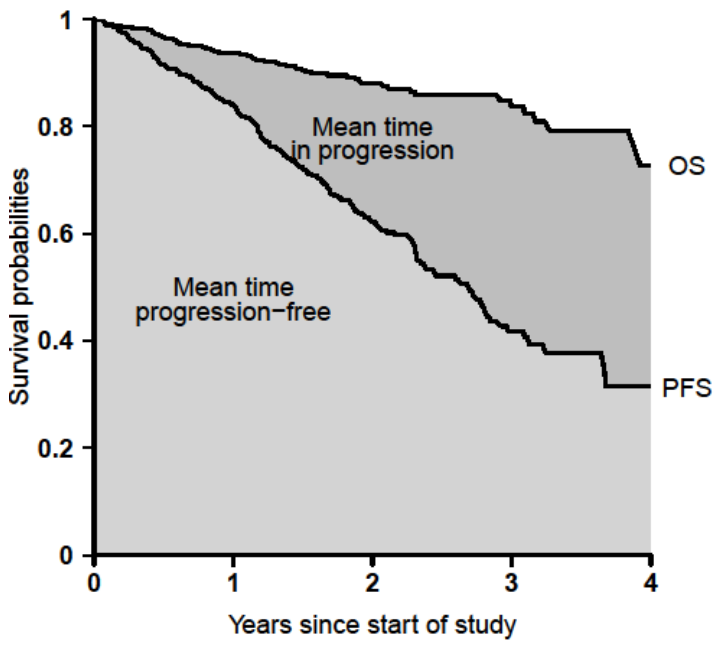


Figure 6.2 Illustration of partitioned survival

The area between the curves provide estimates of the mean duration in the relevant health state. Therefore, the mean duration in the progression state can be estimated from the area between the overall survival (OS) and progression-free survival (PFS) curves. Similarly, an estimate of the mean duration progression-free can be obtained from the area under the progression-free survival curve.

However, not every patient in the study in this illustration was observed long enough to reach progression and/or die. Therefore the progression-free survival and overall survival curves did not represent the entire lifetime of patients.

When the follow-up in a study stops before every patient has experienced the relevant outcomes, the estimates of mean duration in relevant states are said to be restricted means. In order to obtain estimates of the mean time in relevant states, instead of restricted means, parametric survival regression is one approach that can be used to extrapolate survival curves until the whole lifetime is represented. This is described later in this chapter in section 6.5.

Step 3: Comparison of treatments

The final step in the partitioned survival approach is to compare treatments in terms of the Q-SURV outcome described at the end of Step 1. Estimates of Q-SURV are obtained for each treatment separately. The mean durations in the relevant states required for Q-SURV are calculated from the areas under the curves as detailed in Step 2. The treatment effect is then estimated from the difference between Q-SURV in the treatment group and that in the control group.

6.2 Markov decision-analytic modelling

Another common approach in health economic modelling is Markov decision-analytic modelling. Instead of deriving mean times in relevant states from survival outcomes, as in partitioned survival, Markov decision-analytic modelling models the transitions between states directly [Sonnenberg and Beck (1993), Briggs and Sculpher (1998), Briggs et al. (2006) and Sun and Faunce (2008)]. It being a form of state transition modelling means it shares many similarities with the multi-state modelling which was the focus of the previous two chapters. Alongside describing the Markov decision-analytic approach this section will emphasis key differences between it and multi-state modelling.

6.2.1 Define the relevant health states

As with partitioned survival, the first step in Markov decision-analytic modelling is to define the health states of interest. Continuing with the illustrative example, Figure 6.3 shows the transition diagram for the Markov decision-analytic model.

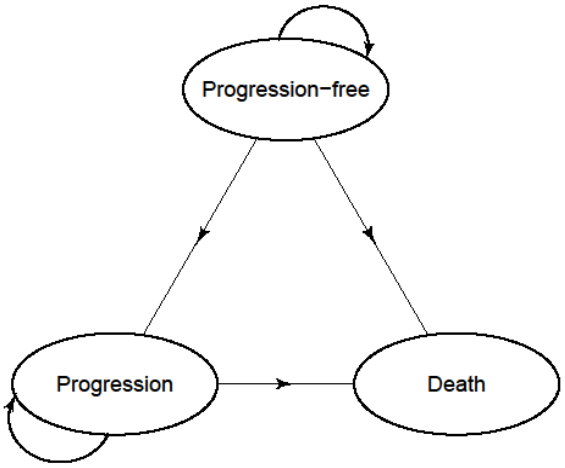


Figure 6.3 Markov decision-analytic model transition diagram

The three states of interest are progression-free, progression and death. Initially, all patients are in the progression-free state. The three transitions progression-free → progression, progression-free → death and progression → death are modelled. Therefore the composite event outcome progression or death, from the progression-free survival outcome used in the partitioned survival approach, is split into the two transitions of progression-free → progression and progression-free → death. There are now two routes to death – either directly or via progression.

In general, there is a decision to be made about which transitions are to be included in the model. This is also the case for multi-state modelling. However, all the transitions that are possible mathematically may not be of interest or make practical sense. For example, as already mentioned in section 4.2 for multi-state models, it is possible for models to have reversible transitions. That is to say once a patient reaches a state it is possible to revert to a previous one. This could be particularly useful for illnesses with a cure/remission phase. However, for instance, it would not make sense for a model to include a transition that involved movement from the death state.

The time horizon, i.e. the finite time frame over which to model, is split into discrete cycles. These cycles need to be sufficiently short to capture changes in clinical outcomes and costs. In practice, models typically use monthly or annual cycles. The circular arrows in Figure 6.3 next to the progression-free and progression states represent patients that stayed within that state for the length of a cycle.

Therefore patients in the progression-free state could either move to the progression state, move direct to the death state without entering the progression state or remain in the progression-free state. Patients in the progression state could either move to the death state or remain in the progression state. Obviously patients who reach the death state cannot move from that state once there, hence the name "absorbing state" given to death. Patients must be in only one of the states at the end of each cycle.

However, the notion of cycles is specific to Markov decision-analytic modelling because it measures time discretely. The use of cycles is not an issue for multi-state modelling because it measures time continuously. That is to say, it uses the individual patient level data and therefore the exact time of transition is known, rather than the typical cohort simulation of Markov decision-analytic modelling over a series of cycles. Therefore, the circular arrows in Figure 6.3 that represent staying in a state within a cycle are only a feature of transition diagrams for Markov-decision analytic modelling and not multi-state modelling.

6.2.2 (Relaxation of) the Markov property

The Markov property – often called the "memoryless property" - is assumed to hold in Markov models. This is the condition that movement from the present state is not dependent on the previous states visited or the length of the visits. Therefore Markov models are said to have no memory of the previous history of patients. (This was previously mentioned in chapter 4 because multi-state models can also be Markov).

However, the Markov property can be relaxed by including tunnel states in the modelling. Figure 6.4 illustrates the use of tunnel states.

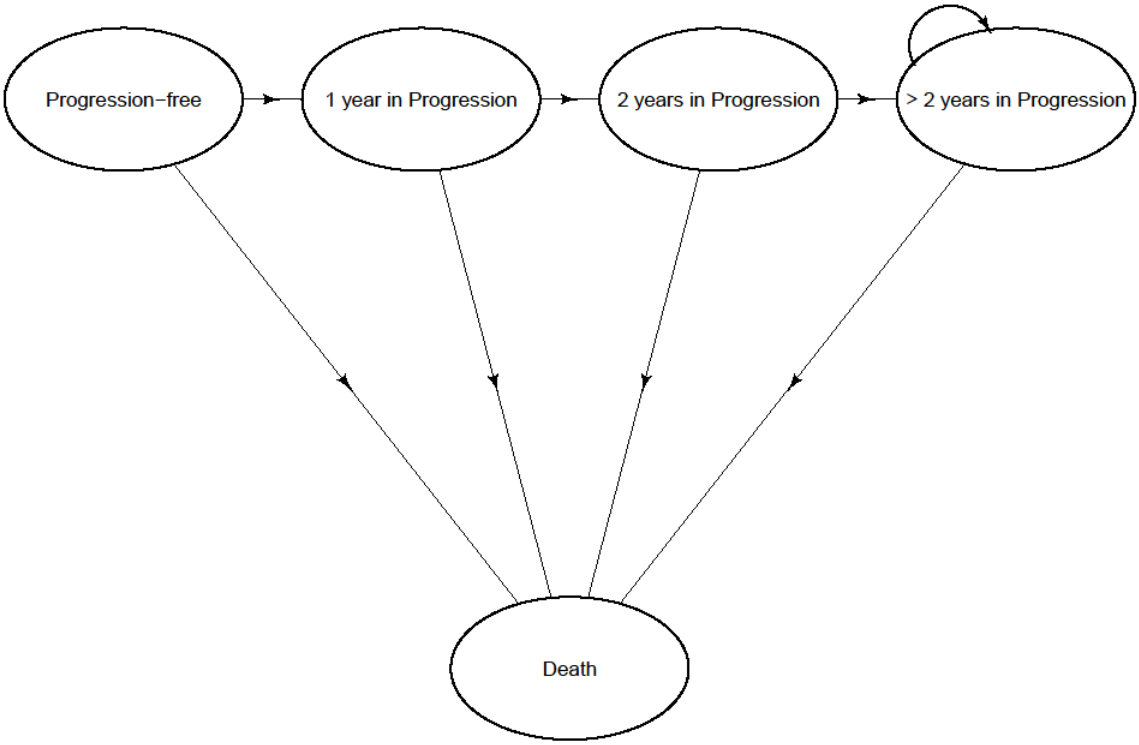


Figure 6.4 Illustration of tunnel states in Markov decision-analytic modelling

A tunnel state is a temporary state that is only occupied for one cycle. The "tunnel" the patients pass through indicates both the state occupied and the number of previous cycles spent in that state. Modelling incorporating tunnel states is known as "semi-Markov", reflecting the relaxation of the Markov property.

In contrast, tunnels states are not needed to build semi-Markov multi-state models. Semi-Markov models in multi-state modelling were described in section 4.4 of chapter 4.

6.2.3 Assignment of transition probabilities/obtaining results

Probabilities of occurrence are assigned to each of the transitions before modelling commences. These can be obtained from several sources including

published literature, empirical modelling and expert opinion. Published literature could include clinical trials, meta-analyses or observational studies. The transition probabilities are stored in an n × n matrix, with n being the number of states. However, only the entries of the matrix representing transitions will be non-zero because, as previously mentioned in sub-section 6.2.1, not all transitions between states will be relevant. In contrast, with multi-state modelling as previously mentioned in chapter 4, transition probabilities are calculated by appropriately combining hazards of individual transitions. Typically, this is after fitting survival regression models for the transition hazards using individual patient level data.

Transition probabilities (Markov decision-analytic modelling) /transition hazards (multi-state modelling) can be constant over time (time homogeneous) or be time-dependent (time inhomogeneous).

Standard approaches for obtaining results from Markov decision-analytic modelling are covered extensively elsewhere and are not described in this thesis. For example, the contribution by Sonnenberg and Beck (1993) provides a succinct summary of the matrix algebra solution, cohort simulation and Monte Carlo simulation.

6.3 Review of multi-state models for health economic modelling in the literature

In recent years, there has been an increase in contributions involving the continuous-time multi-state modelling framework in the health economics literature, and in a health economics context in the more general medical literature. This section summarises and critiques such contributions. The purpose of this review was to explore to what extent the methodology and applications were illustrated and to identify any barriers to adoption of the approach. It also sought to identify areas where multi-state modelling had not been used to its full potential. Due to it being a form of state-transition modelling, multi-state modelling has huge potential as an alternative to conventional spreadsheet-based approaches for cost-effectiveness modelling but it is not widely applied. It is syntax-based providing a transparent record of the analysis and it makes errors

easier to spot. Furthermore, it explicitly allows testing of the Markov property by simply adding a relevant covariate and semi-Markov modelling does not require tunnel states. Of particular note, Markov models can be built in seconds and therefore their creation is a lot less cumbersome and time-consuming.

A search strategy was developed to allow the review to take place. An initial search took place when this research began in 2011. OvidSP was searched for titles, abstracts and keywords with the terms:

(("multi-state" AND "?Markov" AND cost*) OR ("multistate" AND "?Markov" AND cost*) OR ("multi state" AND "?Markov" AND cost*) OR ("illness-death" AND "?Markov" AND cost*) OR ("illness death" AND "?Markov" AND cost*) OR ("disability model*" AND "?Markov" AND cost*) OR (continuous-time AND ?Markov AND cost*) OR (continuous AND time AND ?Markov AND cost*) OR ("multi-state model*" AND cost*) OR ("multistate model*" AND cost*) OR ("multi state model*" AND cost*) OR ("illness-death model*" AND cost*) OR ("illness death model*" AND cost*) OR ("disability model*" AND cost*)).ti,ab,kw.

However, this search has been an ongoing process. Email alerts of new articles were set up using OvidSP and Web of Knowledge. A final OvidSP search using the strategy above was then conducted to search for contributions up to 05/10/2016. Three health economic journals known to have relevant contributions were also specifically targeted – Medical Decision Making, Value in Health and Health Economics. These journals were searched again for “multi-state model*” OR “multistate model*” OR “multi state model*” OR “continuous-time” OR "continuous time"

in titles/abstracts/keywords to capture those contributions that were not found from the other searches.

In addition, email notifications of new content alerts were also set up for these three journals at the beginning of my research period and any relevant contributions were included. I also checked the reference lists of articles I had already reviewed and articles that cited papers that I already knew had contributed to this area of research.

In addition the NICE website (www.nice.org.uk) was searched on 05/10/2016 using the terms:

"multi-state" OR "multi-state Markov" OR "multi-state semi-Markov" OR

"multistate" OR "multistate Markov" OR "multistate semi-Markov" OR

"multi state" OR "multi state Markov" OR "multi state semi-Markov" OR

multi-state OR multistate OR "continuous-time" OR "continuous time"

and found 6 technology appraisals (TAs). However, each of these 6 TAs did not involve continuous-time multi-state modelling, and were excluded for that reason. Furthermore, a search was performed on the National Institute for Health Research (NIHR) Journals Library website www.journalslibrary.nihr.ac.uk on 05/10/2016 within the journal *Health Technology Assessment* specifically, and using the NICE website's NICE Evidence Search facility with source set to NIHR Journals Library, using the search terms:

multi-state OR multistate OR "multi state" OR "multi-state" OR "multistate" OR "continuous-time" OR "continuous time"

and found 7 NIHR Health Technology Assessment Reports. However, 4 of these did not involve multi-state modelling. The other 3 were found in the main OvidSP search but did not involve the continuous-time multi-state modelling framework and were excluded for that reason.

Each of the above search strategies collectively found 27 contributions relevant to continuous-time multi-state modelling (Appendix VIII), which will be discussed in the commentary in the rest of this section. Of note though, the rejection of the contributions from the two sources of health technology assessments highlights there seems to be some inconsistences in terminology with regards to multi-state modelling between the fields of medical statistics/epidemiology and health economics. In the former, the term appears to be used exclusively in the context of continuous-time, unless otherwise explicitly stated, whereas there is a lot more ambiguity in the latter where it appears to be used both in a continuous-time and discrete-time context.

6.3.1 Methodological issues: continuous-time multi-state modelling contrasted with Markov decision-analytic models

It is recognised that, in health economics, the vast majority of state-transition modelling is carried out using Markov decision-analytic models within a discrete-time framework, instead of a continuous-time multi-state modelling approach [e.g. van Rosmalen et al. (2013)]. Section 6.2 described Markov decision-analytic modelling. Briefly, it typically involves following a cohort of patients from an initial health state over a series of discrete cycles, with movement between states based on pre-assigned transition probabilities. van Rosmalen et al. (2013) and Cao et al. (2016) highlight that a half-cycle correction is often undertaken to reduce measurement bias due to the assumption that transitions only occur at discrete times. van Rosmalen et al. (2013) stress that bias can be particularly apparent with long cycles but that the shorter cycle lengths required to overcome the bias can be computationally intensive. However, the continuous-time multi-state modelling framework uses the exact time of transitions and therefore, as stated by Cao et al. (2016), there is no such bias and therefore no need for a half-cycle correction.

Begun et al. (2013) discuss the point that continuous-time models are preferable to discrete-time models, because the former allow for transitions that may have a small probability and therefore cannot be observed during a small time unit. They also emphasise that continuous-time models are more realistic as they allow for state transitions to occur at any moment. A conference proceeding by Tichy (2014) also highlights that a continuous-time framework can be preferable due to overcoming the need for cycles.

In Markov modelling, the Markov “memoryless” property is assumed to hold. This is the condition that movement from the current state does not depend on the previous states visited or the time spent in any previous states. In the discrete-time framework, one way of relaxing the Markov assumption is to incorporate tunnel states, that represent previous cycles spent in a state, into the model. Hawkins et al. (2005) developed an alternative way of representing tunnel states using multi-dimensional transition matrices to incorporate dependency on the previous cycles spent in states. While it was still in the discrete-time framework,

it was however an important contribution demonstrating how to conduct the analysis in the open-source language R, and hence a departure from the reliance on spreadsheet-based packages.

For the calculation of transition probabilities, many of the contributions have commented that, transition probabilities are used as inputs in the discrete-time framework, whilst in the continuous-time framework transition hazards (rates) are used (Welton and Ades (2005), Begun et al. (2013) and van Rosmalen et al. (2013)). Welton and Ades (2005) emphasise that transition rates are commonly used for combining information from different studies. Therefore, the use of transition rates, rather than transition probabilities, provides a natural mechanism to conduct evidence synthesis. Price et al. (2011) stress the need to synthesize evidence from all available trials to assess which is the most effective treatment. In their article, Price et al. (2011) describe how to conduct a meta-analysis to compare different treatments across several trials in a multi-state modelling context. They show how to estimate the treatment effects needed for each of the transitions in the modelling.

6.3.2 Approaches to calculating transition probabilities in continuous-time multi-state modelling

The research describing the methodology of the continuous-time multi-state modelling approach mainly focuses on calculating transition probabilities by solving Kolmogorov's forward differential equations using matrix algebra based around matrix exponentials [Welton and Ades (2005), Begun et al. (2013) and van Rosmalen et al. (2013)]. Welton and Ades (2005) and Begun et al. (2013) each provide worthwhile contributions that demonstrate how the continuous-time multi-state modelling framework can be used with both fully and partially-observable data. Begun et al. (2013) illustrate an approach based on Cox semi-parametric regression. The contribution provides a useful introduction to the concept. However, it does not include code or details of a software package to allow others to adopt the approach. Extending the Cox regression to fully parametric regression would have allowed for extrapolation of survival outcomes that were not observed in all patients. Welton and Ades (2005) present a Bayesian approach to solving the forward equations which has the benefit of

being able to propagate uncertainty through the model, and include WinBUGs code in an appendix. When proposing the continuous-time approach, van Rosmalen et al. (2013) also provide code as an appendix that is based around matrix exponentials. They distinguish multi-state modelling from discrete-event simulation in that, whilst both use a continuous-time framework, the former is deterministic (exact) and the latter is stochastic and therefore subject to simulation errors. Furthermore, van Rosmalen et al. (2013) offer multi-state modelling as an alternative to discrete-event simulation or microsimulation. In particular, they emphasise its advantages over simulation approaches of not requiring dedicated software, needing less computation time and not being subject to simulation error. However, they also acknowledge that the required programming of matrix exponentials for their proposed approach is a drawback compared to the more user-friendly spreadsheets used in the discrete-time Markov decision-analytic modelling.

6.3.3 Some applied uses of multi-state modelling in the health economics literature

The applied use of multi-state modelling for lung cancer by Bongers et al. (2016) adds value to the existing contributions in that it helps to highlight multi-state modelling as an alternative to microsimulation models, albeit in a roundabout way. The article makes use of the `mstate` package in `R` to undertake multi-state modelling. However, rather than using the facility in `mstate` to calculate transition probabilities, it only goes so far as to calculate transition hazards which are then used to inform assumptions in a microsimulation model. Another very recent contribution also carried out multi-state modelling in `R`, but then proceeded to build a conventional Markov decision-analytic model informed in part from the transition probabilities of the multi-state model (Scotland et al. (2016)). Therefore, the implementation in these two contributions fell short of showcasing the full capabilities of `mstate` for carrying out multi-state modelling. Anything that can be built in Excel can be coded in statistics software. Therefore, there is scope to avoid the two-step process and perform all the health economic modelling using the same software.

Another contribution that made partial use of `mstate` was the piece by Egger et al. (2016). They created a multi-state model to investigate the effect of adherence to clinical guidelines on outcomes for elderly patients admitted to hospital with community-acquired pneumonia. The model had 14 cycles. However, since each cycle represented a day, it could be considered to be using a continuous-time framework. The authors used `mstate` to derive transition probabilities, for a given covariate profile, from Cox cause-specific hazard modelling. These transition probabilities were then transformed into beta distributions in the model using the estimates and standard errors from the Cox modelling. However, it was not clear how this transformation occurred. An alternative approach could have been to carry out regressions assuming the hazards followed parametric distributions, and using the hazards from these in conjunction with `mstate` to calculate state occupancy probabilities. The research in this thesis will involve an adaption of `mstate` to allow such parametric regressions instead of Cox modelling.

Other applied uses of continuous-time multi-state modelling have also emerged in the literature in the areas of heart failure (Cao et al., 2013), coronary artery disease (Asaria et al., 2016), diabetes (Jensen et al., 2014), kidney disease (Begun et al., 2013), mRSA (De Angelis et al., 2011), HIV (Le Pen et al. (2001)) and colorectal cancer (Castelli et al., 2007).

Cao et al. (2013) provide a useful illustration of how continuous-time multi-state modelling can be used for cost-effectiveness analysis. Their contribution is from the perspective of producers of health technology as decision makers, rather than a societal perspective. As such, instead of focusing on Cost per quality-adjusted life year, their analysis concentrated on “commerical headroom”. This is identical to the net monetary benefit which is a standard concept described in (Drummond et al., 2005). They use the `msm` package in `R` alongside bootstrapping. Furthermore, their contribution also includes probability sensitivity analysis.

Asaria et al. (2016) illustrate an applied use of a state-transition model which is more akin to the continuous-time multi-state modelling framework, than a

discrete-time Markov decision-analytical approach. However, the paper did not model the transitions simultaneously using the individual patient level data in the usual manner of multi-state modelling. Instead, they developed risk equations for each of the transitions separately and then later incorporated them into a competing risks framework. Therefore, it was not clear whether their post-equation competing risk adjustment appropriately took account of the competing risks scenario which was present.

The demonstration of an applied use of multi-state modelling by Jensen et al. (2014) was a quite complex illustration. The authors built a multi-state model of adherence to long term medication in those with Type 2 diabetes. The assignment of states and transitions between states involved an algorithm which was not easy to follow. While the authors state they used SAS for their analysis, they did not provide any code in an appendix to allow others to adopt a similar approach. This was unfortunate especially considering its complexity.

6.3.4 Multi-state modelling applied to infections

Several authors provide helpful demonstrations that showcase how multi-state modelling can be used to estimate the mean length of stay in various health states, which could then be used in the calculation of costs for hospital-acquired infections in particular [De Angelis et al. (2011), Stewardson et al. (2012) Stewardson et al. (2013), Stewardson et al. (2015), Stevens et al. (2015), Ndir et al. (2016a) and Ndir et al. (2016b)]. This is in contrast to the more typical use of calculating mean (quality-adjusted) life years in relevant health states, which is particularly relevant for chronic and/or progressive diseases. They are a useful addition to the literature, because as De Angelis et al. (2011) highlight, the mean length of stay approach has potential for acute, non-recurrable diseases.

In a follow-up to the work of De Angelis et al. (2011), some of the same authors [Macedo-Vinas et al. (2011) and Macedo-Vinas et al. (2013)] use the excess length of stay calculated to derive excess bed-days which were then used to estimate attributable costs. The work by De Angelis et al. (2011) and Macedo-Vinas et al. (2013) studied MRSA. In other research, which included De Angelis as an author, a similar approach was used to investigate excess length of stay with

bloodstream infections in adults in a Swiss hospital by Stewardson et al. (2012) and Stewardson et al. (2013) and in wider Europe by Stewardson et al. (2015). A different group of authors also used the approach to investigate bloodstream infections, but in a paediatric setting in Senegal (Ndir et al. (2016a)). Furthermore, Ndir et al. (2016b) used the approach to investigate infections caused by extended-spectrum beta-lactamases producing Enterobacteriaceae.

In other research, Stevens et al. (2015) used the approach to investigate excess length of stay attributable to Clostridium difficile infection and Arefian et al. (2016) used it to investigate excess length of stay and costs due to health care-associated infections in general. All of the contributions in this section have advocated the use of multi-state modelling because it explicitly takes into account the time of infection by including infection as a state. Furthermore, Mitchell and Gardner (2012) carried out a review of methods used to investigate length of stay for Clostridium difficile infection. However, the methods used were prior to the uptake of multi-state modelling and therefore introduced bias due to not taking into account the timing of the infection. The authors concluded that multi-state modelling was a more appropriate method, and encouraged collection of the timing of the events that would constitute states in the model. In a related review, Nelson et al. (2015) investigated the magnitude of bias introduced by failing to take into account the timing of the infection. Again these authors recommended using an appropriate method such as multi-state modelling or matching. In another related piece, Barnett et al. ((2016) demonstrated the time-dependency bias when not using multi-state modelling by comparing such a model to a generalised linear model. Shi et al. (2014) investigated the time-dependency bias still further. The authors took into account patient heterogeneity using an approach based on centred residuals from a generalised linear model which were incorporated into a multi-state model. These contributions have been useful in that they have highlighted an aspect of bias that can occur and demonstrated how multi-state modelling can provide a solution.

6.3.5 Multi-state modelling applied to HIV

Le Pen et al. ((2001) 's research is concerned with the effect of highly active antiretroviral therapy in HIV patients. It is an interesting contribution because it raises awareness of the inclusion of irreversible states in multi-state modelling i.e. that patients can revert back to states they have been in previously. Therefore it highlights an alternative approach to the more standard model that normally follows the natural progression of a disease. The states are based on all possible combinations of CD4 cell count and viral load, each of which are grouped into three categories, plus a death state. Therefore there are ten states in the model. The paper makes a valiant attempt to model both progression of disease and improvement in prognosis by modelling transitions between these nine + one health states. However the model is arguably too complex because the model starts with the nine states as initial states. There are too few patients in many of the states even at the initial stage, never mind to follow through to the end of their study. Several other papers used a similar model for HIV based on CD4 cell count and HIV RNA levels (Chaudhary et al. (2010) and Erdesz et al. (2010)) and CD4 cell count and HIV-I viral load (Athanasakis et al., 2014) and therefore also have the same complexity issues.

6.3.6 Multi-state modelling contrasted with partitioned survival

During a conference, Rael et al. ((2016) raised awareness of the advantages of multi-state modelling/state-transition modelling over partitioned survival. They emphasised that modelling the transitions between the states progression-free, post-progression and death can answer more research questions than separate analyses of progression-free survival and overall survival. In particular, they highlighted that treatment effects on pre- and post-progression survival, and whether the observed overall survival benefit is driven solely by the progression-free survival can both by investigated with such models. However, the contribution advocated using patient-level simulations to calculate state occupancy probabilities alongside the multi-state modelling. It failed to recognise that all analysis can be incorporated under the one multi-state modelling framework without the need for a separate simulation model. In addition, it states that a limitation of multi-state modelling is that the data

must be mature enough to estimate the transitions. They could have also noted that maturity is also a requirement of analysis using separate modelling of progression-free survival and overall survival. Lastly, it maintains that software for multi-state modelling is less well developed. A major component of the research carried out during this PhD will address this final point.

6.3.7 Misunderstandings in the literature

van Rosmalen et al. (2013) mention that multi-state modelling is Markov when an exponential distribution is used to model the transition hazards and semi-Markov otherwise. This is a different interpretation to Putter et al. (2007), as explained in section 4.2 of chapter 4, which differentiate Markov from semi-Markov by using the way time is measured: from first entering the initial state for the Markov approach and by resetting the clock back to zero when patients enter a new state with the semi-Markov approach. The contribution by van Rosmalen et al. (2013) fails to recognise that it is possible for the process in both a Markov and semi-Markov model to be time-homogeneous or time-inhomogeneous. Time homogeneous processes involve the use of the exponential distribution with a hazard that is constant over time. In contrast, in a time-inhomogeneous process the hazard varies over time. Parametric distributions other than exponential can be used in these situations. However this does not appear to be well known in the literature, and some incorrect use of terminology has resulted. There are instances where papers describe time-homogeneous processes but then proceed to use distributions other than the exponential; for example, Cao et al. (2016) and Castelli et al. (2007) using log-normal and Weibull distributions respectively.

Another issue that is not always recognised relates to competing risks which often occur in state-transition models. When faced with competing risks and the modelling of hazards, it is not always appropriate to use AICs for model selection. This was discussed previously in section 2.7.2 of chapter 2. However some of the research has used AICs to decide between parametric distributions for the hazard (Cao et al. (2016) and Asaria et al. (2016)). Another important point related to competing risks is apparent in the work by Clarke et al. (2014). The article uses Kaplan-Meier curves as a basis for the transition probabilities

used in their model. It acknowledges that this is likely to introduce bias although it does not specifically attribute it to the censoring of competing risks. Instead, the article acknowledges they may be informative censoring resulting in inflated survival, and therefore it would appear that it states the bias in the wrong direction.

6.3.8 Alternative approaches

Cao et al. (2016) propose an alternative approach to continuous-time multi-state modelling called vertical modelling. However, it does appear to be overly complex. The simulation is difficult to understand and follow, for which no code is provided. It requires separate models to be built for the sojourn times and the future state probabilities. In addition, possible nonlinearity between the two needs to be checked, with fractional polynomials used for that purpose in the illustration given in the paper. The article argues that the vertical modelling approach has advantages over the pattern mixture formulation and cause-specific hazards approaches. From the description of the pattern mixture formulation, the approach suffers from interpretation problems and seems to be unrealistic as it involves conditioning on the future. The paper also argues that the cause-specific hazards approach has interpretation problems as it cannot be directly expressed as probabilities, and therefore would make discussions with clinicians difficult. However I counter this argument as the complexity of the vertical modelling approach would make the communication of results with clinicians problematic. The perceived difficulty with the cause-specific approach can be easily overcome. The facility to calculate state occupancy probabilities using that approach exists in procedures such as those in the `mstate` package in `R`.

Kim and Thompson (2011) ‘s contribution focuses on competing risks specifically rather than multi-state modelling in general. The method proposed in the paper is a continuous-time type approach as the calculations were restricted to 6-month intervals, rather than measuring time continuously. The justification given for this is that it makes it compatible with methods for costs and discounting. However, these methods seem to be just as compatible in a continuous sense. The method involves estimating Life Years gained and cost-

effectiveness using cause-specific, rather than all-cause, mortality. The motivation for this is that it is not always useful to use all-cause mortality for Life Years gained, because only a few causes of death may affect the interventions studied. The paper illustrates, perhaps counter-intuitively, that greater power can be achieved to detect differences in cause-specific mortality than in all-cause mortality. It also finds greater precision, in terms of smaller standard errors, with its estimator compared to using all-cause mortality. However, upon extrapolating the results over a life-time horizon it is found that results using cause-specific and all-cause mortality are similar. In addition, the loss of precision over the period of the study for all-cause mortality is not evident upon extrapolation. The paper concludes that a long-term all-cause mortality outcome can be used, but only when it is appropriate to assume that rates of other-cause mortality are the same across groups. Otherwise, it stipulates a cause-specific approach should be used. It cautions against using the usual competing risks cumulative incidence approach when it is calculated separately for each group. The proposed approach is interesting but it is not likely to be widely-applied in practice. This is because it involves naïve Kaplan-Meier estimates, which is in contrast to all the established competing risks literature that caution against this due to the bias it can introduce.

6.3.9 Summary of the review

This review of the literature has shown limited, but growing, awareness of the continuous-time multi-state modelling framework in the literature for health economics where it has huge potential. However it has highlighted that applications of the approach are scarce, with the field acknowledging that use of the continuous-time framework greatly lags behind its discrete-time counterpart. To my knowledge, there has been no articles illustrating continuous-time multi-state modelling as an alternative by comparing it to commonly-used approaches in cost-effectiveness analysis. Therefore, Chapter 7 of this thesis, and its accompanying publication (Williams et al., 2017b), presents an empirical analysis comparing multi-state modelling to the two more widely-used approaches of partitioned survival and Markov decision-analytic modelling.

6.3.9.1 Code to encourage adoption of the approach

Some of the research has provided code and/or references to software packages (Cao et al. (2013), Welton and Ades (2005), van Rosmalen et al. (2013), Bongers et al. (2016) and Asaria et al. (2016)), overcoming a barrier that previously prevented modellers easily adopting similar approaches. This should help encourage the consideration of the continuous-time multi-state modelling approach, where individual patient level data is available. Such data was available to Clarke et al. (2014) and Hettle et al. (2012), but they opted for a discrete-time approach instead.

However, it was also apparent from the literature that one of the main perceived barriers to adopting the continuous-time multi-state modelling approach was the implementation of the algebra involved together with a lack of readily available software. In order to fill this gap, I created a webpage with customisable `R` functions to allow users to carry out full cost-effectiveness analyses. I also published an accompanying tutorial paper in the Medical Decision Making journal, a journal popular in the health economics field (Williams et al., 2017a). This paper guides the users through the steps involved in a cost-effectiveness analysis using multi-state modelling and encourages adoption of the approach. It demonstrates how to use multi-state modelling to calculate mean (quality-adjusted) life years gained and incremental cost-effectiveness ratios. It also illustrates how the approach can be used in deterministic and probabilistic sensitivity analyses. In using a syntax-based approach, it benefits from the transparency of the code used and analysis that is all contained in one file. Therefore, errors are arguably easier to spot than in spreadsheet packages, where cells can be accidentally changed without the analyst's knowledge. The calculation of the transition probabilities for Markov multi-state modelling was not implemented using matrix exponentials. Instead, an equivalent representation using product integrals, described in more detail in section 4.3 of chapter 4 was used which is less computationally intensive and more intuitive than matrix exponentials. The simulation approach used for semi-Markov multi-state modelling was covered in section 4.4.

6.4 Assessment of fit over the observed period of a study

This section discusses issues worth consideration when assessing the fit to the data observed in a trial or other study. The need to test the proportional hazards (PH) assumption, when considering fitting such models is emphasised [Grieve et al. (2013), Latimer (2013), Bagust and Beale (2014), Latimer (2014)]. One method recommended for this purpose by Latimer (2013) is to plot the log-cumulative hazards versus log time, also known as a log-log plot, for each treatment group. If the lines are reasonably parallel there is no suggestion of a violation of the proportional hazards assumption. Data that follows an exponential distribution will produce a plot that has a straight line with slope 1, whereas data from a Weibull distribution will be represented as a straight line with a slope other than 1. The plot is useful for assessing the hazards at the earlier times rather than the later times when there are fewer events anyway (Latimer, 2014).

Bagust and Beale (2014) prefer a plot of the cumulative hazards versus time to assess whether the proportional hazard assumption holds or not. It is useful for visualising the hazards long-term, rather than in the initial time period during which there can often be a higher hazard which takes this initial period to settle down. Bagust and Beale (2014) argue it is easier to distinguish an exponential distribution from a Weibull distribution with this plot because, rather than both being represented as straight lines as they are in the log-log plot, the cumulative hazard plot can show non-linear trends indicating hazards that either increase or decrease. This can also be seen as a weakness of this plot however because, as Latimer (2014) points out, they are more difficult to assess visually because proportional hazards are represented by lines that diverge at a constant rate rather than parallel lines.

When considering fitting models in the accelerated failure time (AFT) framework (described later in section 6.5), it is recommended to test whether AFT is an appropriate fit by plotting Q-Q plots (Bradburn et al. (2003)). This plots the percentiles/quantiles of the survival times in one treatment group versus the corresponding information in the other treatment group. They should show as a

straight line that passes through the origin with slope equal to the acceleration factor of the model. An alternative approach for assessing whether AFT and/or PH models provide appropriate fits are Cox-Snell residual plots. These check that the survival times follow an exponential distribution with hazard equal to 1, which is indicative of a good-fitting model. The plots show the observed and fitted cumulative hazard and should show a straight line through the origin with a slope equal to one. Cox-Snell residual plots can be used for both AFT and PH models, allowing a direct comparison of such models.

If there is evidence from the plots described above to suggest a violation of the PH or AFT assumption, as appropriate, for the treatment covariate, or indeed any covariate, then such models should not be fitted. Instead, separate fits can be used for each treatment (Bagust and Beale, 2014), although this no longer allows direct estimation of a treatment effect. When the PH/AFT assumption is assumed to hold, separate fits for each treatment can also be carried out as a sensitivity analysis – to check there are no major discrepancies in results between them and the PH/AFT model with treatment as a covariate. It is worth noting however that (Cox) PH models are reasonably robust to minor deviations from the PH assumption. If an inspection of the Kaplan-Meier curves does not show evidence of obvious divergence, convergence or crossing of the curves, then one could assume that hazards are proportional. The hazard ratio can be thought of as the “average” effect across the period of observation.

Visual assessments of the fit to the observed data, such as comparing the predicted survival with the Kaplan-Meier survival estimate and the plots described above, are useful for comparing fits with different distributional assumptions, such as those described later in section 6.5. It can also be worthwhile calculating AICs (Akaike, 1974) or BICs (Schwarz, 1978) because they do not have the subjectivity of the visual inspection of plots. However, assessment of fit should not be based solely on one of these but rather they should both be considered. In addition, the fit to the observed data is only one consideration if extrapolation beyond the observed period is required (extrapolation will be discussed in the next section). In that situation, AICs/BICs can be of limited value because they are only concerned with the fit over the

observed period and not on how reasonable the extrapolation looks. Furthermore, caution is advised when using AICs/BICs with a competing risks scenario (common in state-transition modelling) and the modelling of hazards, as previously discussed in section 2.7.2.

6.5 Extrapolation of outcomes for use in economic evaluations

6.5.1 Motivation for extrapolation

It is often the case that a study does not have sufficient follow-up to observe the outcomes in every patient, e.g. in the example illustrated in Figure 6.2 not every patient was observed long enough to reach progression and/or die. However, in health economics, a lifetime perspective is often needed because an estimate of the mean time in relevant states is needed, to be used with the mean costs, in the calculation of cost-effectiveness measures. Therefore extrapolation of survival beyond the observed period is often necessary. This has received much attention in the literature recently, e.g. Tappenden et al. (2006), Connock et al. (2011) ,Grieve et al. (2013), Latimer (2013), Stevens and Longson (2013), Bagust and Beale (2014) and Latimer (2014). Several of these papers have commented on the need to achieve a balance of both a good fit to the observed data and a sensible and clinically plausible extrapolation.

A useful first step is to use a parametric fit to the observed (Kaplan-Meier) survival curve and extrapolate it until it reaches the desired time horizon. There is a general consensus to consider all the “standard” distributions – i.e. exponential, Weibull, Gompertz, log normal, log-logistic and generalised gamma. Focus should be on trying to achieve sensible extrapolation, in the sense that it should be clinically plausible and/or compatible with external data. If the extended predictions do not adequately represent zero, i.e a lifetime perspective, at the end of the time horizon, then an alternative is to start the extrapolation from the tail of the observed Kaplan-Meier curve. Tappenden et al. (2006) outline an approach based on fitting a linear regression to the Kaplan-Meier curve and then back-transforming to the equivalent using parametric

regression. It involves rearranging the survival function (or function thereof) into a linear function of time that can be used to fit a linear regression. The resultant coefficients are that used in the original parametric survival function after transformation into that scale. This approach could be used to start extrapolation from the tail of the observed curve. However the choice of suitable points will be somewhat arbitrary and therefore sensitivity analysis for this choice should be undertaken. In addition, the tail of the observed curve is where estimates will be least precise. The fact that the confidence interval around the Kaplan-Meier survival estimates demonstrate their precision can be used as an aid to help decide when to start extrapolation. Another sensitivity analysis that is recommended by Grieve et al. (2013) is considering different treatment effects long-term. Specifically, the authors advocate considering no treatment effect in the unobserved period, a reduction in the treatment effect over time and a treatment effect from the observed period that persists over the unobserved period.

6.5.2 Conventional approach: parametric regression

This section summarises some considerations with regards to parametric survival regression as it pertains to extrapolation. The conventional approach for extrapolation of survival is to use parametric survival regression – either fitting a model to all the observed data, or starting the fit from some point on the tail of the Kaplan-Meier curve. Two common ways to fit such models are proportional hazards and accelerated failure time models. In a proportional hazards model, the hazard function can be written as

$$h(t) = h_0(t) \exp(\beta x)$$

where x is the vector of covariates and β is the vector of regression coefficients. $h_0(t)$ is the baseline hazard and is assumed to be parametric. Common parametric distributions used for proportional hazards models are Weibull, exponential and Gompertz.

In an accelerated failure time model, the natural logarithm of the survival time is modelled as a linear function of the covariates i.e.

$$\log(t) = \beta x + \varepsilon$$

where ε is the error term and β and x are as for the proportional hazards model. The distribution used for the error term determines the type of model. Common accelerated failure time models are log normal, log-logistic and generalised gamma. The Weibull can also be expressed as an accelerated failure time model.

Table 6.1 shows the survival and hazard functions for six standard distributions used in health economic modelling. An exponential model is time homogeneous, (i.e. the hazard rate is constant), whereas the other distributions are time inhomogeneous.

For each distribution there is an equation for the survival and hazard functions, allowing survival to be extended to the desired time horizon.

Table 6.1 Common parametric distributions

Distribution	Survival	Hazard $h(t)$ or density $f(t)$, where appropriate
Weibull	$S(t) = \exp(-\lambda t^{\gamma})$	$h(t) = \gamma\lambda t^{\gamma-1}$
	where $\lambda_j = exp(x_j\ \beta)$ and γ is estimated from the data	
exponential	$S(t) = exp(-\lambda t)$	$h(t) = \lambda$
	where $\lambda_j = exp(x_j\ \beta)$	
Gompertz	$S(t) = exp\{-\lambda\gamma^{-1}(e^{\lambda t} - 1)\}$	$h(t) = \lambda\ exp(\gamma t)$
	where $\lambda_j = exp(x_j\ \beta)$ and γ is estimated from the data	
log normal	$S(t) = 1 - \phi\left\{\frac{log(\ t) - \ \mu}{\sigma}\right\}$	$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} exp\left[\frac{-1}{2\sigma^2}\ \{log(t) - \mu\}^2\right]$
	Where $\phi(z)$ is the standard normal cumulative distribution function, $\mu_j = x_j\beta$ and σ is estimated from the data	
log-logistic	$S(t) = \frac{1}{1 + (\lambda t^{1/\gamma})}$	$f(t) = \frac{\lambda^{1/\gamma} t^{1/\gamma-1}}{\gamma\ \{1 + (\lambda t)^{1/\gamma}\}^2}$
	where $\lambda_j = exp(-x_j\ \beta)$ and γ is estimated from the data	
generalised gamma	$S(t) = 1 - I(\gamma, u)$ if κ > 0 $S(t) = 1 - \phi(z)$ if κ = 0 $S(t) = I(\gamma, u)$ if κ < 0	$f(t) = \frac{\gamma^{\gamma}}{\sigma t\sqrt{\gamma}\,\Gamma(\gamma)}\ \ exp(z\sqrt{\gamma} - u)$ if κ ≠ 0 $f(t) = \frac{1}{\sigma t\sqrt{2\pi}} exp\left[\frac{-z^2}{2}\right]$ if κ = 0
	where $I(\gamma, u)$ is the incomplete gamma function, $\gamma = \lvert\kappa\rvert^{-2}$, $u = \gamma\ exp(\lvert\kappa\rvert z)$, z = sign(κ){log(t)- μ }/ σ , $\mu_j = x_j\ \beta$, $\phi(z)$ is the standard normal cumulative distribution function and κ and σ are estimated from the data	

6.5.3 Alternative approaches to extrapolation

As previously mentioned, it is recommended to consider a range of different distributions. Several alternative, more complex, approaches can also be explored; particularly if the standard distributions do not provide an adequate fit, either to the complete observed data or starting from a point on the tail of the Kaplan-Meier curve. Examples include flexible parametric methods involving splines (Andersson et al., 2013), using less standard distributions such as the generalised F (Jackson et al., 2010) or poly-Weibull (Demiris et al., 2015) and using a Bayesian semi-parametric approach with piecewise-constant hazards (Jackson et al., 2010).

Chapter 7 Comparing Markov decision-analytic modelling, partitioned survival and multi-state modelling

7.1 Introduction

This chapter presents an analysis comparing three different approaches to estimating benefit (survival) in a health economics context. It was motivated by an existing Markov decision-analytic model which was created for a manufacturer's submission to NICE for a technology appraisal. However the Evidence Review Group (ERG), working on behalf of NICE, had some concerns about the model and initiated their own sensitivity analysis. Both the original model and the additional analysis to address the ERG's concerns highlighted that there was scope to consider other approaches. Therefore, this chapter uses the manufacturer's economic model as a basis for a case study that compared it to two alternative approaches. These used the individual patient level trial data that was the main source underlying the model. Specifically, the two approaches compared to the existing Markov-decision analytic model were partitioned survival and multi-state modelling. As such, this chapter provides an empirical analysis that complements earlier parts of this thesis that provided a more methodology background to Markov decision-analytic modelling (section 6.2), partitioned survival (section 6.1) and multi-state modelling (chapter 4).

The extrapolations and choice of time horizon used in the comparison of the approaches did not benefit substantially from being informed by external data. This was primarily because the approaches were contrasted with a manufacturer's existing decision-analytic model, and used the same time horizon as that model for comparison purposes. There was no attempt to vary the time horizon, or the extrapolation to this target horizon, and assess their clinical plausibility. This limitation was due to the model building process being carried out by another party, and therefore having no access to the resources and clinical expertise that team would have had. The extrapolations were

instead carried out by making use of parametric survival regressions using six standard distributions. Using external data to strengthen the clinical plausibility of extrapolations, particularly combining randomised clinical trial data with longer-term follow-up in observational routine data, is discussed in chapter 8.

Section 7.2 introduces the decision problem the manufacturer's economic model was designed to address alongside summarising the structure of the model itself. It then continues with outlining some concerns the Evidence Review Group had with the model. This highlighted that there was scope to consider alternative approaches and hence provided the motivation for the rest of this chapter. Since the alternative approaches considered used the individual patient level data from a trial, section 7.3 briefly describes this trial data to put the analysis that follows into context. Next, section 7.4 demonstrates the partitioned survival approach. The chapter then continues with section 7.5 that presents an analysis of the trial data using the multi-state modelling approach. The results of the three approaches are then compared in section 7.6. Sensitivity analyses, for the multi-state modelling specifically, are presented in section 7.7. Then, section 7.8 provides a direct comparison of the Markov decision-analytic modelling with the multi-state modelling. The purpose of this is to demonstrate that multi-state modelling can provide an alternative way of implementing a decision-analytic model, because when the same assumptions are made with each, the results compare well. Finally, the chapter ends with a summary of the results and a discussion in section 7.9.

7.2 Description of the decision problem and the existing Markov decision-analytic model

This chapter presents an empirical analysis using a case study in a health economics context. The case study was based around the economic model created by a manufacturer (Roche, 2008) in their submission to NICE for the specific technology appraisal TA174 (NICE, 2009). This section briefly describes the decision problem that model was designed to address, the structure of the model and the assumptions made by the manufacturer with regards to the transition probabilities.

7.2.1 The decision problem

The decision problem focused on evaluating the clinical and cost-effectiveness of rituximab in combination with chemotherapy for the first-line treatment of chronic lymphocytic leukaemia. As such, the population under consideration was patients with previously untreated chronic lymphocytic leukaemia. Furthermore, the intervention studied was rituximab combined with fludarabine and cyclophosphamide and the comparator was fludarabine and cyclophosphamide alone. In addition, the main outcome to be evaluated by the economic model was the incremental cost per quality-adjusted life year, a standard approach in health economic modelling. Costs and benefits were discounted at a rate of 3.5% annually. Finally, the model took a lifetime perspective with a time horizon of 15 years. The manufacturer's justification for this length of horizon was that only 1.3% of the cohort were estimated to survive beyond 15 years [Roche (2008): p109]. Also, the report by the Evidence Review Group stated that advice given to them thought this to be an appropriate time frame (PenTAG, 2009).

7.2.2 Structure of the model

The main source of data for the model was the CLL-8 trial (Hallek et al., 2010). This trial reported on the outcomes progression-free survival and overall survival for each patient. Data on post-progression survival was also available. This allowed focus to be on the three states progression-free, progression and death & the transitions between them. Specifically, the three transitions that were modelled were progression-free → progression, progression-free → death and progression → death. A state transition diagram for this model was previously presented in Figure 6.3 on page 126. Furthermore, time was measured in monthly cycles.

The manufacturers took the usual approach in Markov decision-analytic modelling of assigning probabilities of occurrence to each of the transitions before modelling started. The assumptions made by the manufacturer for each of the transitions were as follows:

- progression → death

A monthly probability of 0.0405 was used, the same for each arm. It was based on an assumption of a constant death rate that was derived from the inverse of the mean of 24.1791 months from the Kaplan-Meier estimate of post-progression survival.

- progression-free → death

the observed rate of death whilst progression-free, or an age-specific background mortality rate, whichever was largest. The observed monthly probability of death whilst progression-free was 0.0012 and 0.00139 in the RFC and FC arms respectively.

- progression-free → progression

This was calculated by adding together the probability for progression-free → death and the probability of staying in the progression-free state, and then subtracting the result from one. The probability of staying in the progression-free state was based on a Weibull regression fitted to the observed progression-free survival data that was then extrapolated to 15 years.

Of particular note was the assumption for the probability of progression → death – specifically because it was the same for each treatment arm.

7.2.3 Sensitivity analysis initiated by the Evidence Review Group

The three leftmost columns of numbers in Table 7.1 [adapted from PenTAG (2009)] show the original Mean Life Years and QALY results submitted by the manufacturer. However the Evidence Review Group (ERG), working on behalf of NICE, had some concerns about the model presented by the manufacturer (PenTAG, 2009). In particular, the ERG were concerned about the overall survival gain that resulted from the model because this was not demonstrated empirically. Consequently, the ERG asked the manufacturer to carry out two sensitivity analyses to remove the benefit in overall survival. Table 7.1 also shows these two sensitivity analyses initiated by the ERG. The first sensitivity analysis removed the overall survival benefit by increasing the probability of

death in the progressed state in the RFC arm only by 315%. The second sensitivity analysis removed the benefit by decreasing the probability of death in the progressed state in the FC arm only to 57.4% of what it was originally.

With both sensitivity analyses, the mean QALYs gained reduced to 0.24 from the 0.88 in the manufacturer's original submission.

	Manufacturer's Original Submission			Sensitivity Analysis 1			Sensitivity Analysis 2		
	RFC	FC	Incre-mental	RFC	FC	Incre-mental	RFC	FC	Incre-mental
Mean Life Years	5.73	4.65	1.07	4.66	4.65	0.00	5.73	5.72	0.00
Mean Life Years in PFS	4.11	2.93	1.18	4.11	2.93	1.18	4.11	2.93	1.18
Mean Life Years in Progression	1.62	1.73	-0.11	0.55	1.73	-1.18	1.62	2.80	-1.18
Mean QALYs	4.26	3.38	0.88	3.62	3.38	0.24	4.26	4.02	0.24
Mean QALY in PFS	3.29	2.34	0.95	3.29	2.34	0.95	3.29	2.34	0.95
Mean QALY in Progression	0.97	1.04	-0.07	0.33	1.04	-0.71	0.97	1.68	-0.71

Table 7.1 Mean Life Years/QALYs for time horizon of 15 years: Sensitivity analyses initiated by the ERG to remove the benefit in overall survival [adapted from PenTAG (2009)]

Table 7.2 shows the Costs per Life Years/QALYs gained from the manufacturer's submission and the two sensitivity analyses initiated by the ERG (PenTAG, 2009).

It can be seen that, due to the reductions in QALY gained in the sensitivity analyses, the cost per QALYs had increased to just above the £30,000 per QALY gained willingness-to-pay threshold.

	Manufacturer's Original Submission			Sensitivity Analysis 1			Sensitivity Analysis 2		
	RFC	FC	Incre-mental	RFC	FC	Incre-mental	RFC	FC	Incre-mental
Mean Life Years	5.73	4.65	1.07	4.66	4.65	0.00	5.73	5.72	0.00
Mean QALYs	4.26	3.38	0.88	3.62	3.38	0.24	4.26	4.02	0.24
Mean Total Cost	£25,595	£13,978	£11,617	£21,204	£13,978	£7,226	£25,595	£18,367	£ 7,228
Cost per Life Year Gained			**£10,825**			**£ 3,473,529**			**£2,756,887**
Cost per QALY Gained			**£13,189**			**£30,336**			**£30,304**

Table 7.2 Cost per Life Year gained & Cost per QALY gained: manufacturers submission and ERG sensitivity analyses [adapted from PenTAG (2009)]

Table 7.3 shows what the changes in the probability of death after progression in the two sensitivity analyses initiated by the ERG equate to in terms of hazard ratios (before discounting).

	Sensitivity analysis 1	Sensitivity analysis 2
Change to probability of death in progression in RFC arm	315% increase	no change
Change to probability of death in progression in FC arm	no change	57.4% of original
Probability of death in progression in RFC arm	0.168	0.041
Probability of death in progression in FC arm	0.041	0.023
Probability of death in progression in RFC arm expressed as a rate	0.184	0.041
Probability of death in progression in FC arm expressed as a rate	0.041	0.024
Hazard ratio	**4.451**	**1.758**

Table 7.3 ERG sensitivity analyses expressed as hazard ratios

However, fitting a Cox regression to the observed data resulted in a hazard ratio of 1.364. Therefore, the assumptions made in the sensitivity analyses initiated by the ERG resulted in larger effects than were observed in the data, especially so in sensitivity analysis 2. It was the sensitivity analyses initiated by the ERG that motivated the exploration of the alternative approaches presented in the rest of this chapter.

7.3 Description of Rituximab clinical trial dataset

So far in this chapter, the analyses discussed were carried out by the manufacturer or the ERG (with the exception of those presented in Table 7.3). The rest of this chapter will focus on modelling approaches, wholly carried out by myself as part of the research conducted for this PhD, that used the individual patient level data. Consequently, this section summarises this trial data to provide background.

As mentioned earlier in this chapter, the basis for the case study used for illustration was an economic model in a manufacturer's submission (Roche, 2008) to NICE for the specific technology appraisal TA174 (NICE, 2009). The CLL-8 trial (Hallek et al., 2010) was the main source of data used to populate the model. For comparison, the partitioned survival approach (section 7.4) and multi-state modelling (section 7.5) presented later in this chapter used the data from this trial directly.

Table 7.4 summarises the number of events of relevance for the 408 patients in the RFC arm and the 409 patients in the FC arm.

Patients were in the trial for up to 4 years and not all of them were observed to the end of their life. This meant extrapolation of survival was necessary to gain a lifetime perspective. As already explained in section 7.2, a time horizon of 15 years was used in the manufacturer's economic model. To help with the comparison with the two approaches wholly using the individual patient level data from the trial, a time horizon of 15 years was also used with each of them.

	RFC (n=408)	FC (n=409)
progression	106	148
death after progression	23	27
death without progression	21	26

Table 7.4 Summary of number of events in the CLL-8 trial

7.4 Partitioned survival

This section presents analysis using the partitioned survival method, the first approach considered as an alternative to the manufacturer's economic model. It can be thought of as simpler to implement than the state-transition approaches it will be compared to in this chapter, as it only involves Kaplan-Meier curves of the survival outcomes of interest. A background to the approach was given in section 6.1 of the previous chapter. For this illustration of it using the case study, it involved the trial outcomes progression-free survival and overall survival, which are considered in turn in sections 7.4.1 and 7.4.2. For all the approaches compared in this chapter, an objective was to achieve a balance of a good fit to the data over the observed period of the trial and a suitable extrapolation to the time horizon of 15 years. For each of the survival outcomes involved in the partitioned survival, the fit over the observed period of the trial and that over the extrapolated period were initially considered separately.

7.4.1 Progression-free survival

7.4.1.1 Assessment of fit over the observation period of the trial

This section assesses the fit (of survival regressions) over the observed period of the trial for progression-free survival. Some methodological background to this was given in section 6.4 of the previous chapter. Before carrying out survival regression, a standard first step is to check for violations of the proportional hazards assumption. Therefore, initially plots were created to visualise the data and to check for such violations. Figure 7.1 shows the Kaplan-Meier estimates of

progression-free survival, together with 95% confidence intervals, for the treatment groups RFC and FC.

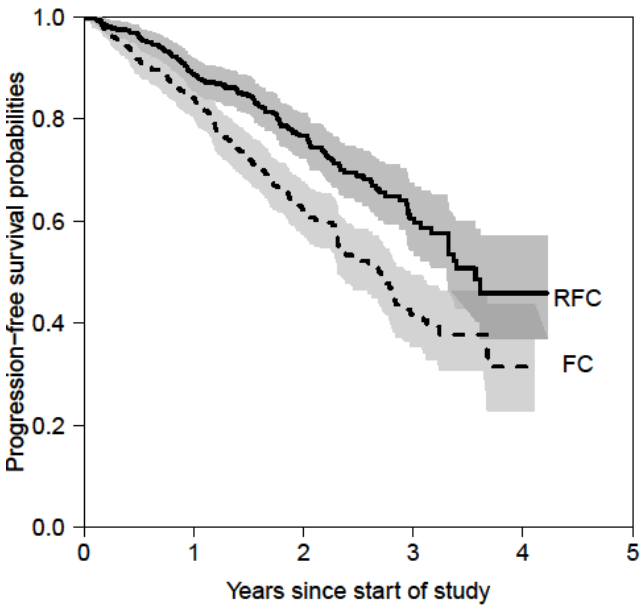


Figure 7.1 Kaplan-Meier estimates of progression-free survival for RFC and FC

While progression-free survival did drop slightly quicker in the initial 1 year period, it settled down after this and the lines were still roughly parallel throughout with no severe convergence or divergence or crossing of the lines.

A log-log plot, including a dashed reference line with slope 1, can be seen in Figure 7.2.

The lines were roughly parallel and appeared to have a slope greater than 1. This indicated the underlying data followed a distribution with hazards that increased with time, suggesting that a Weibull or Gompertz fit (rather than an exponential) was worth consideration.

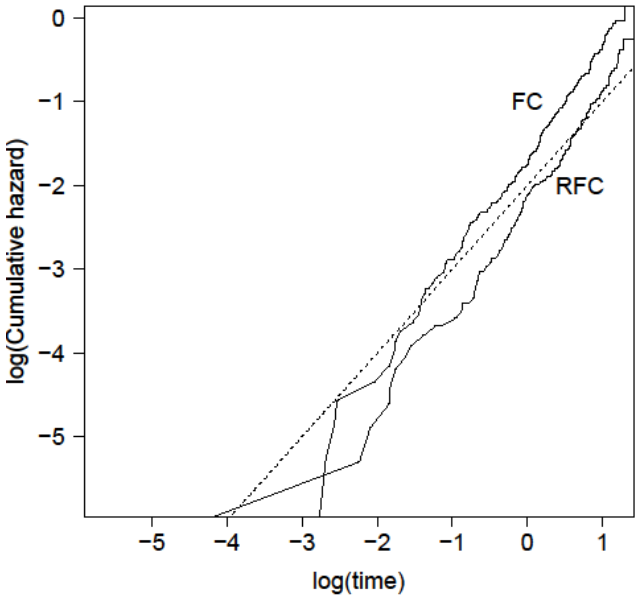


Figure 7.2 log-log plot of progression-free survival

Figure 7.3 shows a cumulative hazard plot.

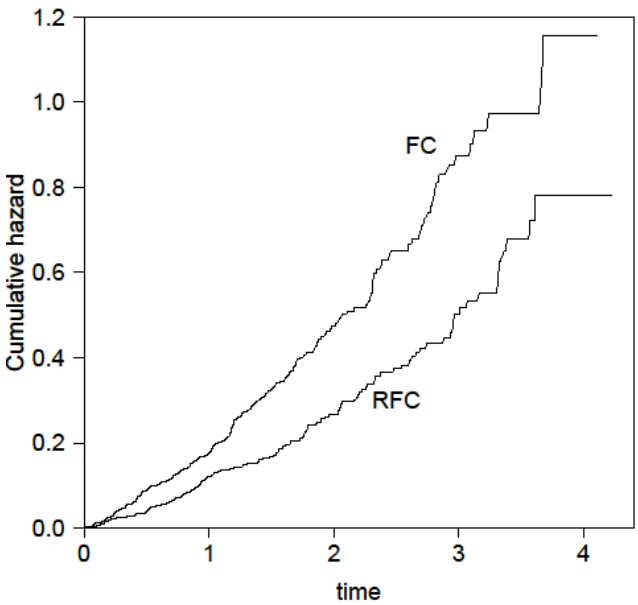


Figure 7.3 Cumulative hazard plot of progression-free survival

The lines appeared to diverge at a constant rate – indicative of hazards that increase over time - suggesting again that a Weibull or Gompertz fit (rather than an exponential) was worth consideration.

Since there was nothing to suggest any severe violations of the proportional hazards assumption for treatment from Figures 7.1 – 7.3, it was worthwhile considering such models. (The appropriateness of accelerated failure time models was assessed at a later stage using the Cox-Snell residuals plot on page 161, thereby allowing a comparison of all accelerated failure time and proportional hazards models at the same time).

Parametric regression models were then fitted to the data over the whole observation period of the trial using each of the standard distributions. Due to no evidence of severe violations of the proportional hazards assumption, models with treatment as a covariate were fitted rather than fitting a separate model for each treatment. Figure 7.4 (a) and (b) show, for RFC and FC respectively, the Kaplan-Meier estimates of progression-free survival and that predicted from each of the models.

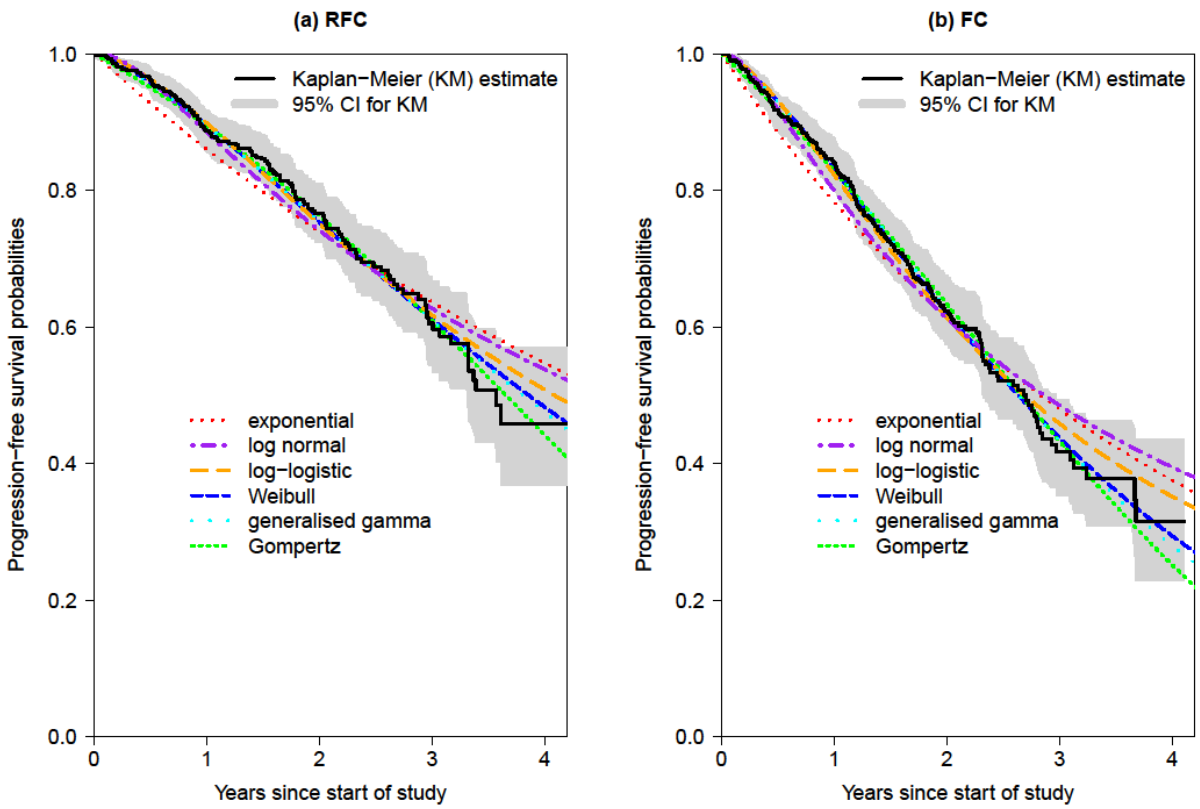


Figure 7.4 Kaplan-Meier and prediction estimates of progression-free survival: (a) RFC (b) FC

Figure 7.4(a) shows the exponential and log normal fitted the RFC data least well. The other distributions all provided a reasonable (similar) fit up to 3 years after which they started to diverge. The Weibull, generalised gamma and Gompertz distributions appeared to provide relatively good fits from year 3 to 4. However it was difficult to choose between distributions over this period due to the greater uncertainty in the Kaplan-Meier estimates. It can be seen in Figure 7.4(b) that again the Weibull, generalised gamma and Gompertz distributions provided the better fits to the FC data, although there was little to choose between them.

Figure 7.5 displays the Cox-Snell residuals following each of the model fits.

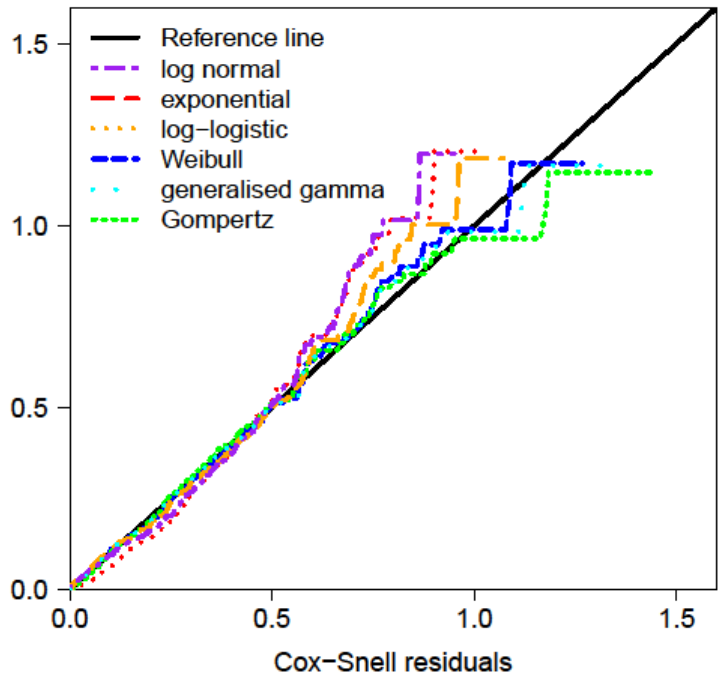


Figure 7.5 Cox-Snell residuals following the parametric regression of progression-free survival

Again, it indicated the fits from Weibull, generalised gamma and Gompertz were more suitable.

Table 7.5 shows the AICs for each of the model fits.

Distribution	AIC
exponential	1575.9
Weibull	**1545.7**
Gompertz	1548.9
log-logistic	1551.5
log normal	1572.6
generalised gamma	1547.4

Table 7.5 AICs from parametric regressions models for progression-free survival

It suggested that the Weibull provided the best fit of all the distributions considered as it had the lowest AIC.

7.4.1.2 Extrapolation of progression-free survival

The assessment of fit then moved on to considering the unobserved extrapolation period from 4 years to the time horizon of 15 years. The conventional approach to extrapolation of fitting parametric regressions using standard distributions to the whole Kaplan-Meier curve, and then extending the fit out to the desired time horizon was used. Extrapolations that represented zero by 15 years were considered reasonable. As with the assessment of the observed fit presently previously, the models had treatment as a covariate rather than fitting separate models for each treatment.

In Figure 7.6(a) and (b) the extrapolation of progression-free survival to 15 years can be seen, for RFC and FC respectively, based on the model fits using each distribution.

It is evident that the Weibull, generalised gamma and Gompertz extrapolations most adequately represented the time horizon of 15 years, although the Gompertz survival reaches zero somewhat earlier than that time point.

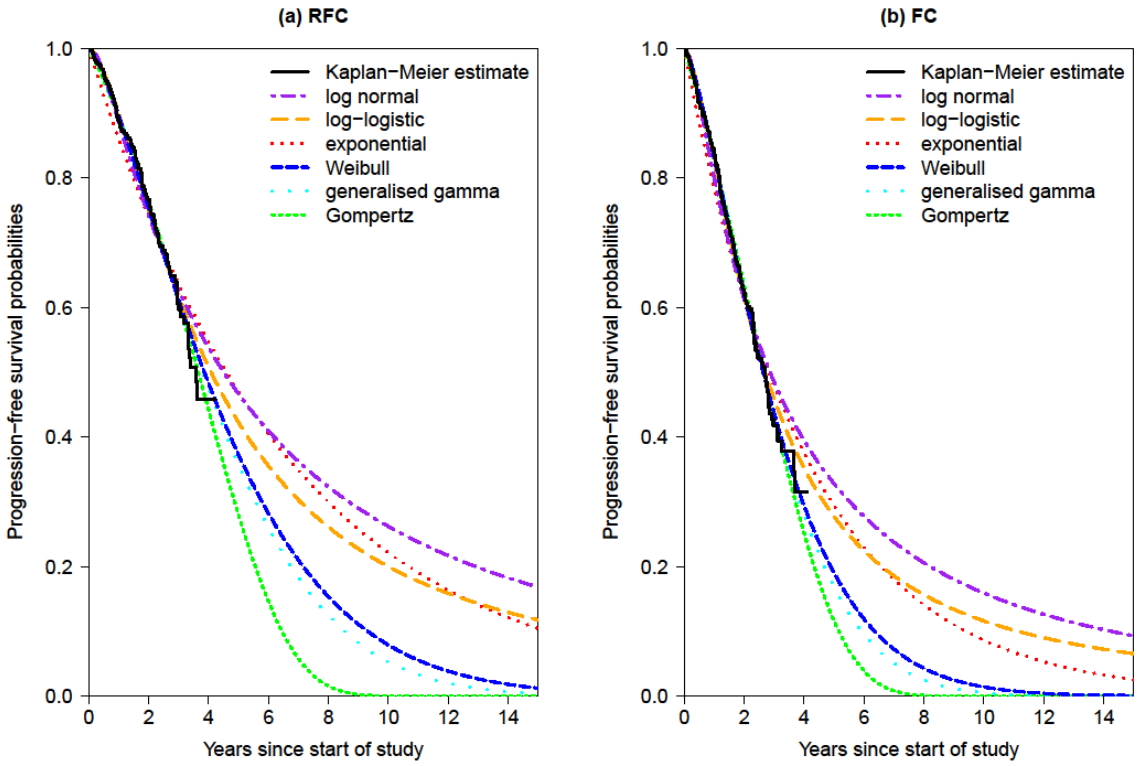


Figure 7.6 Extrapolated progression-free survival to 15 years: (a) RFC (b) FC

7.4.1.3 Choice of base-case model

This part takes into consideration all the results presented in this section so far to decide on a base-case model for progression-free survival. The Weibull, generalised gamma and Gompertz models with treatment as a covariate all seemed to provide a reasonable fit over the observed period of the trial. The AIC indicated the Weibull provided the best fit. Taking into account both these assessments of fit, the Weibull model was considered the best fit to the observed data.

The Weibull also provided a sensible extrapolation in terms of representing a lifetime of 15 years. Sensitivity analysis considering separate Weibull fits for each treatment showed there was very little difference between them and the Weibull model with treatment as a covariate (see below). Therefore the Weibull fit with treatment as a covariate, and then extrapolated to 15 years, was chosen as the base case model for progression-free survival. A summary of the model is shown in Table 7.6.

	coefficient	s.e.	HR (95% CI)	p-value
treatment (RFC vs FC)	-0.519	0.117	0.595 (0.473, 0.748)	<0.001
log(scale)	1.237	0.060		
log(shape)	0.310	0.051		

Table 7.6 Weibull model for progression-free survival

7.4.1.4 Selected sensitivity analyses

Sensitivity analyses were carried out to check there were no substantive differences between the model fits with treatment as a covariate and separate fits for each of the treatments. The fit over the observed period of the trial and that over the extrapolated period were considered separately.

Sensitivity analysis with fit for each treatment – observed period

The black solid line in Figure 7.7 and Figure 7.8 show the Kaplan-Meier estimate of progression-free survival for RFC and FC respectively. All other solid lines in Figure 7.7 and Figure 7.8 show, for RFC and FC respectively, the progression-free survival predicted by models fitted with treatment as a covariate, with colour used to distinguish between different distributions. For each colour, the corresponding dashed line shows the prediction when only the data from the specified treatment were fitted with the same distribution.

Figure 7.7 and Figure 7.8 show there was very little difference between the model types for a given distribution for either of the treatments, with some of the lines being indistinguishable. Please note, the y-axis of Figure 7.7 starts at 0.4, rather than 0, for clarity.

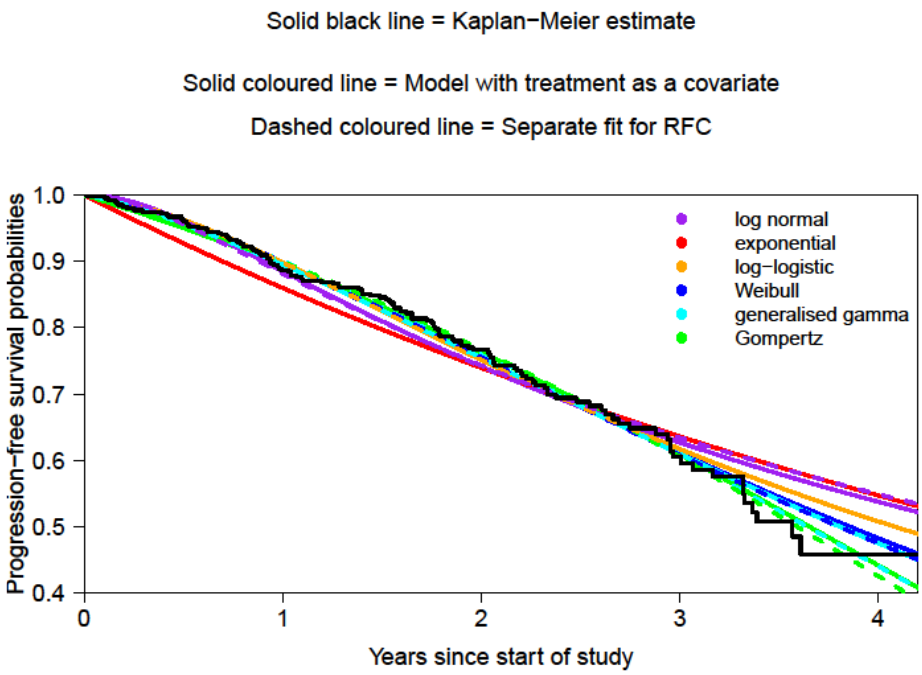


Figure 7.7 Predicted progression-free survival: treatment as a covariate vs separate fit for RFC.

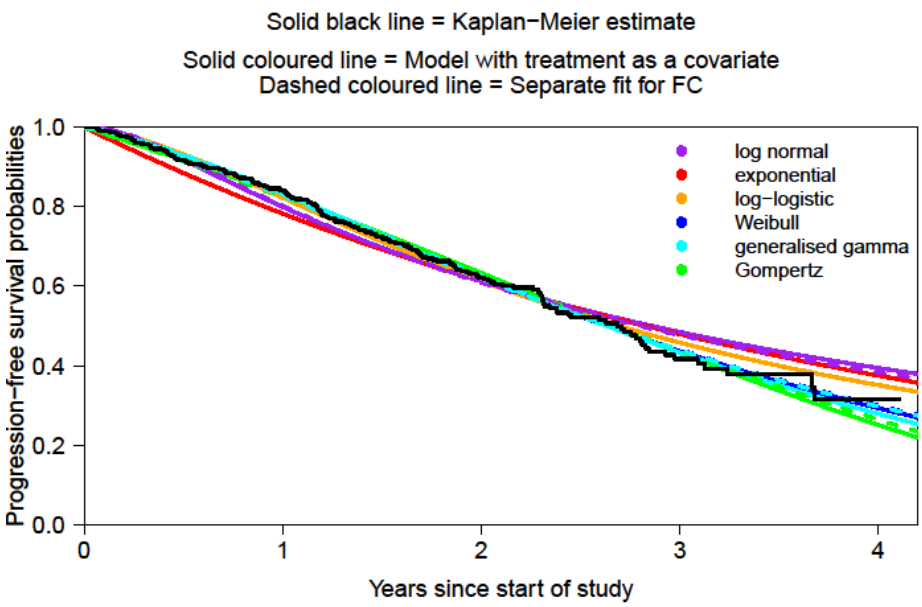


Figure 7.8 Predicted progression-free survival: treatment as a covariate vs separate fit for FC.

Table 7.7 and Table 7.8 show, for RFC and FC respectively, the area under the predicted progression-free survival curve for each of the two types of fits for each distribution. There was very little difference between the two types of fit,

with the exception of log normal for FC which was found to be one of the least reasonable fits anyway.

	with treatment as a covariate in the model	separate fit for each treatment
exponential	3.120	3.120
Weibull	3.109	3.107
Gompertz	3.084	3.079
log-logistic	3.134	3.134
log normal	3.143	3.156
generalised gamma	3.102	3.088

Table 7.7 Progression-free survival (years) AUCs from model with treatment as a covariate vs separate fit for RFC

	with treatment as a covariate in the model	separate fit for each treatment
exponential	2.586	2.586
Weibull	2.567	2.568
Gompertz	2.543	2.548
log-logistic	2.607	2.607
log normal	2.635	2.621
generalised gamma	2.562	2.567

Table 7.8 Progression-free survival (years) AUCs from model with treatment as a covariate vs separate fit for FC

Sensitivity analysis with fit for each treatment – extrapolation period

Figure 7.9 and Figure 7.10 show, for RFC and FC respectively, the results of a sensitivity analysis carried out to assess whether there were any substantive differences between the extrapolation based on the models with treatment as a covariate and those with separate fits for each of the treatments.

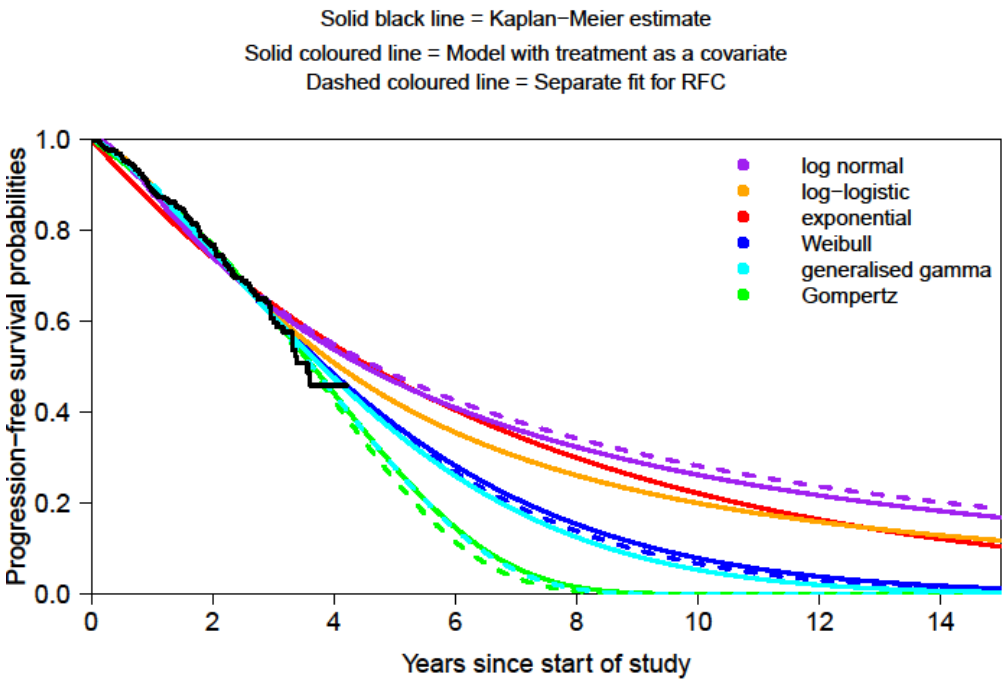


Figure 7.9 Extrapolated progression-free survival: treatment as a covariate vs separate fit for RFC.

Only one red curve is shown in Figure 7.9 for the exponential distribution because both types of model produced identical fits. The curves from the two types of log-logistic model were indistinguishable by eye, and therefore again only one orange curve is evident. The two types of model for the log normal, Weibull and Gompertz produced very similar fits. The generalised gamma was the only distribution for which there was a more marked difference: its separate fit for RFC was indistinguishable from the fit for the Gompertz with treatment as a covariate.

Figure 7.10 displays only one curve for the exponential and log-logistic for the same reason as in Figure 7.9.

The two fits for the Weibull and the separate generalised gamma fit for FC were indistinguishable. Furthermore, the two fits for the log-normal and Gompertz were very similar.

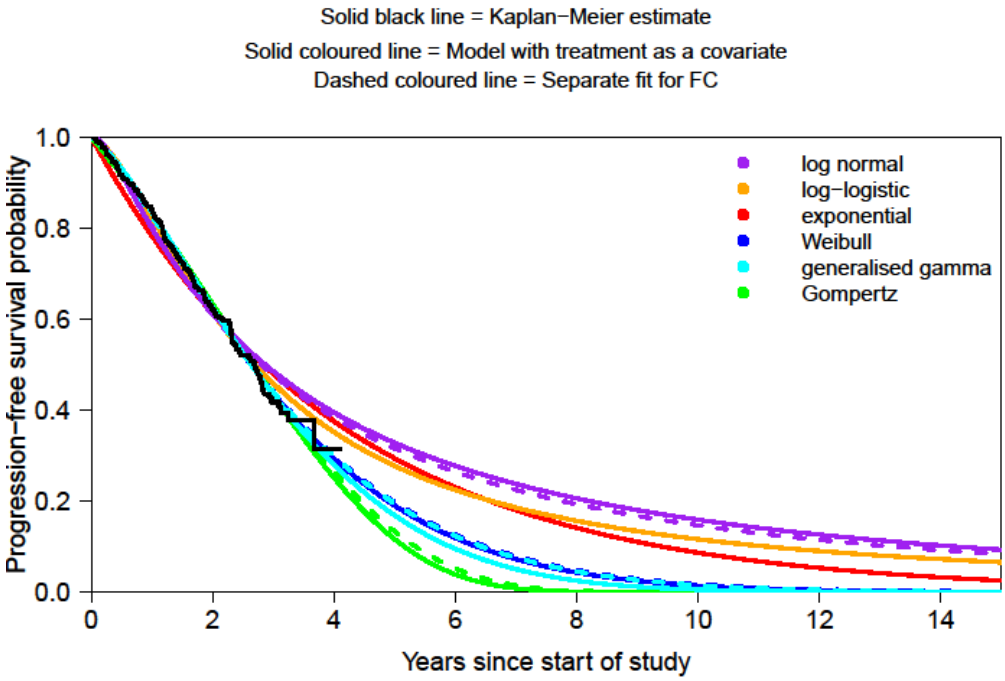


Figure 7.10 Extrapolated progression-free survival: treatment as a covariate vs separate fit for FC.

Table 7.9 and Table 7.10 show, for RFC and FC respectively, the area under progression-free survival curves extrapolated to 15 years for each of the two types of fit for a given distribution.

There was very little difference between the two types of fit for a given distribution, with the exception of log normal - which did not adequately represent the time horizon of 15 years anyway – and generalised gamma.

	with treatment as a covariate in the model	separate fit for RFC
exponential	5.939	5.939
Weibull	4.590	4.473
Gompertz	3.701	3.598
log-logistic	5.706	5.703
log normal	6.322	6.535
generalised gamma	4.363	3.689

Table 7.9 Extrapolated progression-free survival AUCs from model with treatment as a covariate vs separate fit for RFC

	with treatment as a covariate in the model	separate fit for FC
exponential	3.978	3.978
Weibull	3.154	3.196
Gompertz	2.796	2.845
log-logistic	4.210	4.214
log normal	4.704	4.560
generalised gamma	3.023	3.175

Table 7.10 Extrapolated progression-free survival AUCs from model with treatment as a covariate vs separate fit for FC

7.4.2 Overall survival

The focus of the analysis then changed to overall survival, the other survival outcome involved in this illustration of partitioned survival.

7.4.2.1 Assessment of fit over the observation period of the trial

In a similar manner to the beginning of the previous sub-section that focused on progression-free survival, this sub-section focuses on a corresponding assessment of fit over the observed period of the trial for overall survival. As such, initially plots were created to visualise the data and to check for violations of the

proportional hazards assumption. Figure 7.11 shows the Kaplan-Meier estimates of overall survival, together with 95% confidence intervals, for the treatment groups RFC and FC.

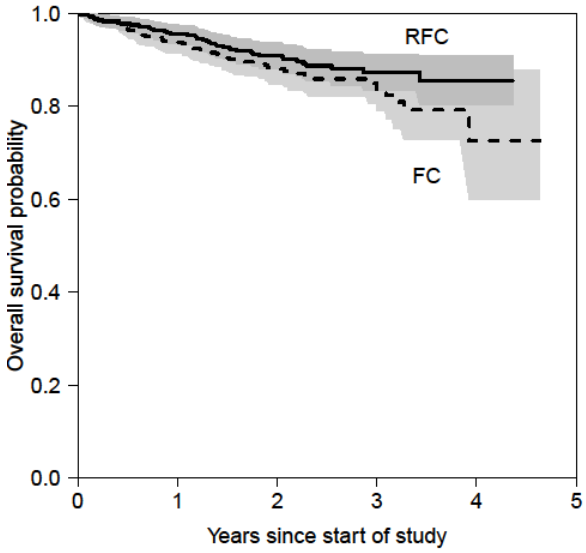


Figure 7.11 Kaplan-Meier estimates of overall survival for RFC and FC

The lines were roughly parallel in the first three years indicating no obvious violation of the proportional hazards assumption over this period. After three years there was more divergence as there was a greater drop in survival in the FC group. However there was a greater degree of uncertainty in the data over this period.

A log-log plot including a dashed reference line with slope 1 can be seen in Figure 7.12.

The lines were roughly parallel and appeared to have a slope of 1. This would suggest the underlying data follow an exponential distribution. The lines did cross over for a short period – however this was not unexpected given the closeness of the lines.

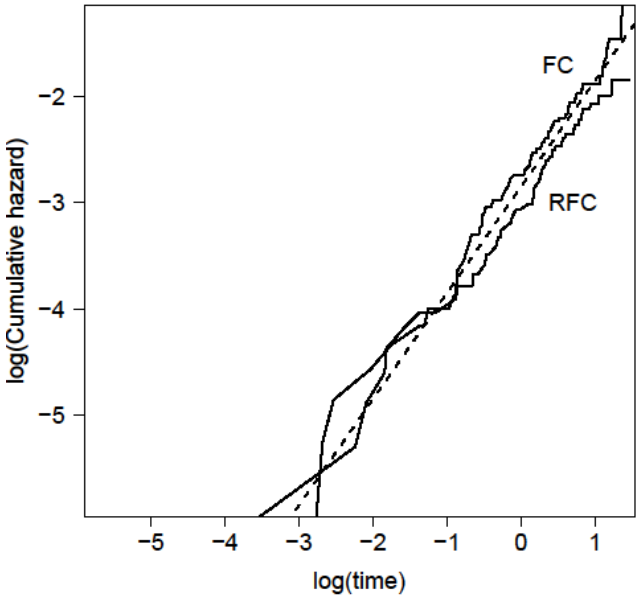


Figure 7.12 Log-log plot of all-cause mortality (1-overall survival)

Figure 7.13 shows a cumulative hazard plot.

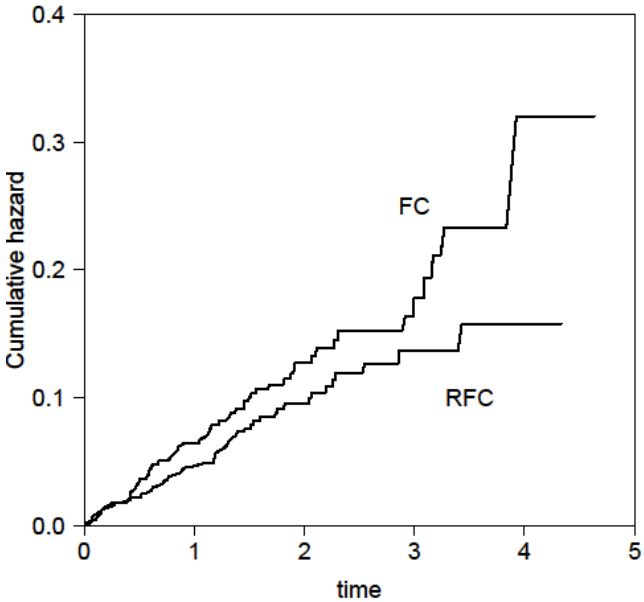


Figure 7.13 Cumulative hazard plot of all-cause mortality (1-overall survival)

The lines appeared roughly parallel (up until 3 years anyway, after which there was greater uncertainty in the estimates) suggestive of a constant hazard of dying. Therefore the exponential was an appropriate distribution for initial consideration. Since there was nothing to suggest any severe violations of the proportional hazards assumption for treatment from Figures 7.11 - 7.13, it was worthwhile considering such models, and in particular the exponential. Again, assessment of the accelerated failure time models was carried out using the Cox-Snell residuals plot (page 173).

Parametric regression models were then fitted over the whole observation period of the trial for each treatment using each of the standard distributions. As with the analysis presented previously for progression-free survival, models were fitted with treatment as a covariate, rather than fitting a separate model for each treatment. Figure 7.14(a) and (b) show, for RFC and FC respectively, the Kaplan-Meier estimate of survival with the 95% confidence interval and the predicted survival from each of the models. Please note, the y-axes of both Figures do not reach zero for clarity.

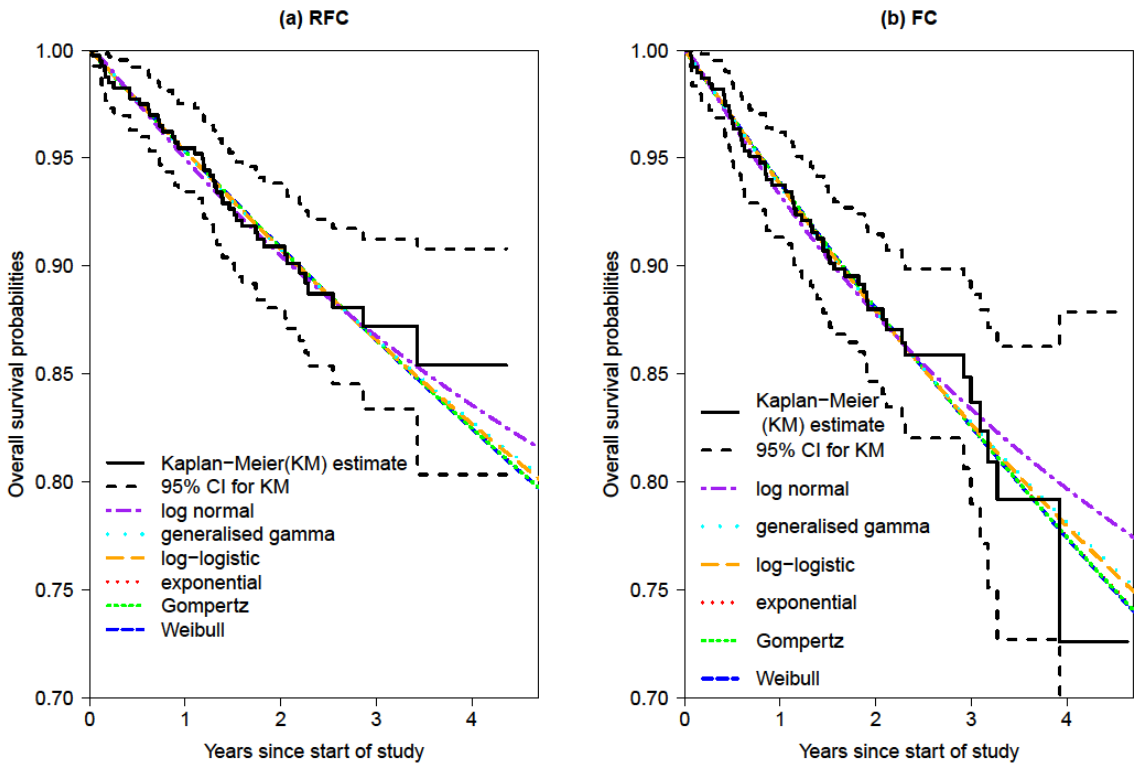


Figure 7.14 Kaplan-Meier and prediction estimates of overall survival: (a) RFC (b) FC

Figure 7.14(a) and (b) show that the fits from each of the six distributions were very similar up to 2.5 years, after which there was some divergence. However, In this latter period there was greater uncertainty in the data over this period, especially for FC.

The Cox-Snell residuals following each of the model fits are presented in Figure 7.15.

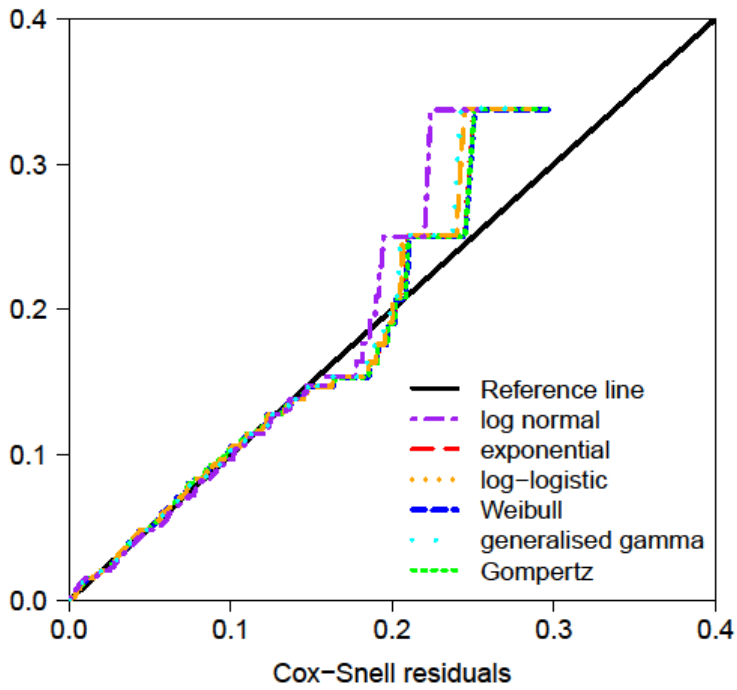


Figure 7.15 Cox-Snell residuals following the parametric regression of overall survival

Again, it demonstrated the fits were, until divergence in the later period, all reasonable as well as being very similar.

Table 7.11 shows the AICs from each of the model fits.

It suggested that the exponential provided the best fit of the distributions considered as it had the lowest AIC value.

Distribution	AIC
exponential	**756.940**
Weibull	758.936
Gompertz	758.939
log-logistic	758.978
log normal	759.930
generalised gamma	760.812

Table 7.11 AICs from parametric regressions models for overall survival

7.4.2.2 Extrapolation of overall survival

The assessment of fit then moved on to considering the unobserved extrapolation period from 4 years to the time horizon of 15 years. The conventional approach to extrapolation of fitting parametric regressions using standard distributions to the whole Kaplan-Meier curve, and then extending the fit out to the desired time horizon was considered initially. Extrapolations that represented zero by 15 years were considered reasonable. As with the assessment of the observed fit presently previously, the models had treatment as a covariate rather than fitting separate models for each treatment.

Figure 7.16 (a) and (b) show, for RFC and FC respectively, the predicted overall survival from each of the models, together with the corresponding observed Kaplan-Meier estimate. It can be seen that each of the models were far from producing overall survival of zero at 15 years.

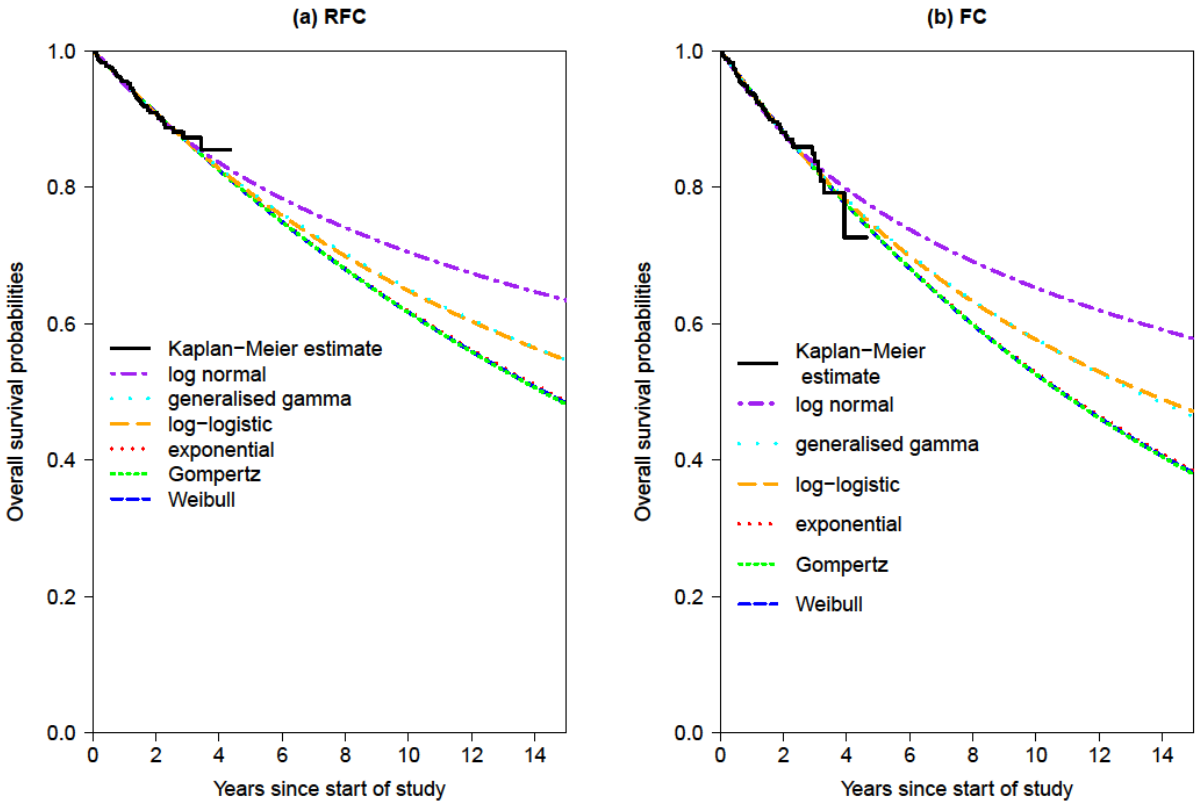


Figure 7.16 Extrapolated overall survival to 15 years: (a) RFC (b) FC

Therefore focus changed to starting the extrapolation from the tail of the Kaplan-Meier curve. Each of the observed time points were considered in turn as the starting point for extrapolation. Weibull models were fitted following the approach of Tappenden et al. (2006) mentioned previously in section 6.5.

Each of the treatment groups were considered separately.

Exponential models were considered but were not fitted. This was due to the intercept in the linear regression acting as a nuisance parameter that caused problems in transforming the fitted line to an exponential survival function. Gompertz models were also considered using the approach of Tappenden et al. (2006) but an assumption of these models, namely that the logarithm of the hazard was linear with time, was not met and therefore they were not fitted to the data. Tappenden et al. (2006)'s article did not include corresponding details for the accelerated failure time framework and were not derived for this thesis.

Extrapolation from the tail of the overall survival Kaplan-Meier curve: FC

Table 7.12 and Figure 7.17 summarise the results of fitting Weibull extrapolation models for FC, showing the results based on a wide range of starting points for extrapolation – those closest to 0, 0.5, 1, 1.5, 2, 2.5, 3 and 3.5 years & the one that best represented a time horizon of 15 years, namely a starting point of 3.663 years.

Starting point of extrapolation (years)	n at risk at starting point of extrapolation	survival at 15 years
0	409	36.9%
0.498	372	42.4%
0.999	351	40.2%
1.487	319	39.2%
1.996	201	32.5%
2.489	120	10.6%
2.773	95	4.7%
2.995	73	11.0%
3.491	32	2.0%
3.663	23	0.3%

Table 7.12 Extrapolation of overall survival for FC using Weibull models - Summary using a selection of starting points

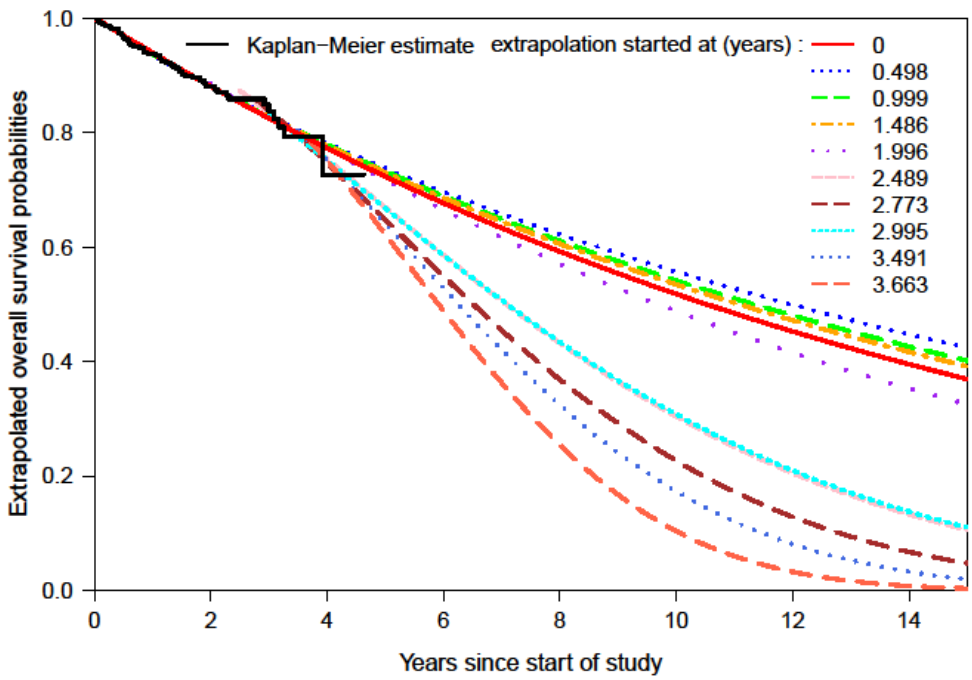


Figure 7.17 Extrapolation of overall survival for FC using Weibull models - Summary using a selection of starting points

Starting time points in the ranges 2.727 – 2.828 and 3.455 - 3.762 years achieved survival of less than 5% upon extrapolation to 15 years. While there was less uncertainty in the Kaplan-Meier estimate in the time points over the range 2.727 – 2.828, the survival they achieved was as high as 4.7% at best upon extrapolation to 15 years.

However, time points over the range 3.455 - 3.762 provided better representation of the time horizon, with the smallest survival estimate at 15 years being 0.3% when the extrapolation started at 3.663 years. Therefore, a Weibull fit to the Kaplan-Meier tail beginning at 3.663 years was the basis for the extrapolated section of the base case model of overall survival for FC.

Extrapolation from the tail of the overall survival Kaplan-Meier curve: RFC

Table 7.13 and Figure 7.18 summarise the results of fitting Weibull models for RFC.

Starting point of extrapolation (years)	n at risk at starting point of extrapolation	survival at 15 years
0	408	52.2%
0.479	389	56.6%
0.983	375	61.3%
1.495	356	63.9%
1.993	240	66.7%
2.486	147	70.0%
2.998	85	70.2%
3.493	44	85.4%
3.973	10	85.4%

Table 7.13 Extrapolation of overall survival for RFC using Weibull models - Summary using a selection of starting points

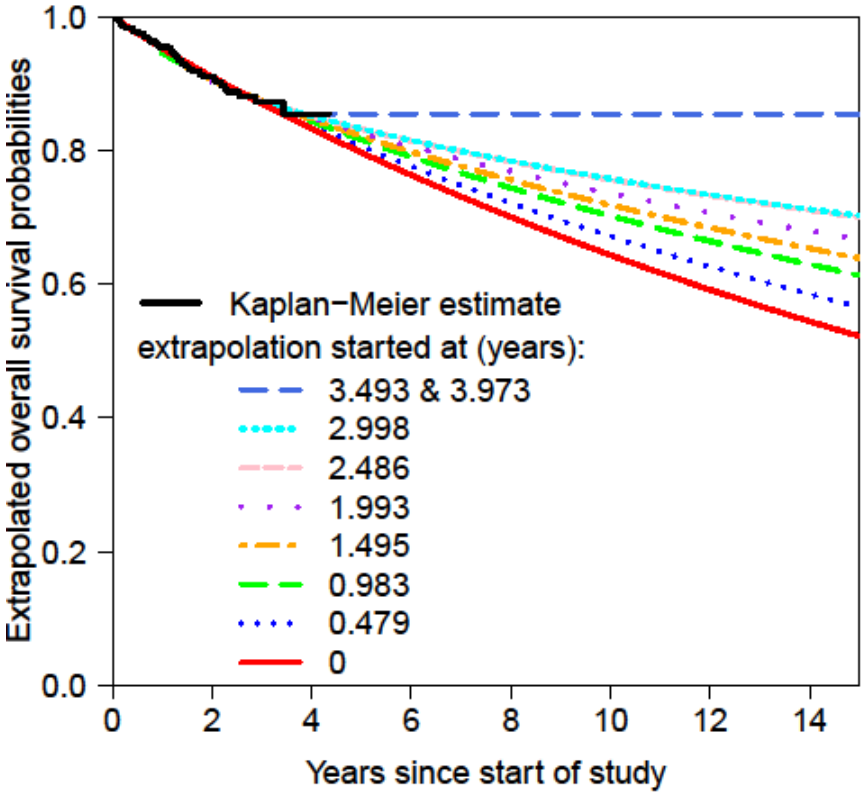


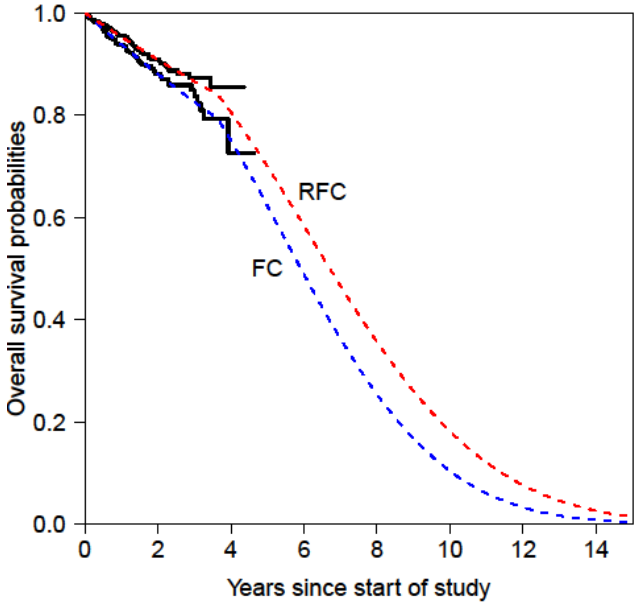
Figure 7.18 Extrapolation of overall survival for RFC using Weibull models - Summary using a selection of starting points

It can be seen that none of the models provided any fits that came anywhere close to reaching zero at 15 years, and that there was no improvement on the original fit to the whole Kaplan curve (i.e. the red line representing the extrapolation that started at 0 years) with any of the distributions.

Since neither a parametric fit to the whole – or a tail of any length of the – Kaplan-Meier curve adequately represented survival of zero upon extrapolation to 15 years, it was decided to derive the extrapolated survival probabilities for RFC by applying a hazard ratio to the survival probabilities already predicted for FC.

As a starting point, the hazard ratio of 0.753 from the exponential model over the observed period (the model found to be the most reasonable fit to the observed data) was applied to the extrapolated survival probabilities already

predicted for FC. Figure 7.19 shows the predicted overall survival for both RFC and FC. Using a hazard ratio of 0.753 resulted in a reasonable representation of



the time horizon with survival of 1.4% at 15 years for RFC.

Figure 7.19 Overall survival to 15 years for RFC and FC

7.4.2.3 Base-case models used for overall survival

Table 7.14 shows a summary of the models used for overall survival.

	parameter
exponential fit over observed period (0 - 3.663 years)	
lambdaFC	0.064
lambdaRFC	0.048
Weibull fit used for extrapolation (3.663 - 15 years)	
gamma	2.257
lambdaFC	0.013

Table 7.14 Base case models for overall survival

An exponential regression with treatment as a covariate was used for predictions over the observed period of 0-3.663 years. Therefore, the predictions for both treatment arms were derived using the same model.

Over the extrapolated period, a Weibull fit was derived for the FC arm, using the aforementioned approach of Tappenden et al. (2006) based on linear regression. The extrapolation for the RFC arm was derived by applying a hazard ratio of 0.753 to the extrapolated survival probabilities for the FC arm.

7.4.2.4 Selected sensitivity analyses

Sensitivity analyses were carried out to check there were no substantive differences between the model fits with treatment as a covariate and separate fits for each of the treatments. Only the fit over the observed period of the trial was considered. Such a sensitivity analysis over the extrapolated period was not required because the base-case extrapolation was already based on separate fits for each treatment.

The results of the aforementioned sensitivity analyses are shown in Figure 7.20 (RFC) and Figure 7.21 (FC). Please note, the y-axes of both Figures do not reach zero for clarity.

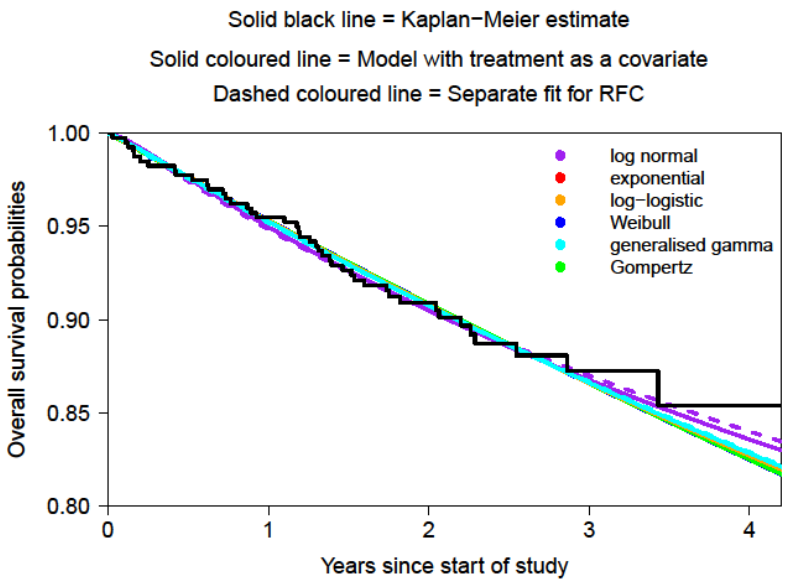


Figure 7.20 Predicted overall survival: treatment as a covariate vs separate fit for RFC.

The exponential, log-logistic, Weibull, generalised gamma and Gompertz curves in Figure 7.20 from the two types of model were very similar to the extent they were difficult to distinguish by eye. The two types of model using the log normal distribution also produced very similar fits.

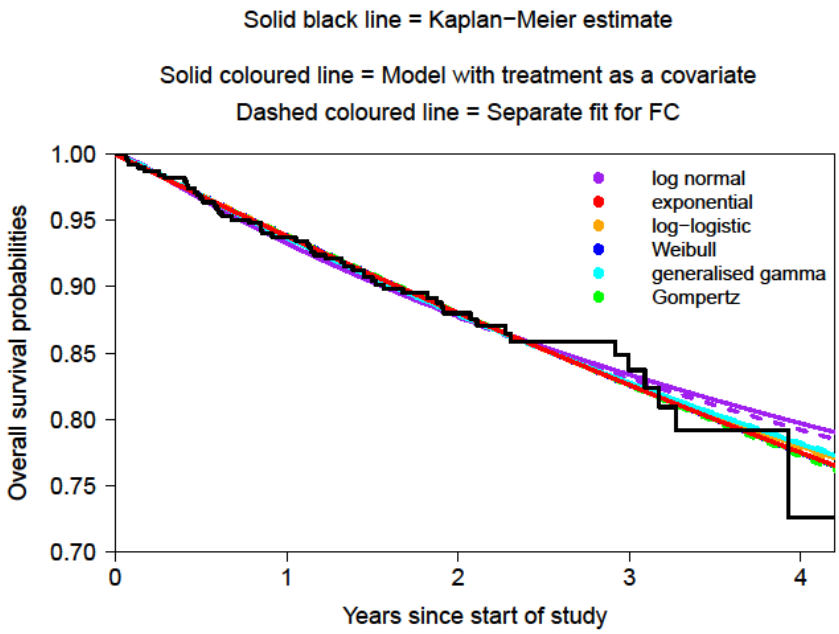


Figure 7.21 Predicted overall survival: treatment as a covariate vs separate fit for FC.

As in Figure 7.20, it was difficult to distinguish between the exponential, log-logistic, Weibull, generalised gamma and Gompertz curves from the two types of model (Figure 7.21). The two fits using the log-normal distribution were again very similar.

Table 7.15 and Table 7.16 show, for RFC and FC respectively, the area under the overall survival curves for each of the two types of fits for each distribution. There was very little difference between the two types of fit.

Table 7.15 Overall survival AUCs from model with treatment as a covariate vs separate fit for RFC

	with treatment as a covariate in the model	separate fit for each treatment
exponential	3.631	3.631
Weibull	3.634	3.625
Gompertz	3.633	3.618
log-logistic	3.617	3.607
log normal	3.550	3.541
generalised gamma	3.607	3.607

Table 7.16 Overall survival AUCs from model with treatment as a covariate vs separate fit for FC

	with treatment as a covariate in the model	separate fit for each treatment
exponential	3.037	3.037
Weibull	3.041	3.052
Gompertz	3.039	3.056
log-logistic	3.007	3.019
log normal	2.912	2.926
generalised gamma	2.995	2.985

7.4.3 Compatibility of fits for progression-free and overall survival

While carrying out the partitioned survival approach, each of the health states that overall survival was partitioned into were modelled separately. Figure 7.22(a) and (b) show the resulting base case progression-free survival curve and overall survival curve (section 7.4.1 and 7.4.2) for FC and RFC respectively.

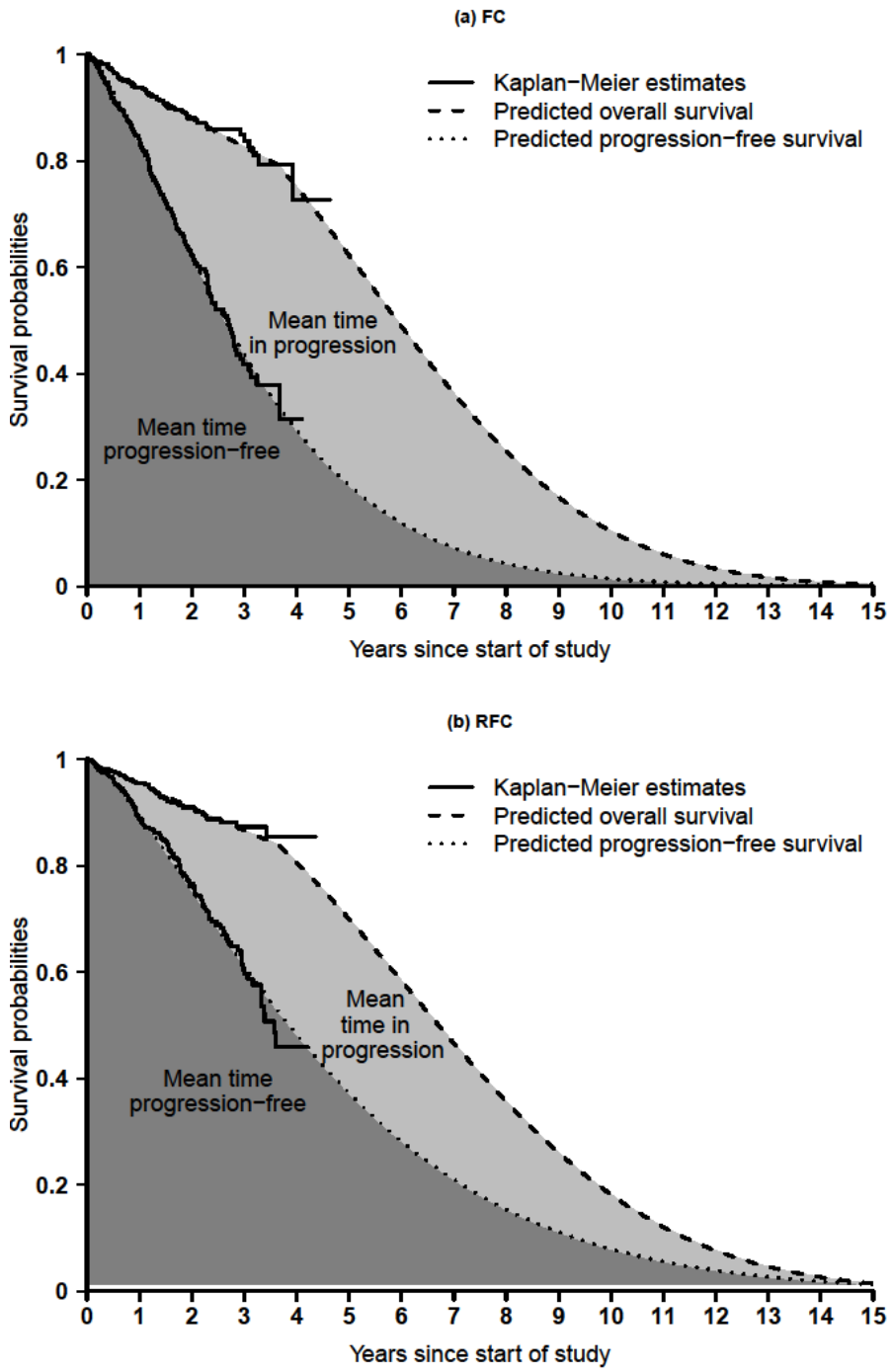


Figure 7.22 Base case partitioned survival for (a) FC and (b) RFC

They provide a visual check that each of the fits were compatible, which may not always be the case because they were modelled separately. It can be seen that the fits were compatible for each of the treatment arms, in the sense that overall survival was greater than or equal to progression-free survival throughout the 15 years.

7.4.3.1 Sensitivity analysis for treatment effect in extrapolated period

A sensitivity analysis was then carried out, reducing the treatment effect upon extrapolation (Table 7.17). There was a slight improvement in the representation of the whole lifetime upon reduction of the treatment effect, in the sense of overall survival reaching zero. However, none of the reductions provided a sensible extrapolation to 15 years, with regards to the compatibility specified above, as they all resulted in survival at 15 years which was less than the 1.2% survival for progression-free survival.

hazard ratio	overall survival at 15 years
0.753	1.4%
0.8	1.1%
0.85	0.8%
0.9	0.6%
0.95	0.5%

Table 7.17 Effect of reducing the treatment effect on extrapolated overall survival to 15 years for RFC.

It was not thought appropriate to consider increasing the treatment effect upon extrapolation.

7.5 Multi-state modelling

7.5.1 Introduction

This section presents analysis of the health economics case study using the multi-state modelling approach. It was the second approach considered as an alternative to the manufacturer's economic model, and the main focus of the research in this thesis. It shares many similarities with the original manufacturer's economic modelling due to them both being state-transition modelling approaches. The structure of the model is the same in the sense the health states and the transitions between them that are modelled are identical. However, there are also some important differences. In particular, in this illustration of multi-state modelling each of the transitions between states were based on regression modelling using the individual patient level trial data. This

use of multi-state modelling as an alternative approach was motivated by the ERG's concern with the original manufacturer's model that the overall survival gain it demonstrated was not seen empirically.

This section, and those that follow, will demonstrate that anything that can be produced using conventional Markov-decision analytic modelling can just as easily be performed using multi-state modelling. The implementation of multi-state modelling for cost-effectiveness in R is described elsewhere in one of my accompanying publications to this chapter (Williams et al., 2017a). This paper also outlines the advantages of using a syntax-based approach with statistical software instead of the spreadsheet-based approaches that are more commonly employed for state-transition modelling.

In sub-section 7.5.2, just as with the partitioned survival demonstrated in the previous section, the appropriateness of the proportional hazards assumption was considered before building any models. Then, whether the Markov property held was investigated in sub-section 7.5.3. Next, the distribution to use for each of the transitions was explored in sub-section 7.5.4. Following on from this, in sub-section 7.5.5, the model used for the base-case analysis is summarised.

7.5.2 Consideration of the appropriateness of the proportional hazards assumption

This section assesses, prior to carrying out any modelling, the suitability of the proportional hazards assumption. In Figure 7.23(a), (b) and (c) a log-log plot for treatment is shown for each of the respective transitions. A log-log plot is also shown for time in previous state for the progression → death transition (Figure 7.23(d)). This was because this covariate was to be included in this transition in the modelling carried out in the next section, section 7.5.3. For Figure 7.23(d), the median time progression-free in those who experienced death after progression was used as a cut-off to provide two groups that were balanced in terms of the number of events.

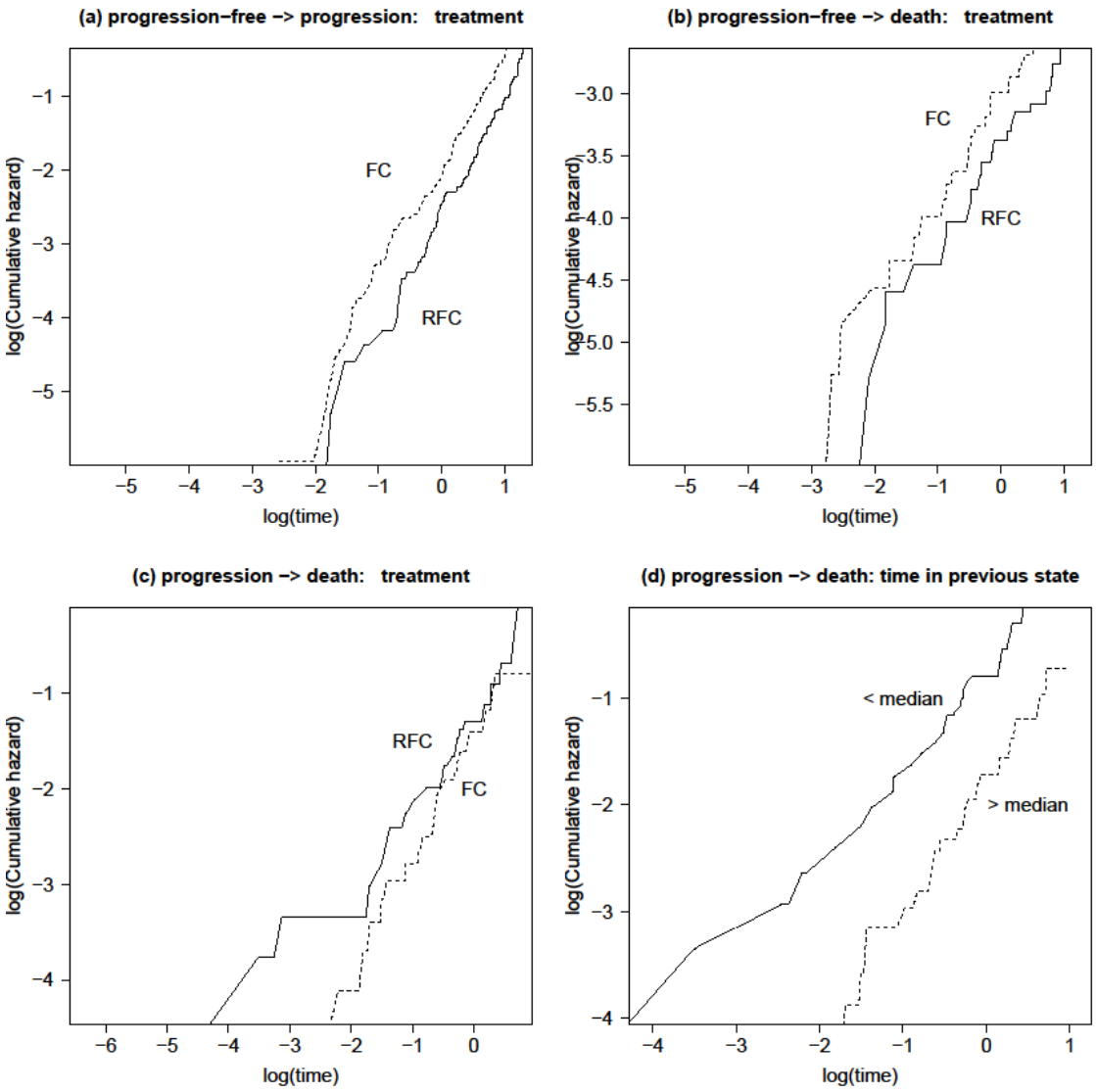


Figure 7.23 Log-log plots for the transitions in the multi-state model

It can be seen from Figure 7.23 (a), (b) and (c) that the lines in the treatment plots for each transition were roughly parallel, with any crossing of the lines due to the lack of a treatment effect rather than any major violation of the proportional hazards assumption. Figure 7.23(d) showed lines that were again roughly parallel for time in the previous state for progression → death, with no suggestion of any major violation of the proportional hazards assumption.

Figure 7.24 shows cumulative hazard versus time plots for treatment for each of the transitions. Again, a corresponding plot is also shown for time in the previous state for the progression → death transition.

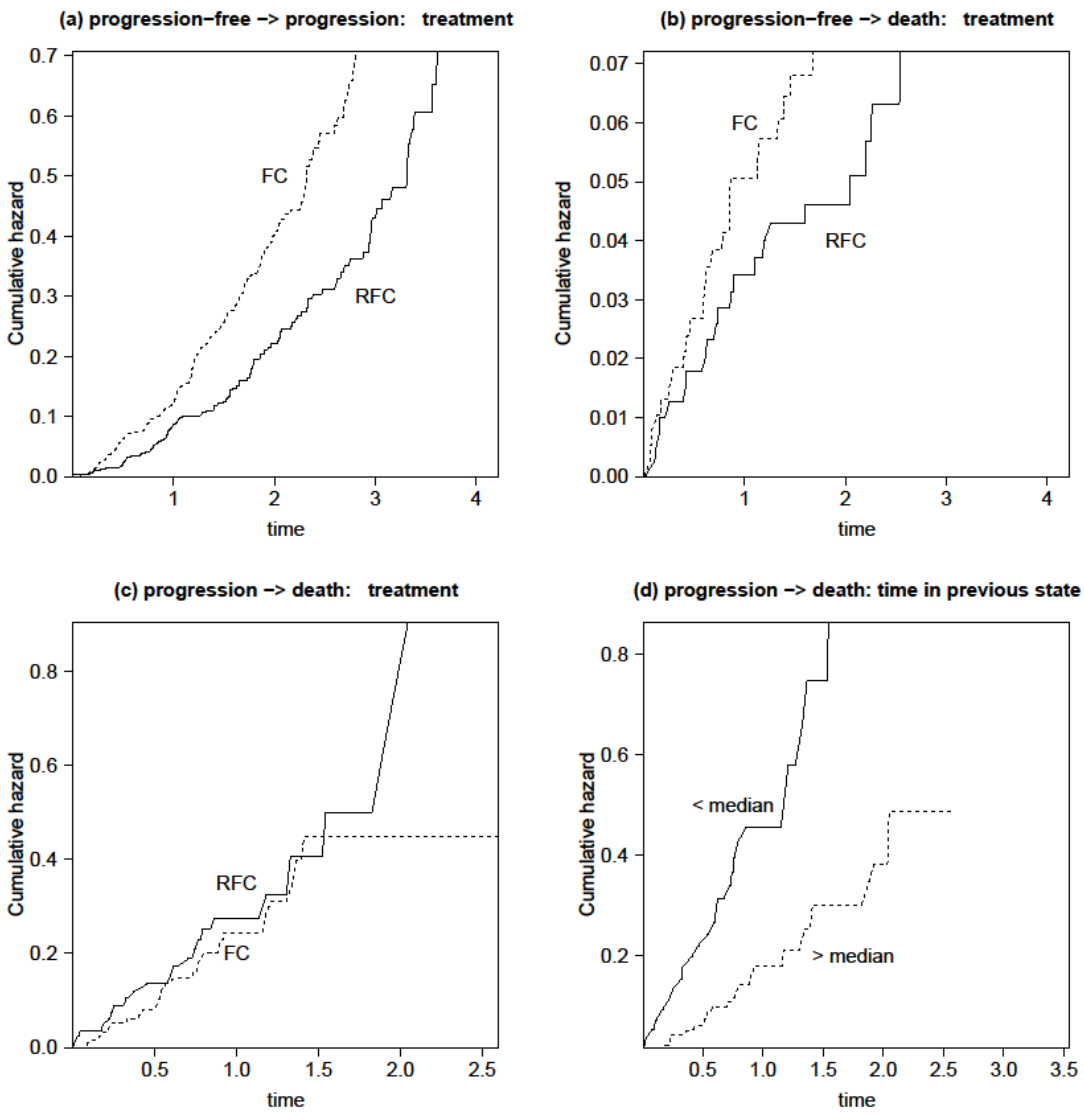


Figure 7.24 Cumulative hazard versus time plots for the transitions in the multi-state model

In Figure 7.24 (a), (b) and (d) the lines diverged suggesting that it would be worthwhile considering a distribution that allows increasing hazards over time. Figure 7.24 (c) was indicative of the lack of a treatment effect rather than any major violation of the proportional hazards assumption.

7.5.3 Deciding on the appropriateness of the Markov property

To help decide whether the Markov property held, a Cox Markov state-arrival extended model for progression → death was initially fitted. Table 7.18 shows the results of fitting this model.

	hazard ratio (95% CI)	p-value
treatment (RFC vs FC)	1.555 (0.874, 2.766)	0.133
time spent progression-free	0.413 (0.215, 0.794)	0.008

Table 7.18 Results of a Cox Markov state-arrival extended model for progression → death

As previously mentioned in Williams et al. (2017a), the time spent progression-free was found to have a statistically significant association with death after progression (p-value = 0.008). The hazard ratio point estimate and 95% confidence interval were below one, indicating that the longer the time spent progression-free, the lower the risk of death after reaching the progression state. For each increase of one year in the progression-free state the hazard of death reduced by 58.7%. The point estimate of 58.7% equates to, for those in the FC treatment arm who spent one year in the progression-free state, an absolute risk of death of 72.9% 4 years after reaching the progression state. The corresponding figure for those in the FC arm who spend two years in the progression-free state was 41.7%. The equivalent figures for the RFC arm were 86.9% and 56.8%. The effect of time in the previous state was of a size likely to be of practical importance, both in relative and absolute terms. Therefore there was evidence to suggest the Markov property did not hold, indicating that a semi-Markov model was more appropriate than a Markov approach.

7.5.4 Choice of distribution for each transition

This section assesses the fit to each of the states and relevant transitions using six standard distributions with a semi-Markov multi-state modelling approach. In

a similar manner to the partitioned survival approach presented earlier in this chapter, fits were assessed over both the observed period of the trial and the period of extrapolation. Also, similar to the previous analyses, models with treatment as a covariate were used rather than models for each treatment separately. This was thought appropriate due to the lack of any severe violation of the proportional hazards assumption presenting in section 7.5.2.

For each of the transitions in the multi-state model, the fit over the observed period of the trial and that over the extrapolated period were considered separately.

7.5.4.1 Progression → death

The fit to the progression → death transition was considered initially as this was the simplest transition to assess in the sense that the observed data could be expressed as a (standard) Kaplan-Meier survival curve.

Assessment of fit over the observed period

Figure 7.25 (a) and (b) show, for RFC and FC respectively over the observed period of the trial, the cumulative incidence (1-Kaplan-Meier) estimate of death after progression, its 95% confidence interval (CI) and estimates predicted by parametric survival regressions using standard distributions.

It can be seen from Figure 7.25 (a) and (b) that there was little to choose between the distributions for either RFC or FC, especially as they only started to diverge at the tail of the cumulative incidence curve where there was the most uncertainty.

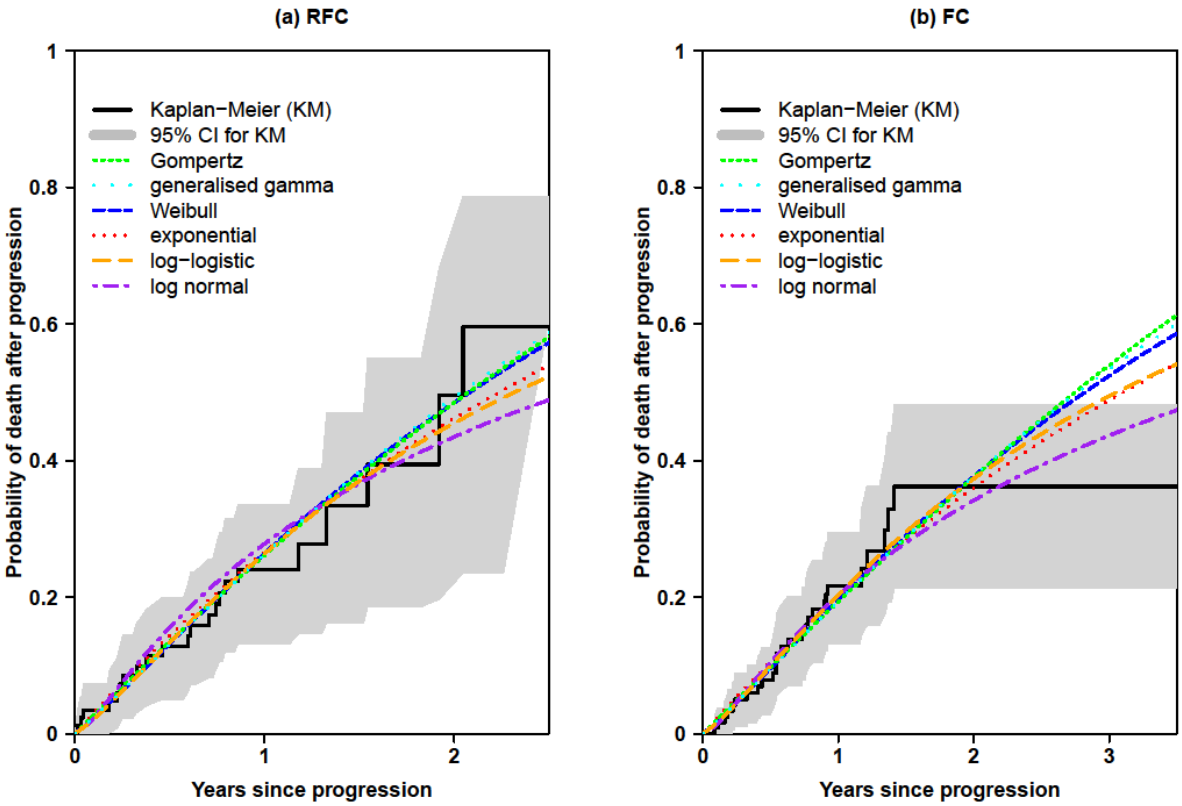


Figure 7.25 Progression → death for over trial observation period: (a) RFC and (b) FC

Figure 7.26 on the next page displays the Cox-Snell residuals following each of the model fits.

There was very little to choose between the distributions in terms of the Cox-Snell residuals.

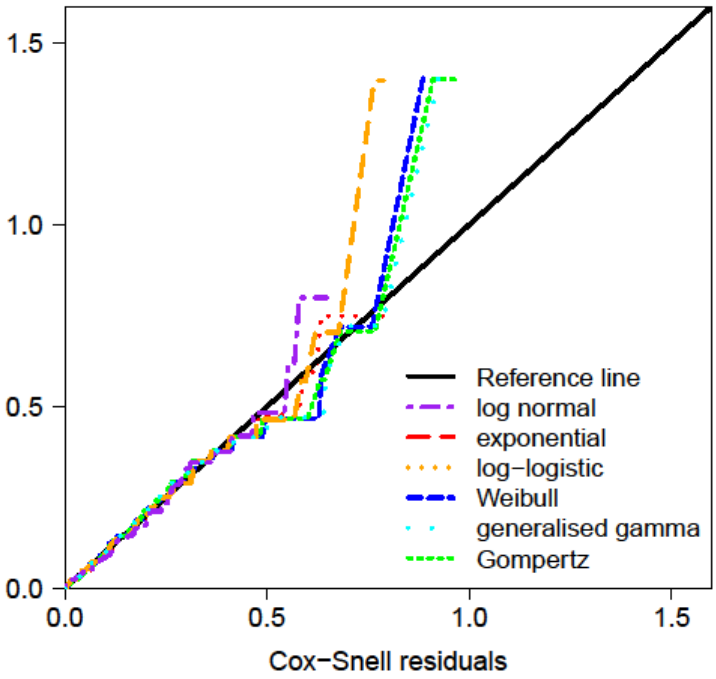


Figure 7.26 Progression → death : Cox-Snell residuals

Table 7.19 shows the AICs from each of the model fits.

	AIC
Weibull	240.2
exponential	**239.0**
Gompertz	240.5
log-logistic	240.8
log normal	243.1
generalised gamma	242.2

Table 7.19 AICs from parametric regression models for progression -> death

It suggests the exponential provided the best fit over the observed period of the trial as it had the lowest AIC value, although there was little to choose between the distributions.

Assessment of fit over the period of extrapolation

The assessment of fit then moved on to considering the unobserved extrapolation period from 4 years to the time horizon of 15 years. The conventional approach to extrapolation of fitting parametric regressions using standard distributions and then extending the fit out to the desired time horizon was used. Extrapolations that represented zero by 15 years were considered reasonable. As with the assessment of the observed fit presently previously, the models had treatment as a covariate rather than fitting separate models for each treatment.

Figure 7.27 (a) and (b) show, for RFC and FC respectively, the cumulative incidence (1-Kaplan-Meier) estimate of death after progression and that predicted by parametric survival regressions using standard distributions with extrapolation to 15 years.

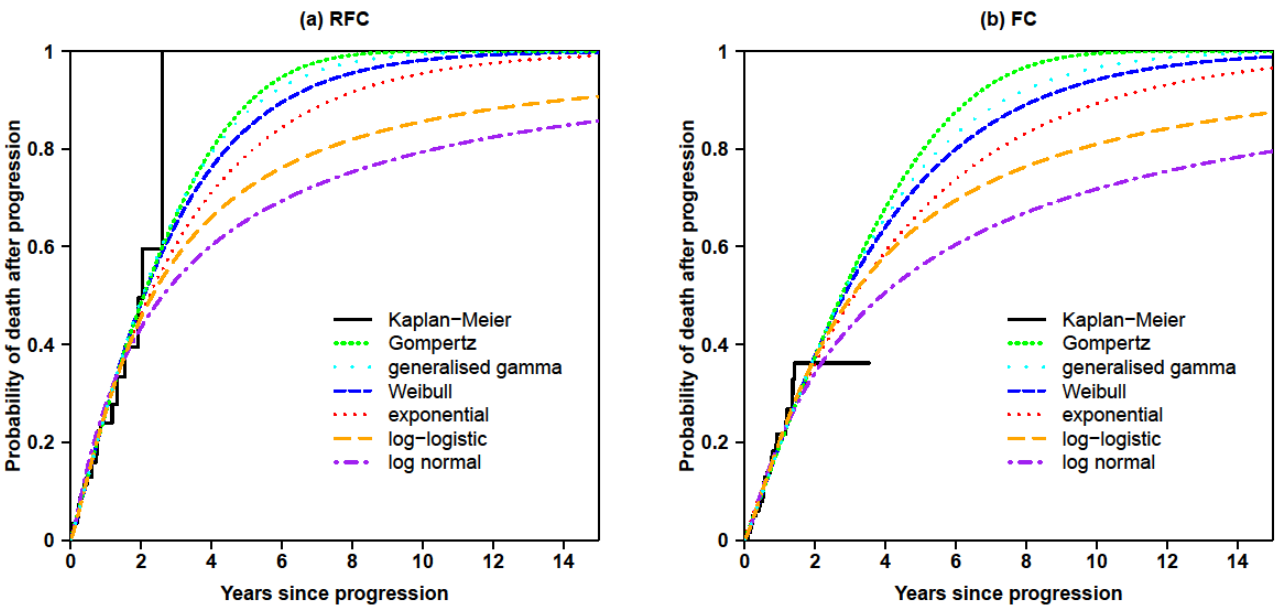


Figure 7.27 Progression→ death extrapolated to 15 years: (a) RFC and (b) FC

The Gompertz, generalised gamma and Weibull distributions each provided the required extrapolation by 15 years for both treatment arms. The exponential

distribution also provided a sensible extrapolation for the RFC arm. Given that this plot was based on death since progression, rather than from the start of the study, the Gompertz fits were probably most plausible. That was because everyone seemed to have died by 10 years, rather than taking the whole time horizon of 15 years to do so, thereby still leaving some time to be attributed to the time they would spent progression-free before reaching progression.

Choice of base-case model

Consequently, for the base-case analysis of the progression → death transition, the Gompertz distribution was chosen as it seemed to provide the best balance between the observed data and extrapolation to 15 years.

7.5.4.2 Progression-free → progression

The second transition for which the fit was assessed was progression-free → progression. This was carried out by examining the fit for the probability of being in the progression state at any given time. However this state was an intermediate state and as such had flow in (progression-free → progression) and out (progression → death) of it. To focus on comparing distributions for the transition progression-free → progression specifically, a Gompertz fit was used for progression → death throughout. Other parametric distributions for progression → death were considered in the sensitivity analyses in section 7.7.1.

The fit over the observed period of the trial and that over the extrapolated period were considered separately.

Assessment of fit over the observation period of the trial

In Figure 7.28 (a) and (b) the probability of being in progression at any given time over the observed period of the trial can be seen for RFC and FC respectively. The coloured lines each show a different distribution fitted to the progression-free → progression transition. Furthermore, each of the different shades of a colour show when a different distribution was used for progression-free → death. Therefore, given that 6 distributions were used for each

transition, each line in a particular shade of a particular colour represents one of the 36 possible ways of using the distributions for modelling the transitions.

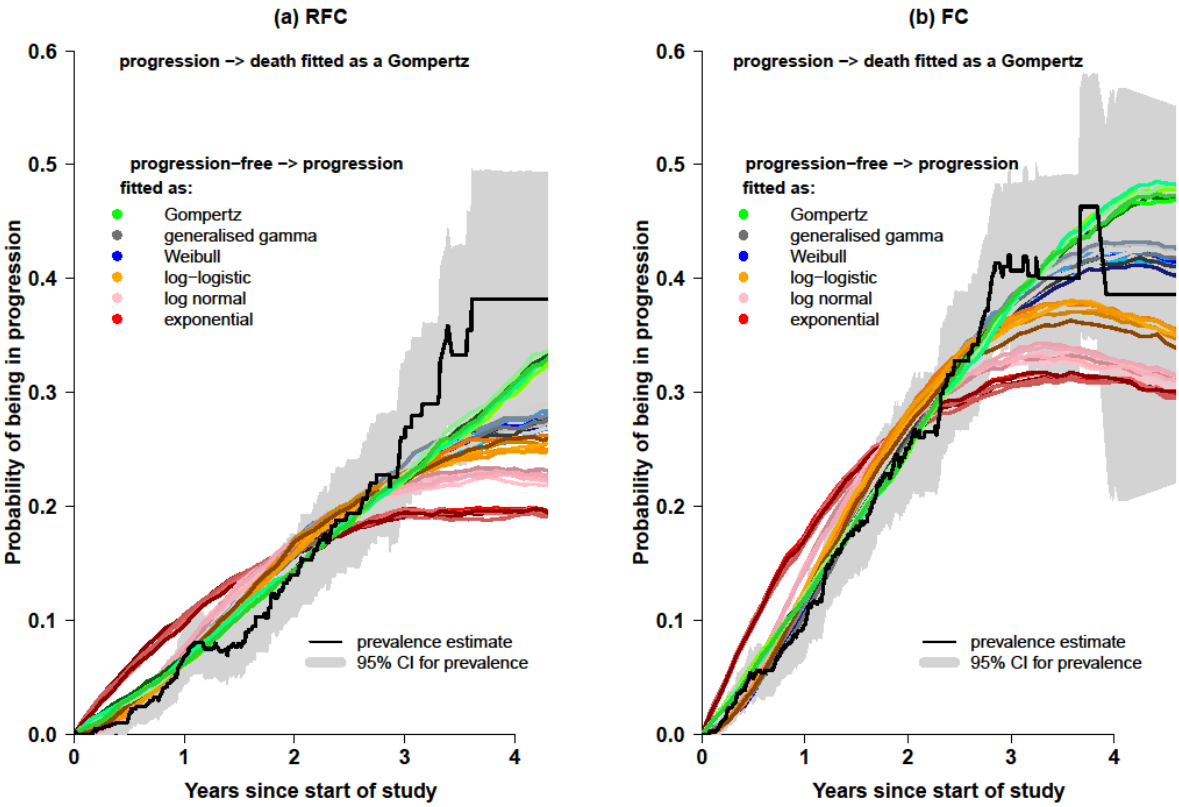


Figure 7.28 Probability of progression over trial observation period for (a) RFC and (b) FC Different distributions for progression-free →progression are distinguished by colour Shades of each colour represent a different distribution used for progression-free → death

Each of the different shades of a particular colour gave similar fits. Therefore the fit to progression-free → progression did not seem to be affected by the distribution used for progression-free → death.

Figure 7.28(a) indicates that, over the observed period of the trial, the Gompertz fit for progression-free → progression provided the best fit for the probability of progression for the RFC treatment arm. It can be seen in Figure 7.28(b) that a reasonable fit for the probability of progression for the FC arm was achieved when the Gompertz, generalised gamma and Weibull distributions were used for progression-free → progression.

Assessment of fit over the extrapolated period

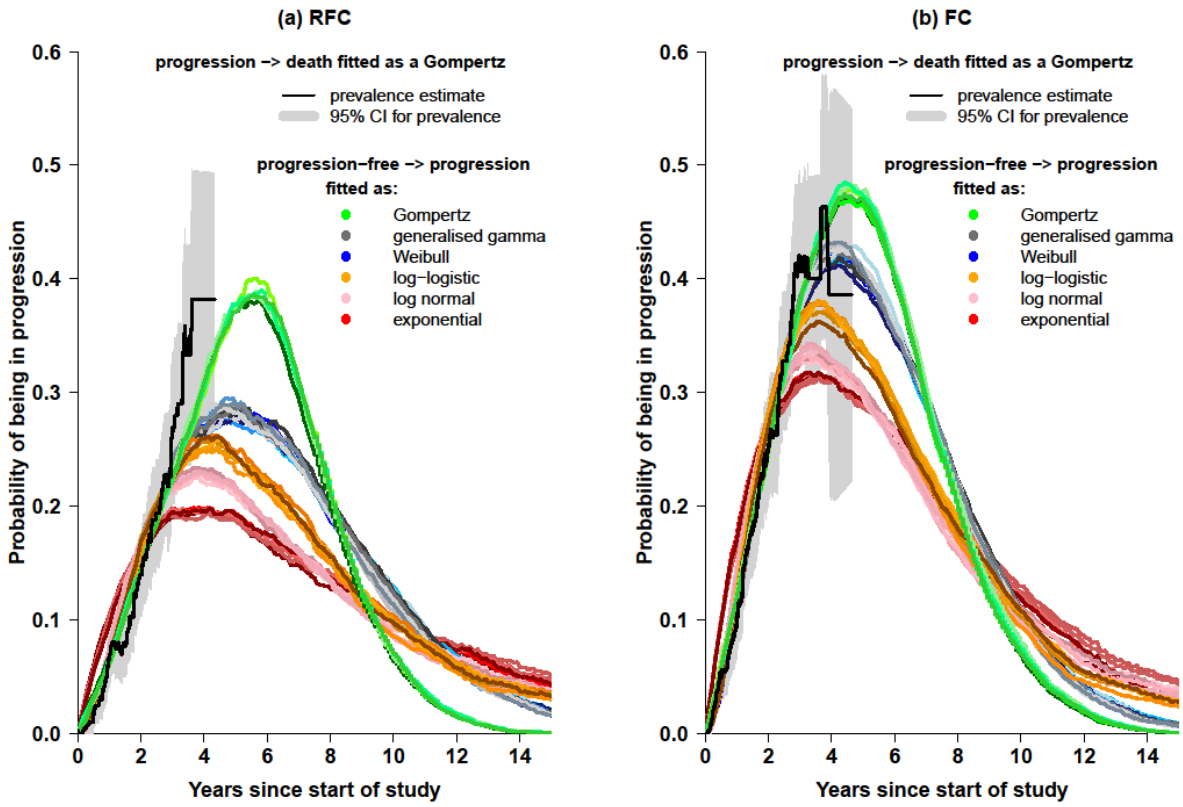


Figure 7.29 Probability of progression extrapolated to 15 years: (a) RFC and (b) FC
Different distributions for progression-free →progression are distinguished by colour
Shades of each colour represent a different distribution used for progression-free → death

Figure 7.29 (a) and (b) indicate the Gompertz fit for progression-free → progression provided the most reasonable extrapolation for each treatment arm. It can be seen in Figure 7.29 (b) that the fit for FC was also reasonable upon extrapolation with the Weibull and generalised gamma models.

Choice of base-case model

For the base-case analysis of the progression-free → progression transition, the Gompertz distribution was chosen as it seemed to provide the best balance between the fit to the observed data and the extrapolation to 15 years.

7.5.4.3 Progression-free → death (without progression)

The third and final transition that was assessed was progression-free → death. To focus on comparing the fits of different distributions for this transition specifically, the progression → death and progression-free → progression transitions were fitted using a Gompertz distribution throughout. Alternative fits for each of the transitions are considered in the sensitivity analysis presented in section 7.7.1.

Figure 7.30 (a) and (b) show, for RFC and FC respectively over the observed period of the trial, the competing risks (CR) cumulative incidence estimate of progression-free → death and that predicted by parametric survival regressions using standard distributions. It can be seen in Figure 7.30 (a) the generalised gamma fit probably provided the best fit over the observed period for RFC, although the Gompertz provided a very similar fit over the first 2 years. Figure 7.30 (b) shows the Gompertz and log normal fits seemed to be the most suitable over the observed period for FC.

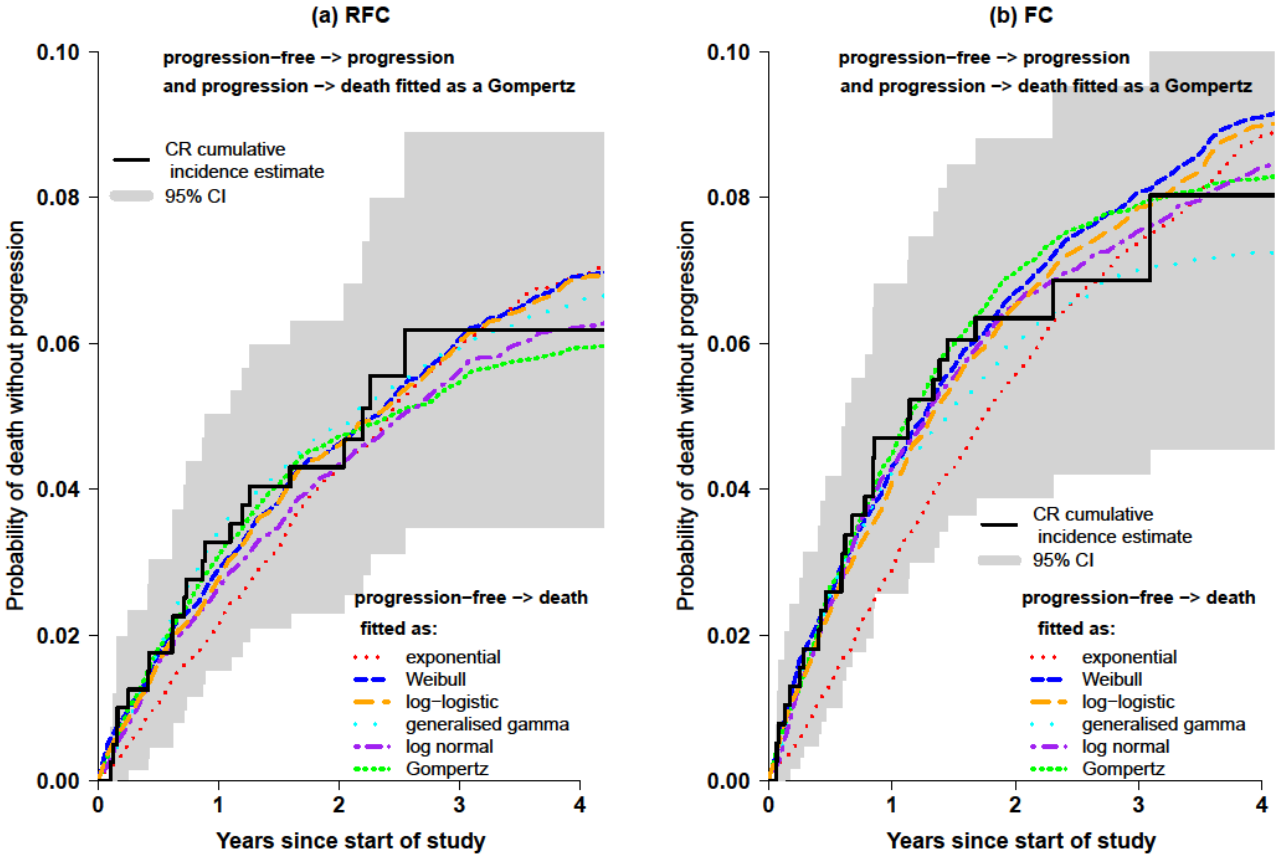


Figure 7.30 Progression-free → death without progression over the trial observation period: Gompertz used for progression → death and progression-free → progression abbreviation used: CR=competing risks

Figure 7.31 shows the same information as Figure 7.30, only this time the estimates have been extrapolated to 15 years.

Figure 7.31 (a) and (b) show that the fits over the extrapolated period all seemed to be flat. Whilst this does not seem very realistic, indicating that the vast majority of patients had reached progression and possibly died after progression over the period, the generalised gamma and Gompertz were probably the best fit of all the distributions considered for RFC as they also provided the best fits over the observed period. Similarly, the Gompertz and log normal fits probably provided the best extrapolations for FC. Taking all this into account, the Gompertz was chosen for the base-case analysis for progression-free → death with progression, with alternative distributions considered in the sensitivity analysis in section 7.7.1.

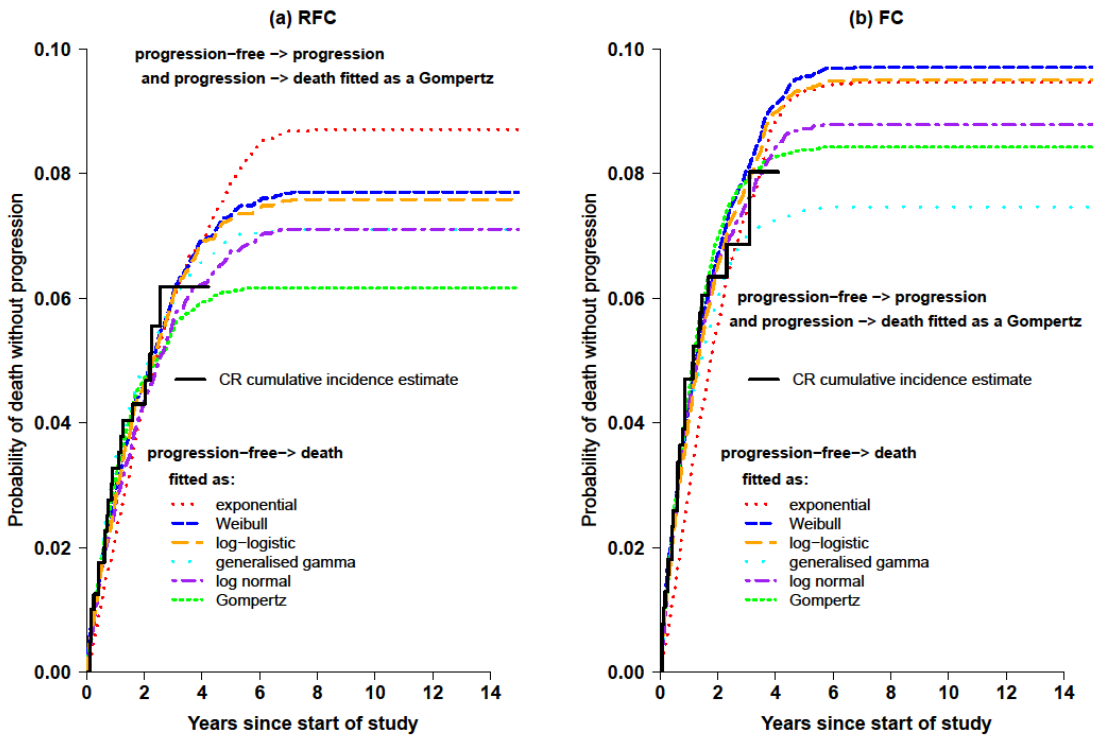


Figure 7.31 Progression-free → death without progression extrapolated to 15 years
(a) RFC (b) FC
Gompertz used for progression → death and progression-free → progression
abbreviation used: CR=competing risks

7.5.5 Summary of base-case model for each transition

Table 7.20 shows, for each transition, a summary of the Gompertz models used for the base case analysis.

Alternative distributions for each transition are considered in the sensitivity analysis presented in section 7.7.1.

	coefficient	s.e.	HR (95% CI)	p-value
progression-free→ progression				
treatment (RFC vs FC)	-0.558	0.128	0.572 (0.446, 0.735)	<0.001
constant	-2.187	0.130		
gamma	0.474	0.068		
progression-free → death				
treatment (RFC vs FC)	-0.343	0.294	0.710 (0.399, 1.262)	0.243
constant	-2.825	0.265		
gamma	-0.487	0.207		
progression → death				
treatment (RFC vs FC)	0.342	0.285	1.408 (0.806, 2.461)	0.229
constant	-1.627	0.267		
gamma	0.174	0.244		

Table 7.20 Base case models (all Gompertz) for the three transitions using the multi-state modelling approach

7.6 Comparison of the three approaches

In this section the results using the Markov decision-analytic model, partitioned survival and multi-state modelling are compared. It is based largely on a previous version of a similar comparison that was presented in Williams et al. (2017b).

Firstly, sub-section 7.6.1 visually assesses the fits using each of the approaches. Secondly, the survival estimates in terms of mean Life Years and QALYs are compared (sub-section 7.6.2). Thirdly, the costs used in each approach are described in sub-section 7.6.3. Finally, sub-section 7.6.4 summarises and compares the incremental cost-effectiveness ratios.

7.6.1 Visual assessment of the fits

In this sub-section each of the modelling approaches are compared by informally visually assessing the fits. The assessment was based on a balance of achieving a good fit to the observed data and extrapolation that represented a lifetime since entry into the trial of 15 years.

7.6.1.1 progression-free state

Figure 7.32 shows the predictions of being in the progression-free state using each of the approaches.

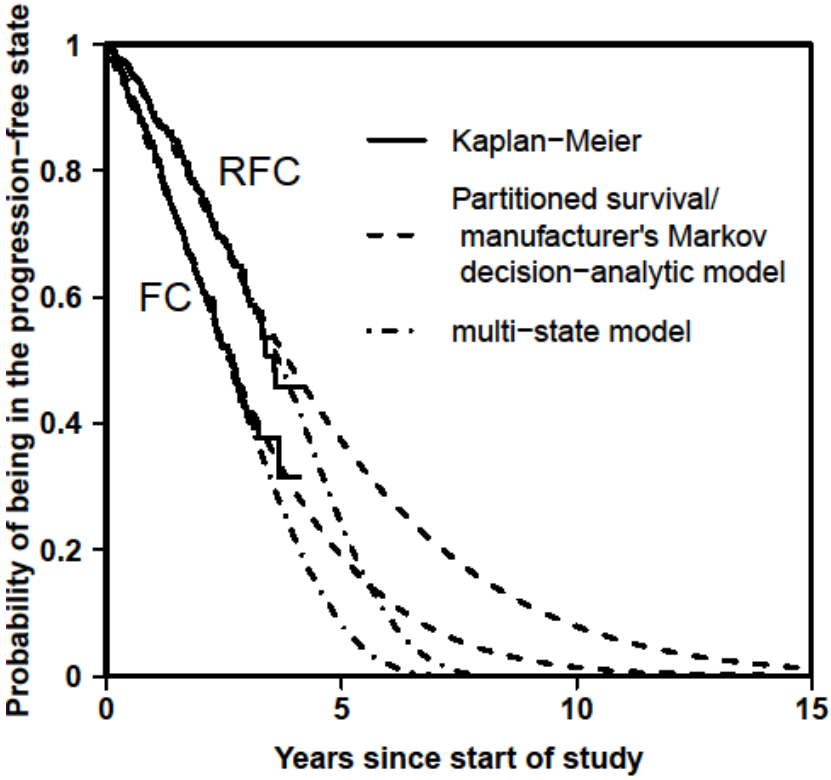


Figure 7.32 Probability of being in the progression-free state using each approach

The partitioned survival approach and the Markov decision-analytic modelling were based on the same fit to progression-free survival. For both treatment arms, each of the three approaches appeared to be well-fitting to the Kaplan-Meier estimate. However, the extrapolation using the multi-state modelling approach reached zero somewhat earlier than the other approaches. This seems plausible as it allows those patients who reach the progression state to spend some time there before reaching the end of their lives.

7.6.1.2 progression state

Figure 7.33 (a) and (b) show, for the RFC and FC arm respectively, the predictions of being in the progression state with each of the approaches.

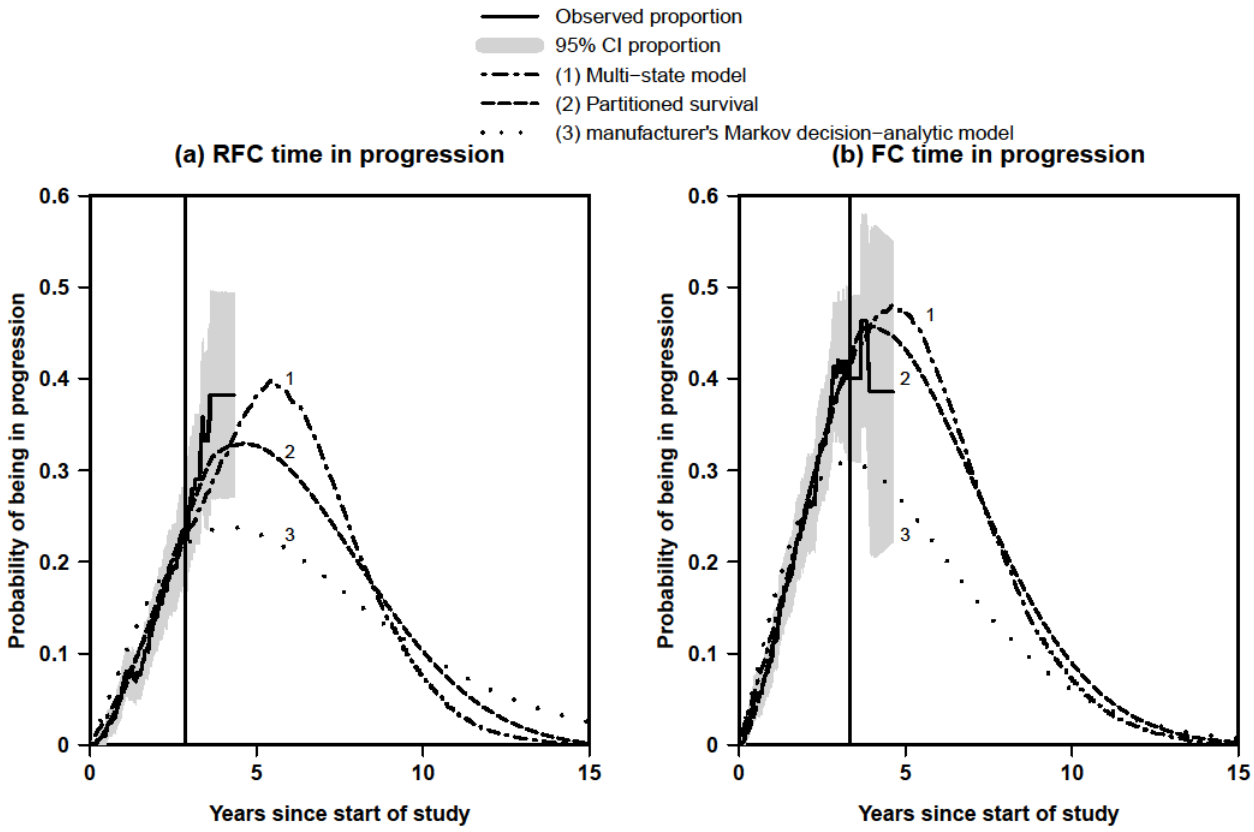


Figure 7.33 Probability of being in the progression state using each approach (a) RFC (b) FC

The vertical solid lines on Figure 7.33 (a) and (b) are positioned where there were less than 20 patients at risk of death after progression, an indication of more uncertainty in the observed proportion estimates. They therefore provide a dividing line between the periods of observation and extrapolation. To further emphasise the uncertainty, the shaded areas show the 95% confidence intervals for the observed proportions. They were created using 5000 bootstrapped samples.

For the RFC arm, the fit to the observed data was good from all approaches up until the vertical line (Figure 7.33 (a)). Subsequently however, there was a marked difference between the approaches in where and when the predictions of being in progression peaked. Furthermore, the partitioned survival approach and the multi-state modelling were the only methods to reach zero by 15 years.

For the FC arm, the fit to the observed data from the partitioned survival approach and multi-state modelling was good fit up until the vertical line (Figure 7.33 (b)). However, this could only be said of the Markov decision-analytic modelling up to 2.3 years. Again, there was a marked difference between the approaches in the peaks. Furthermore, all approaches reached zero by 15 years as required.

7.6.1.3 death state

Figure 7.34 (a) and (b) show the Kaplan-Meier estimates of being in the death state, together with the predictions from each of the modelling approaches, for the RFC and FC arms respectively.

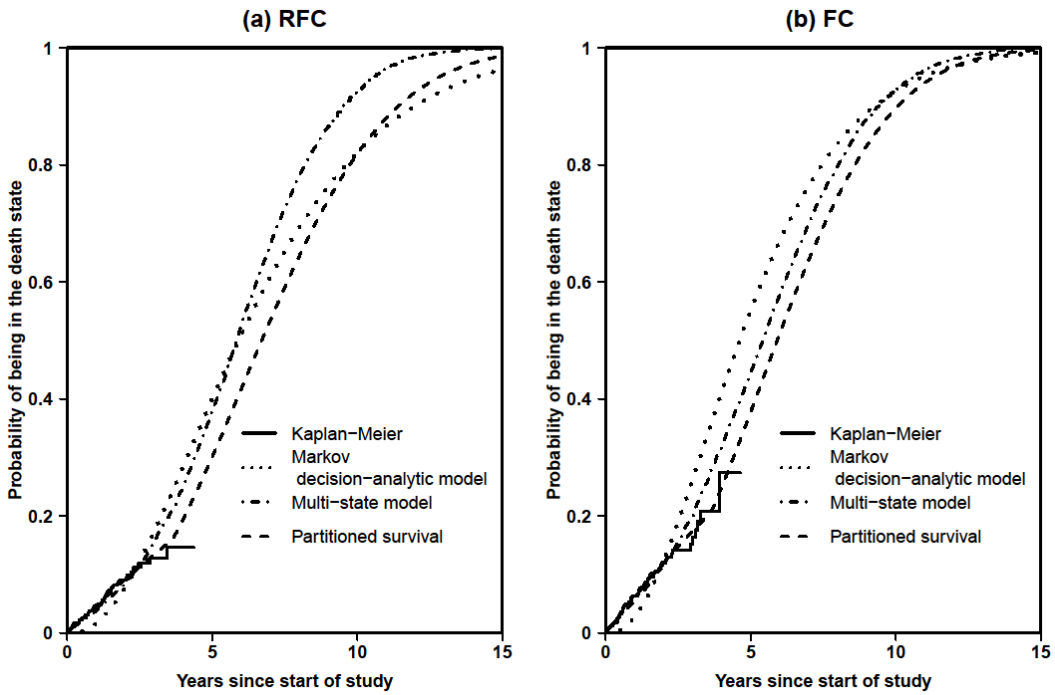


Figure 7.34 Probability of being in the death state using each approach (a) RFC (b) FC

The multi-state modelling and partitioned survival approach fitted the observed data over the first 4 years reasonably well. However, the Markov decision-analytic modelling underestimated death over the first 2 years. All approaches

reached a probability of one by 15 years for FC (Figure 7.34 (b)). However the multi-state modelling was the only approach to do so for RFC (Figure 7.34 (a)).

7.6.1.4 death without progression

Figure 7.35 shows, for each treatment arm, the competing risk cumulative incidence estimate of death without progression together with the predictions from the multi-state modelling.

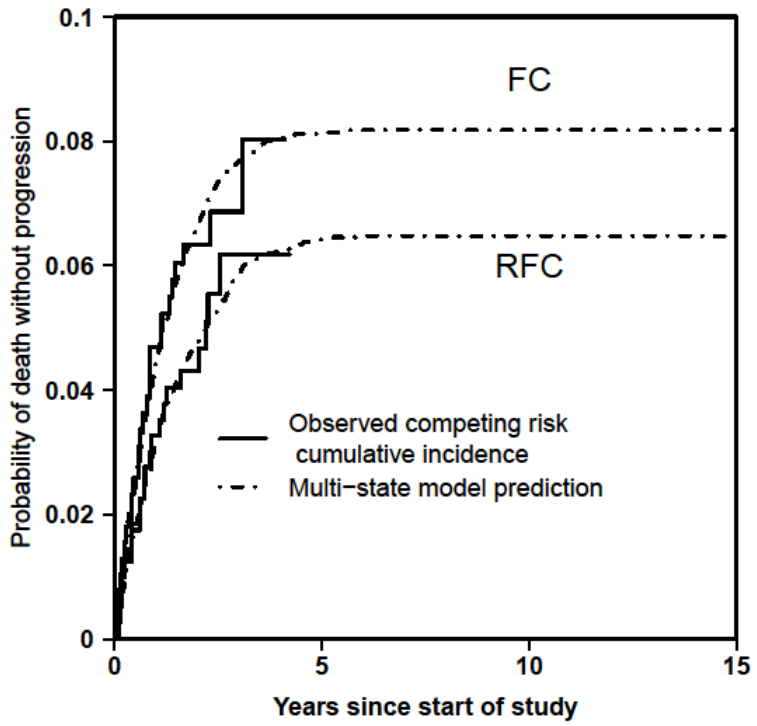


Figure 7.35 death without progression

It can be seen that the multi-state modelling fitted the observed data fairly well. These predictions were possible because the multi-state modelling allowed the state occupancy probabilities of death to be split into death without progression and death after progression. An equivalent Figure from the Markov decision-analytic model was not available.

Furthermore, the partitioned survival approach did not model death without progression.

7.6.1.5 death after progression

Figure 7.36 shows, for each treatment arm, the Kaplan-Meier estimate of death once in the progression state, together with the predictions from the multi-state modelling and the manufacturer's assumption.

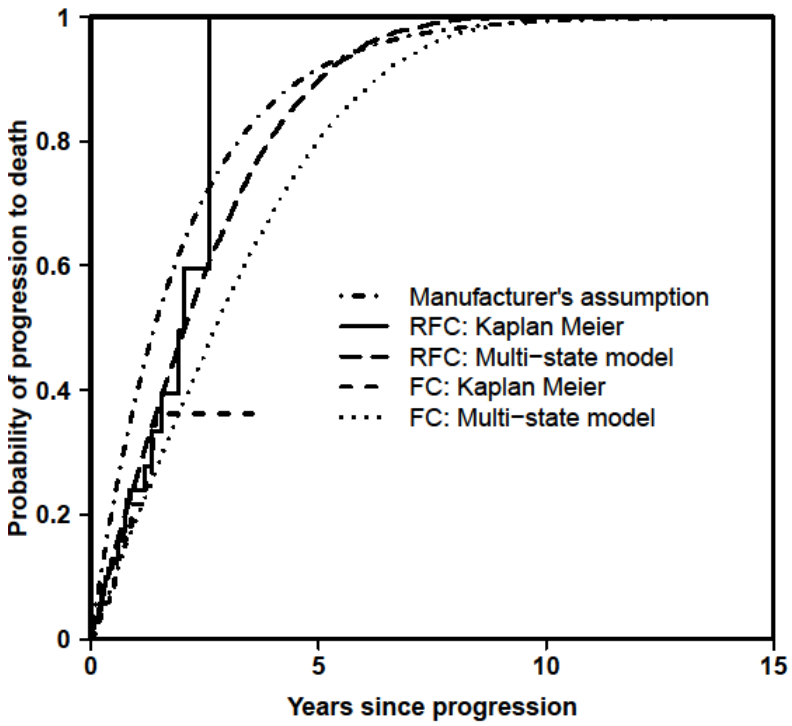


Figure 7.36 post-progression death

Each of the predictions from the multi-state modelling were close to their corresponding Kaplan-Meier estimate. It can be seen that the extrapolation for both methods was fairly good at reaching one by 15 years. However, the manufacturer's assumption was less well-fitting to each of the treatment arms.

The partitioned survival approach did not model death after reaching the progression state.

7.6.2 Mean Life Years/QALYs

Table 7.21 shows the mean Life Years and QALYs results for each relevant health state using each of the three approaches. Mean QALYs were calculated by assuming a utility of 0.8 for the time spent progression-free and 0.6 for the time

spent in progression, the approach used by the manufacturer in the Markov decision-analytic model. All of the information in the table was discounted at 3.5%.

	Markov decision – analytic modelling			Partitioned survival			Multi-state modelling		
	RFC	FC	Incre-mental	RFC	FC	Incre-mental	RFC	FC	Incre-mental
Mean Life Years	5.73	4.65	1.07	5.96	5.31	0.65	5.24	4.96	0.28
Mean Life Years Progression-free	4.11	2.93	1.18	4.10	2.92	1.18	3.35	2.55	0.81
Mean Life Years in Progression	1.62	1.73	-0.11	1.86	2.39	-0.53	1.89	2.42	-0.53
Mean QALYs	4.26	3.38	0.88	4.40	3.77	0.63	3.82	3.49	0.33
Mean QALYs Progression-free	3.29	2.34	0.94	3.28	2.34	0.95	2.68	2.04	0.65
Mean QALYs in Progression	0.97	1.04	-0.07	1.11	1.43	-0.32	1.13	1.45	-0.32

Table 7.21 Mean Life Years and QALYs results: all three approaches

It can be seen in Table 7.21 that the mean Life Years/QALYs Progression-free results were similar for the partitioned survival and Markov decision-analytic modelling approaches, with the multi-state modelling having a smaller benefit.

Furthermore, each approach found a decrement in mean Life Years/QALYs in Progression. The decrements were largest with the partitioned survival and multi-state modelling approaches with mean Life Years (QALYs) of -0.53 (-0.32). The corresponding decrement in mean Life Years (QALYs) of -0.11 (-0.07) with the Markov decision-analytic modelling was somewhat smaller. This was due to the assumption of no treatment effect used for the progression → death transition in that model.

In terms of mean Life Years/QALYs overall, the largest benefit was seen with the Markov decision-analytic modelling due to the smallest decrement in time in progression with that approach. In contrast, the benefits were far smaller with

the multi-state modelling approach due to the combination of having the smallest benefit Progression-free and the largest decrement whilst in Progression.

7.6.3 Description of the costs

Table 7.22 shows a breakdown of the mean costs used with each of the three approaches. Most of the mean costs were not related to the time spent in relevant health states. However, cost of supportive care whilst in PFS, cost of supportive care whilst in progression and cost of 2nd line and subsequent therapy whilst in progression were all associated with time spent in relevant states. Therefore, the mean life years in the appropriate states were used in the calculation of these costs. All other costs were taken from the original manufacturer's submission [Roche (2008), pp127-131].

	Markov decision-analytic modelling			Partitioned survival			Multi-state modelling		
	RFC	FC	Incremental	RFC	FC	Incremental	RFC	FC	Incremental
Mean cost of PFS	**18965**	**6891**	**12074**	**18963**	**6890**	**12074**	**18711**	**6764**	**11947**
Costs of Rituximab	10113	0	10113	10113	0	10113	10113	0	10113
Administration costs of Rituximab	1224	0	1224	1224	0	1224	1224	0	1224
Cost of Fludarabine	2776	2790	-14	2776	2790	-14	2776	2790	-14
Administration costs of Fludarabine	1109	1115	-6	1109	1115	-6	1109	1115	-6
Costs of Cyclophosphamide	21	22	0	21	22	0	21	22	0
Administration costs of Cyclophosphamide	1109	1115	-6	1109	1115	-6	1109	1115	-6
Cost of supportive care in PFS	1381	983	398	1379	982	397	1127	855	272
Cost of Bone Marrow Transplantation	592	360	231	592	360	231	592	360	231
Cost of Blood Transfusions	640	507	133	640	507	133	640	507	133
Mean cost of Progression	**6630**	**7088**	**-458**	**6406**	**8233**	**-1827**	**6450**	**8248**	**-1798**
Cost of supportive care in progression	1630	1742	-111	1873	2407	-534	1905	2436	-531
Cost of 2nd line and subsequent therapy	5001	5344	-340	4533	5825	-1293	4546	5812	-1267
Mean Total Cost	**25595**	**13978**	**11617**	**25369**	**15122**	**10247**	**25161**	**15012**	**10149**

Table 7.22 Costs breakdown: all three approaches

7.6.4 Incremental cost-effectiveness ratios with each approach

Table 7.23 shows the incremental cost-effectiveness ratio that resulted with each approach. It can be seen in Table 7.23 that the partitioned survival and Markov decision-analytic modelling deemed the RFC treatment cost-effective. However when the multi-state modelling approach was used, the Cost per QALY gained was £30,702, in excess of the commonly used willingness to pay threshold of £20,000 - £30,000 per QALY gained.

	Markov decision-analytic modelling			Partitioned survival			Multi-state modelling		
	RFC	FC	Incre-mental	RFC	FC	Incre-mental	RFC	FC	Incre-mental
Mean Life Years	5.73	4.65	1.07	5.96	5.31	0.65	5.24	4.96	0.28
Mean QALYs	4.26	3.38	0.88	4.40	3.77	0.63	3.82	3.49	0.33
Mean Total Cost	£25,595	£13,978	£11,617	£25,369	£15,122	£10,247	£25,161	£15,012	£10,149
Cost per Life Year Gained			**£10,825**			**£15,696**			**£36,049**
Cost per QALY gained			**£13,189**			**£16,310**			**£30,702**

Table 7.23 Incremental cost effectiveness ratios: all three approaches

To provide more insight into how the approaches compared, the mean Life Years/QALYs results in Table 7.23 were split into those over the observed period (Table 7.24) and those over the unobserved extrapolation period (Table 7.25).

It can be seen in Table 7.24 that the approaches were reasonably comparable over the observed period of the trial. However, in the unobserved extrapolation period (Table 7.25) there was an increment in mean Life Years/QALYs gained while in progression with the Markov decision-analytic modelling. This was in contrast to the rest of the approaches which found decrements.

	Markov decision – analytic modelling			Partitioned survival			Multi-state modelling		
	RFC	FC	Incre-mental	RFC	FC	Incre-mental	RFC	FC	Incre-mental
Mean Life Years	3.42	3.27	0.16	3.42	3.32	0.10	3.40	3.28	0.12
Mean Life Years Progression-free	2.85	2.44	0.41	2.85	2.42	0.43	2.86	2.37	0.49
Mean Life Years in Progression	0.57	0.82	-0.25	0.57	0.90	-0.33	0.53	0.91	-0.37
Mean QALYs	2.62	2.45	0.18	2.62	2.48	0.15	2.61	2.44	0.17
Mean QALYs Progression-free	2.28	1.95	0.33	2.28	1.94	0.34	2.29	1.90	0.39
Mean QALYs in Progression	0.34	0.49	-0.15	0.34	0.54	-0.20	0.32	0.54	-0.22

Table 7.24 Mean Life Years and QALYs: trial observation period of 0-4 years

	Markov decision– analytic modelling			Partitioned survival			Multi-state modelling		
	RFC	FC	Incre-mental	RFC	FC	Incre-mental	RFC	FC	Incre-mental
Mean Life Years	2.30	1.39	0.92	2.54	1.99	0.55	1.85	1.68	0.16
Mean Life Years Progression-free	1.26	0.48	0.78	1.26	0.50	0.75	0.49	0.17	0.32
Mean Life Years in Progression	1.05	0.90	0.14	1.29	1.49	-0.20	1.36	1.51	-0.15
Mean QALYs	1.63	0.93	0.71	1.78	1.29	0.48	1.21	1.04	0.16
Mean QALYs Progression-free	1.01	0.39	0.62	1.00	0.40	0.60	0.39	0.14	0.25
Mean QALYs in Progression	0.63	0.54	0.08	0.77	0.89	-0.12	0.81	0.91	-0.09

Table 7.25 Mean Life Years and QALYs: extrapolation over 4-15 years

7.7 Sensitivity analyses for the multi-state modelling approach

In this section, results of sensitivity analyses are presented to check the robustness of the multi-state model from deviations from the base-case assumptions. Only sensitivity analyses for the multi-state modelling are considered, it being the main focus of this thesis. The section is based largely on the sensitivity analyses presented in Williams et al. (2017a). However, the data underlying the analyses were different. This was because the published tutorial paper was not based on the actual trial data but on data digitised from published curves.

Three sub-sections of sensitivity analyses are shown. In sub-section 7.7.1 alternative distributions for each of the transitions are considered. Sub-section 7.7.2 explores other one-way sensitivity analyses and sub-section 7.7.3 presents the results of probabilistic sensitivity analyses.

7.7.1 Using alternative distributions for each of the transitions in the multi-state modelling

When building the multi-state models, six parametric distributions were considered for each of the three transitions resulting in 6^3 = 216 combinations of distributional fits to assess. Table A1 in the Appendix IX shows the results – in terms of incremental QALYs, incremental costs and Cost per QALYs gained – for each of these combinations. The results are shown sorted by Cost per QALYs gained, with the exception of the base case model which is shown first.

It can be seen in Table A1 that the Cost per QALYs gained ranged from £11,393 to £344,585. However, many of the combinations did not provide a reasonable fit to at least one transition for at least one treatment arm. When this was the case, the Cost per QALYs were shaded in grey. The scenario each particular shade represents are described in the remaining part of this section. In common with earlier sections of this chapter, fits were evaluated by considering both the observed fit to the data and whether a reasonable extrapolation of survival was achieved in terms of reaching zero by 15 years.

Figure 7.25 (a) and (b) on page 190 showed, for the RFC and FC treatment arms respectively, that there was little to choose between the distributions for progression → death over the observed period of the trial. However Figure 7.27 (a) and (b) on page 192 showed that the log-logistic and log normal fits did not provide sensible extrapolations for either arm. Therefore the Costs per QALYs gained in Table A1 for each combination that used either a log-logistic or log normal distribution for progression → death were shaded in grey.

Only the remaining four distributions were considered any further for progression → death. Figure 7.37, focusing on the RFC arm, shows the probability of being in the progression state at any given time, with progression → death fitted using (a) Gompertz, (b) exponential, (c) Weibull and (d) generalised gamma.

Figure 7.38 shows the equivalent information for the FC arm.

Figure 7.37 and Figure 7.38 show that fitting progression-free → progression using the Gompertz distribution provided the best fit for the probability of progression upon extrapolation, for the RFC and FC arm respectively, regardless of the distribution used for progression → death. Furthermore, the Gompertz was the only distribution that, when fitted to progression → death, resulted in an extrapolation to zero by 15 years for the probability of being in the progression state. When a Gompertz was used for progression → death, the Weibull and generalised gamma also provided sensible extrapolations for the probability of being in the progression state, but only for the FC arm. In addition, the distribution used for progression-free → death had very little influence on the fits, as indicated by the very similar curves for the different shades of a particular colour.

Consequently, the Cost per QALYs gained in Table A1 for each combination that did not use a Gompertz for progression-free → progression or a Gompertz for progression → death - that was not already highlighted in grey – was highlighted in light grey.

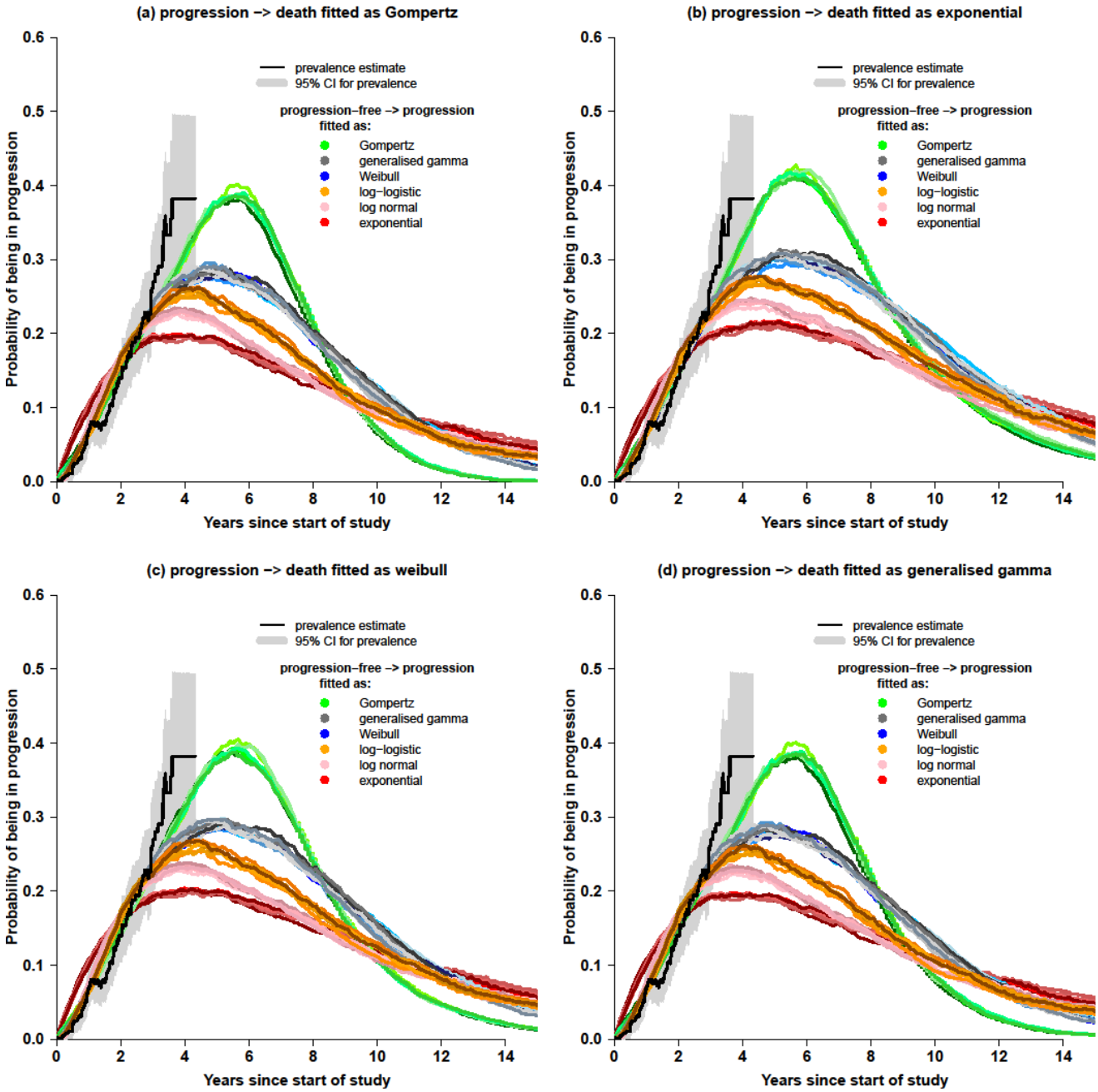


Figure 7.37 **Probability of progression for RFC: extrapolation to 15 years**

Progression → death fitted using: (a) Gompertz, (b) exponential (c) Weibull and (d) generalised gamma

Different distributions for progression-free →progression are distinguished by colour

Shades of each colour represent a different distribution used for progression-free → death

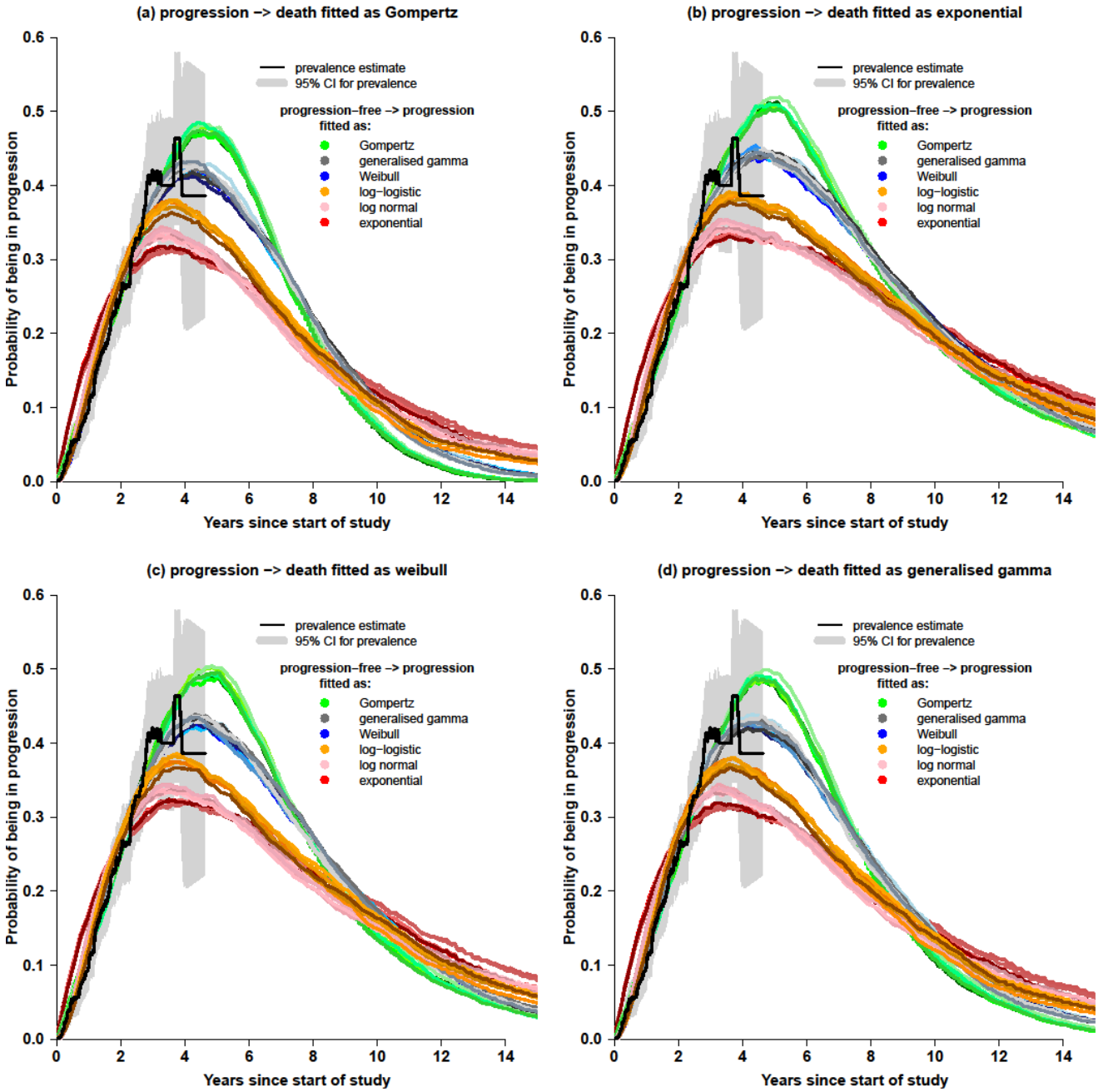


Figure 7.38 Probability of progression for FC: extrapolation to 15 years

Progression → death fitted using: (a) Gompertz, (b) exponential (c) Weibull and (d) generalised gamma

Different distributions for progression-free →progression are distinguished by colour

Shades of each colour represent a different distribution used for progression-free → death

The only Costs per QALYs gained in Table A1 that were not highlighted either in grey or light grey were those that involved a Gompertz fit for progression-free → progression and progression → death. For this scenario, the effect of the

distribution used to fit progression-free → death on the probability of being in the progression-free and progression states – the two states relevant for the calculation of the Cost per QALYs gained - was considered.

Figure 7.39 shows the aforementioned effect on the probability of remaining in the progression-free state (a) over the observed period of the trial and (b) upon extrapolation to 15 years. Figure 7.39 (c) and (d) show the corresponding information for the effect on the probability of being in the progression state. Each of the coloured lines in the plots represent a different parametric distribution used to fit progression-free → death.

It can be seen from each of the four plots in Figure 7.39 that reasonable (similar) fits for the probability of being in the aforementioned states were achieved regardless of the distribution used for progression-free → death.

Because the non-highlighted Cost per QALYs gained in Table A1 could not be disregarded, in the sense they did not result from unreasonable fits for the transitions/states, the robustness of the fit for progression-free → death was considered as part of the one-way sensitivity analyses in the following sub-section.

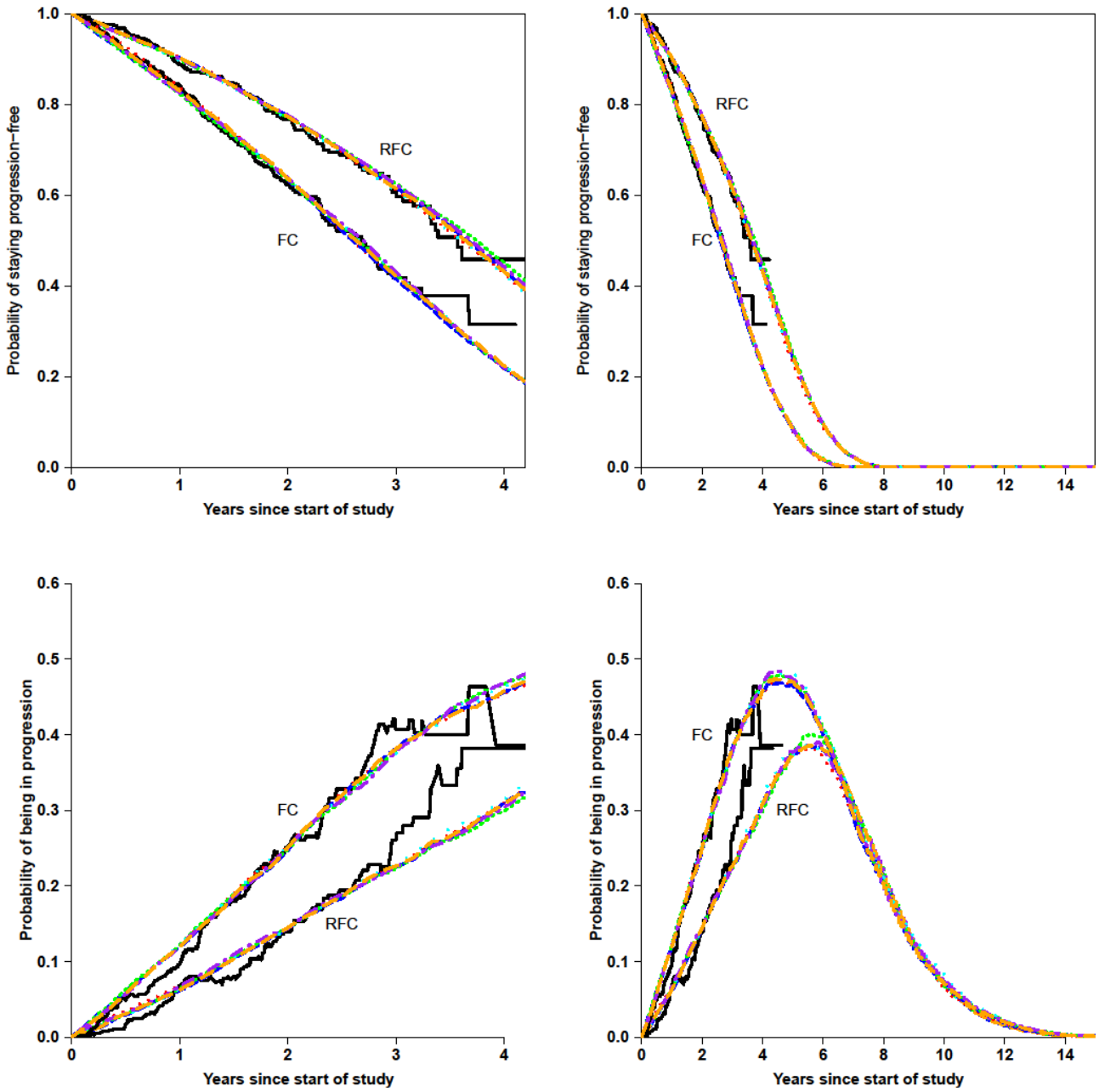


Figure 7.39 **Effect of fit to progression-free → death on probability of being in the:**

(a) progression-free state over the observed period of the trial

(b) progression-free state extrapolated to 15 years

(c) progression state over the observed period of the trial

(d) progression state extrapolated to 15 years

7.7.2 One-way sensitivity analyses in multi-state modelling

In this sub-section results of one-way sensitivity analyses are presented in a tornado plot. The sensitivity analyses carried out are described below:

1) To allow computation of the Gompertz cumulative hazards, the time points used in the base-case calculation (measured in years) were in 1/12 increments until 7.5 years, followed by 1/144 increments until 12 years and then 1/600 to 15 years. The one-way sensitivity analyses included varying the number of time points. An increase, equivalent to using a shorter cycle length, was considered by using 1/144 increments between 7.5 and 11 years and then 1/600 to 15 years. Less time points were also considered by using 1/12 increments up to 118/12 (9.83 years) followed by the same increments as the base-case analysis.
2) In the base-case analysis it was assumed that the treatment effect observed during the trial persisted to the time horizon. Two sensitivity analyses are presented where the treatment effect no longer persists in the extrapolation period. Firstly, the probability in the extrapolation period for one treatment was calculated by applying a hazard ratio of 1 to the probability for the other treatment. Alternatively, the model used in the extrapolation period did not include treatment as a covariate.
3) The previous sub-section found there were combinations of distributions that could not be disregarded due to a poor fit for a state or a transition. This was the case when Gompertz distributions were used for the progression-free → progression and progression → death transitions. Therefore, the sensitivity analyses include varying the distribution used for progression-free → death when the other two transitions were fitted using Gompertz distributions.
4) The base-case analysis had a 15-year time horizon. An analysis was carried out with a 20-year time horizon.
5) The remaining sensitivity analyses were based on the one-way sensitivity analyses carried out by the manufacturer as part of their economic modelling [Roche (2008),p145]

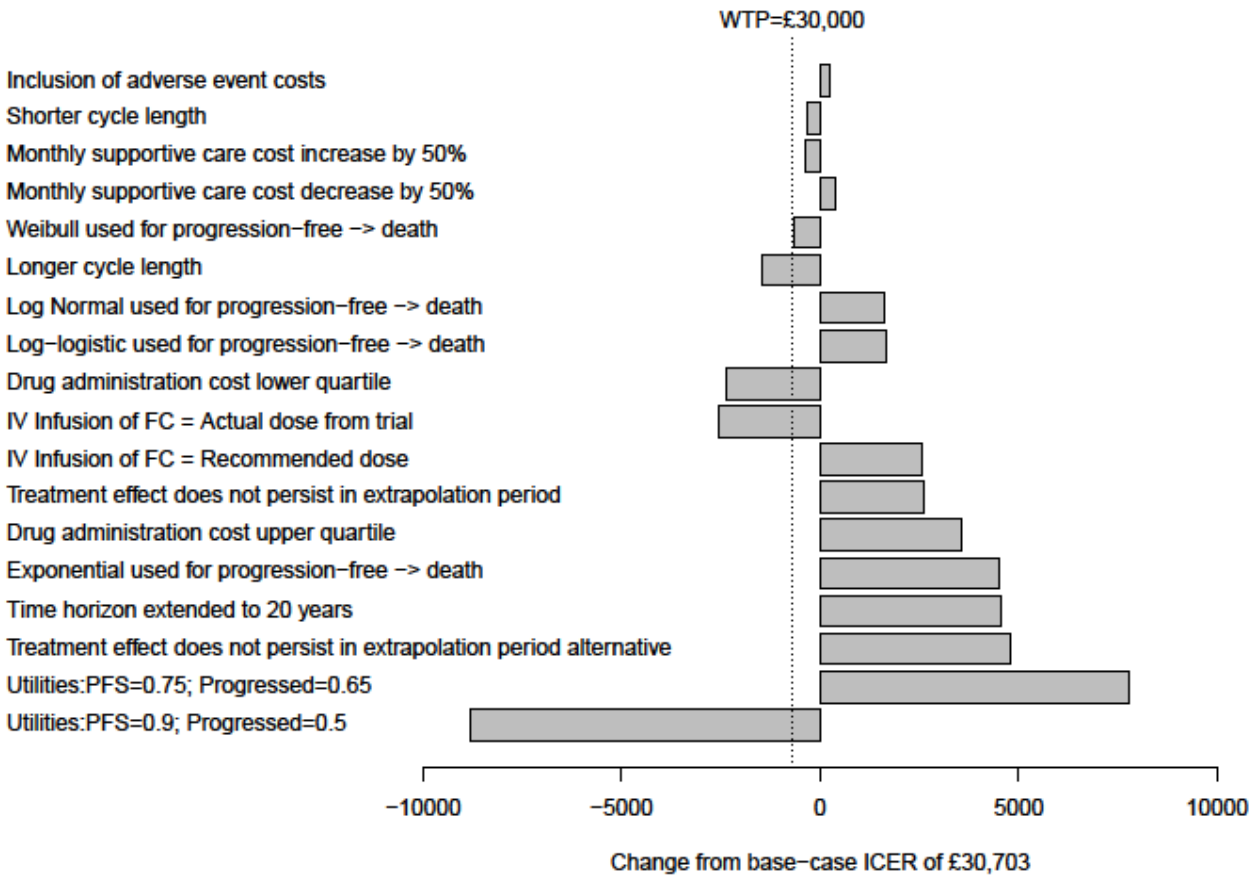


Figure 7.40 Tornado plot: one-way sensitivity analyses

Figure 7.40 shows that varying the utility conveyed the most uncertainty with an incremental cost-effectiveness ratio (ICER) £7768 higher and £8832 lower than the base-case, when the gap between the utilities for the health states narrowed and widened respectively. Furthermore, only seven of the eighteen one-way sensitivity analyses lead to decreases from the base-case ICER. In particular, only four of those led to ICERs that did not exceed the commonly used threshold of £30,000 per QALY gained.

7.7.3 Probabilistic sensitivity analysis for multi-state modelling

A probabilistic sensitivity analysis was carried out for clinical effectiveness resulting in state occupancy probabilities for each draw. In addition, each of the cost parameters were assumed to follow Beta Pert distribution in line with the

economic modelling carried out by the manufacturer. Table 7.26 shows the mean base-case estimates together with the ranges used to generate the distributions. The particular Beta Pert distribution chosen for the cost of monthly supportive care and 2nd line and subsequent therapy whilst in progression was dependent on the mean life years in progression. All other distribution parameters values were as presented by the manufacturer [Roche (2008), pp137-138].

Costs	Base case	Minimum	Maximum
Monthly supportive care cost whilst in PFS	£28	£14	£42
Monthly supportive care and 2nd line and subsequent therapy cost whilst in progression	£259.89	£129.94	£389.83
Administration - Deliver exclusively Oral Chemotherapy	£280	£174	£482
Administration - Deliver complex Chemotherapy, including prolonged infusional treatment at first attendance	£430	£210	£795
Bone marrow transplant	£47,565.05	£34,318.25	£54,646.47
Blood transfusion	£289.73	£173.84	£405.62
1 Unit of blood	£161.11	£96.67	£225.26

Table 7.26 Beta Pert distributions used in probabilistic sensitivity analysis for cost parameters

In Figure 7.41 the cost-effectiveness plane can be seen. The probabilistic sensitivity analysis involved 1000 draws with 10% excluded due to computational difficulties. All draws resulted in the RFC treatment being more costly than FC, therefore only the NW and NE quadrants are shown.

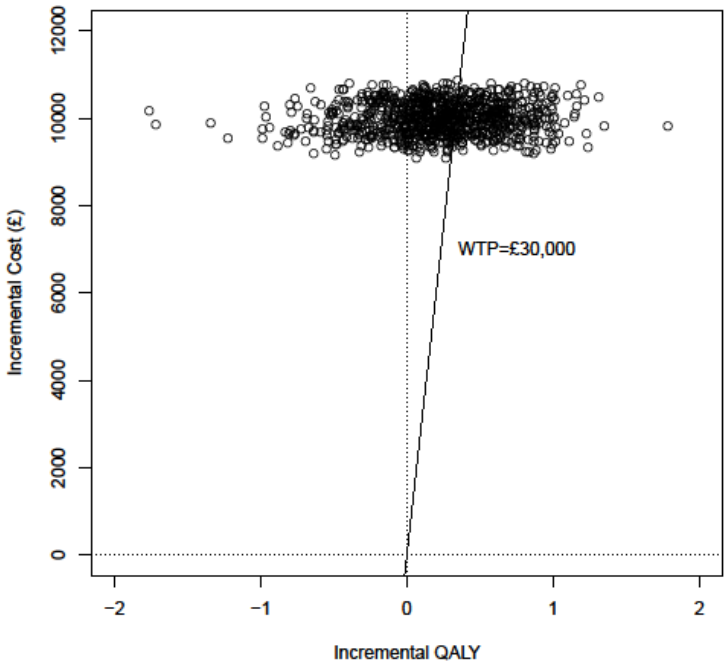


Figure 7.41 Cost-effectiveness plane

A cost-effectiveness acceptability curve is shown in Figure 7.42.

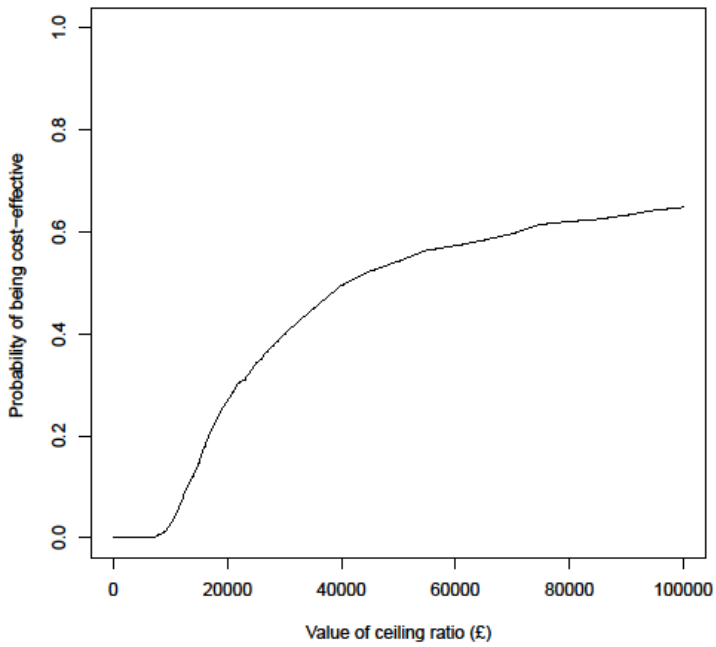


Figure 7.42 Cost-effectiveness acceptability curve

It can be seen in Figure 7.42 that, given the common range in the UK of a willingness to pay of £20,000 - £30,000 per QALY gained, RFC only had a

probability of 0.27-0.40 of being cost-effective. Even at a willingness to pay of £100,000 the probability that RFC was cost-effective compared to FC was only 0.65.

7.8 Direct comparison of the manufacturer's Markov decision-analytic model with multi-state modelling

In the analysis in this chapter up to now, different assumptions were used with regards to transition probabilities/hazards with each of the approaches. The purpose of this section is to show that multi-state modelling can produce equivalent results to Markov decision-analytic modelling, if the same assumptions are used for each approach. Much of this section is based on an appendix in one of my accompanying publications to this chapter (Williams et al., 2017b).

In order to facilitate the direct comparison, the assumptions used for the transition probabilities in the manufacturer's Markov decision-analytic modelling were converted to transition hazards for use with the multi-state modelling. For comparison purposes transition probabilities, rather than the transition hazards on which multi-state modelling is based, were also used. However, the Markov decision-analytic modelling did not assign probabilities to the progression-free → progression transition/progression state directly. Instead the probability for the progression state was derived from the probabilities for progression → death, progression-free → death and the probability of staying in the progression-free state. The transition hazards and probabilities for use with the multi-state modelling were obtained in a similar manner.

Figure 7.43 shows the predictions of staying in the progression-free state for each of the treatment arms.

It can be seen in Figure 7.43 that the predictions from each approach were very similar for both treatments.

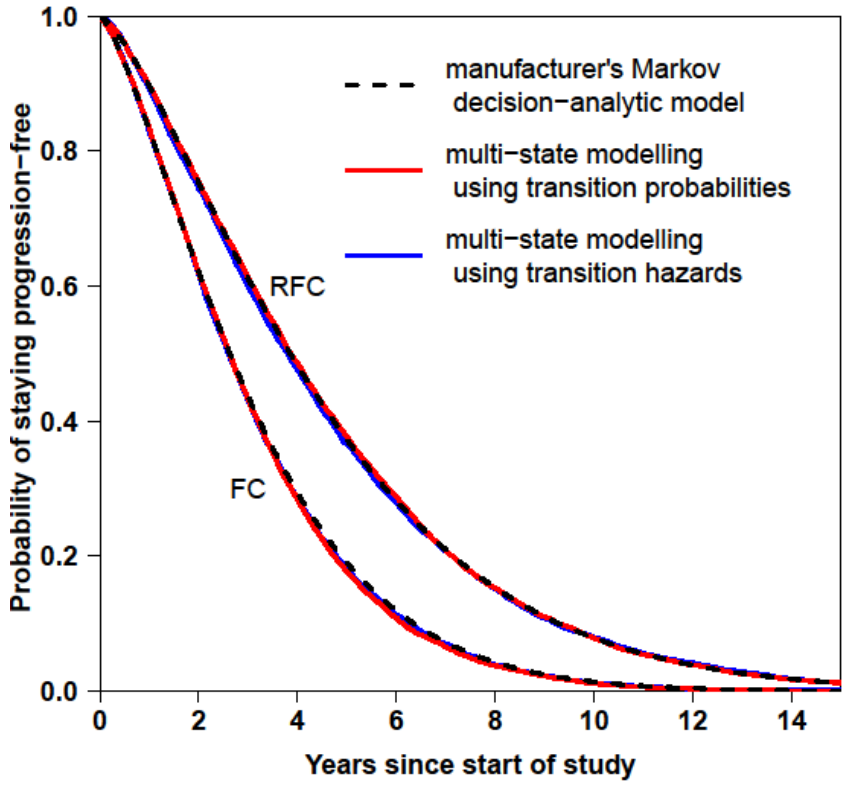


Figure 7.43 Probability of staying in the progression-free state: equivalent assumptions

Figure 7.44 (a) and (b) show the predictions of being in the progression state for RFC and FC respectively.

It can be seen in Figure 7.44 (a) and (b) that the predictions from the manufacturer's Markov decision-analytic modelling and the multi-state modelling using transition probabilities were similar. They had a higher peak than with the multi-state modelling using transition hazards.

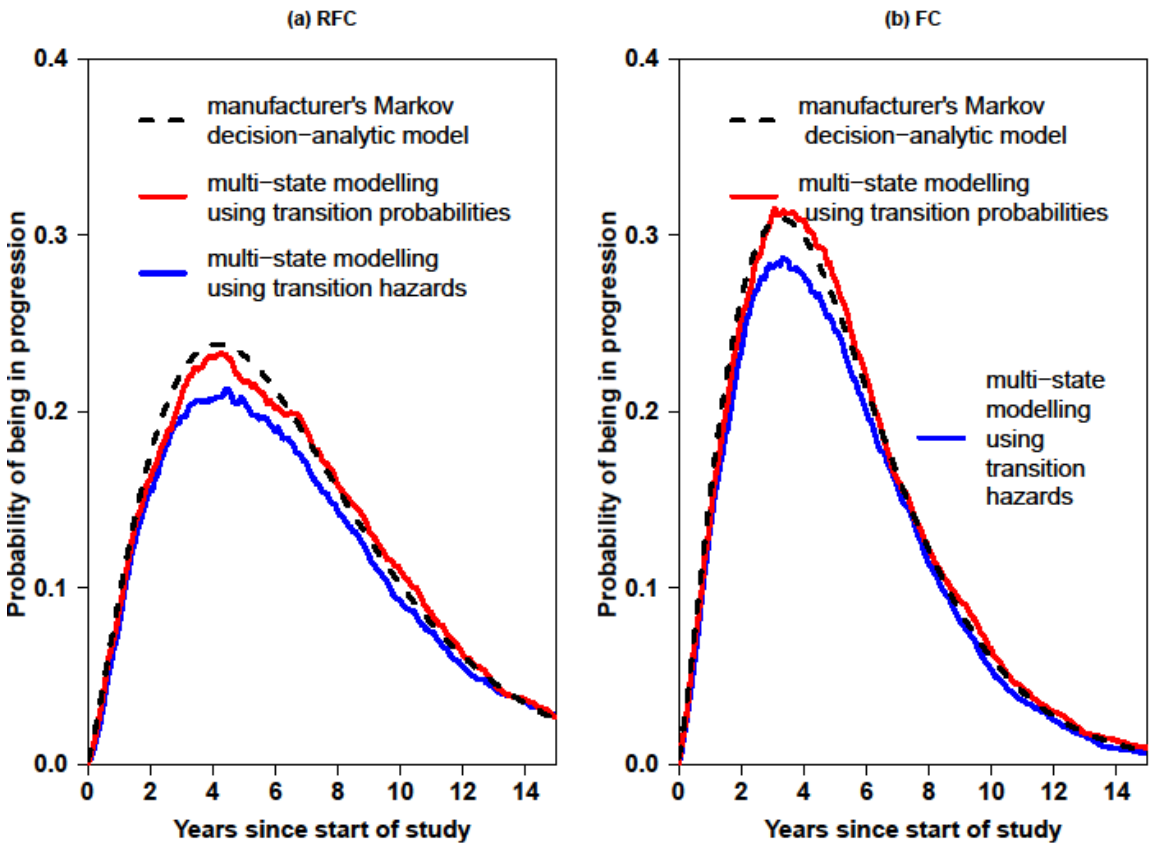


Figure 7.44 Probability of being in the progression state: equivalent assumptions (a) RFC (b) FC

In Figure 7.45 the predictions of being in the death state can be seen for RFC and FC.

Figure 7.45 shows that the predictions from the manufacturer's Markov decision-analytic modelling and multi-state modelling using transition probabilities were reasonably comparable, with slightly higher predictions with the multi-state modelling using transition hazards.

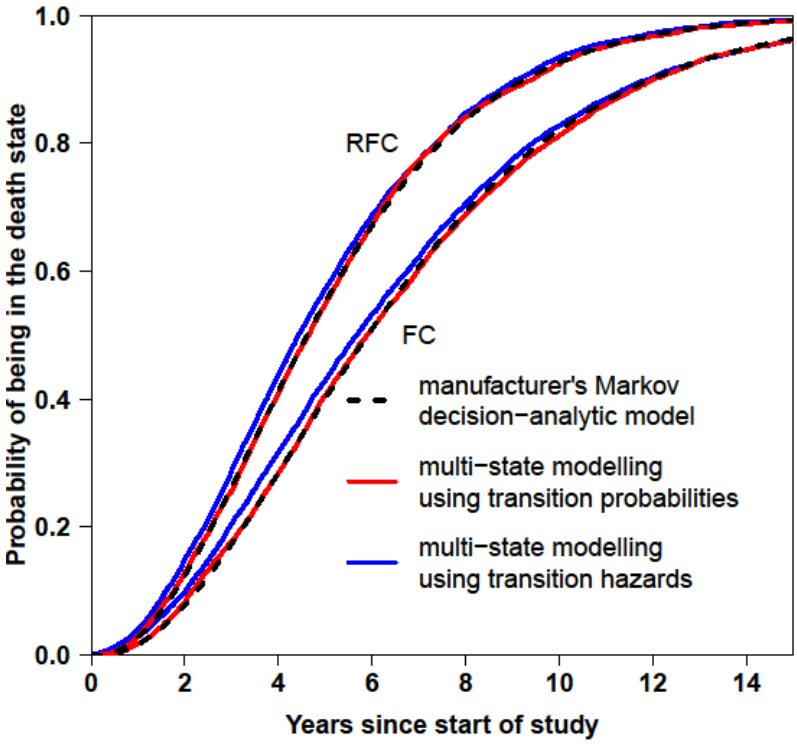


Figure 7.45 Probability of being in the death state: equivalent assumptions

Table 7.27 shows the incremental mean Life Years in each of the relevant health states using each of the approaches. Two methods of calculating the mean Life Years are shown. Firstly, the trapezoidal rule is used to calculate the area under the curve, the approach used throughout in this chapter for the multi-state modelling. Secondly, the probabilities at each time point were summed together, the approach used by the manufacturer in their Markov decision – analytic model. The actual results presented by the manufacturer are shown in bold.

It can be seen in Table 7.27 that the results for the incremental mean Life Years in Progression were not overly sensitive to the method used to calculate the means for any of the approaches. Furthermore, it can be seen that the results for the individual treatments from the multi-state modelling with transition probabilities were closer to that for the Markov decision-analytic model than the corresponding results using transition hazards. However, in terms of incremental

	mean Life Years calculated using: area under the curve with trapezoidal rule			sum of transition probabilities over time		
	RFC	FC	Incremental	RFC	FC	Incremental
Mean Life Years Progression-free						
multi-state modelling using transition hazards	4.07	2.89	1.18	4.11	2.93	1.18
multi-state modelling using transition probabilities	4.11	2.87	1.24	4.15	2.92	1.24
Markov decision modelling without a half-cycle correction	4.10	2.92	1.18	4.15	2.96	1.18
Markov decision modelling with a half-cycle correction	4.07	2.88	1.18	**4.11**	**2.93**	**1.18**
Mean Life Years in Progression						
multi-state modelling using transition hazards	1.45	1.58	-0.13	1.45	1.58	-0.13
multi-state modelling using transition probabilities	1.58	1.73	-0.15	1.59	1.73	-0.15
Markov decision modelling without a half-cycle correction	1.61	1.73	-0.11	1.62	1.73	-0.11
Markov decision modelling with a half-cycle correction	1.62	1.73	-0.11	**1.62**	**1.73**	**-0.11**

Table 7.27 Incremental mean Life Years in each health state: equivalent assumptions

results, the transition hazards approach produced means that were nearer the Markov decision-analytic modelling results. This was because, whilst there was more discrepancy between the individual treatment means, the differences were comparable for each treatment resulting in a similar incremental effect.

Table 7.27 also shows that the results for the incremental mean Life Years Progression-free were sensitive to the method used to calculate the means. Regardless of the method used, the multi-state modelling approach using

transition probabilities produced incremental results which were higher than any of the (similar) corresponding results from the other approaches. Furthermore, it can be seen that the transition hazards approach was most comparable to the actual Markov decision-analytic modelling when the summing of probabilities method was used. When the trapezoidal rule was used, the results using the multi-state modelling with transition hazards and the Markov decision-analytic modelling were similar. However it was the Markov decision-analytic modelling without the half-cycle correction that most represented the actual Markov decision-analytic modelling when the trapezoidal rule was used. This suggests that using the trapezoidal rule for calculating the means (based on probabilities without a half-cycle correction) may be equivalent to the summing of probabilities for which there was a half-cycle correction.

7.9 Summary of the results and Discussion

This chapter demonstrated and compared three different approaches to estimating benefit (survival) for use in economic evaluations – Markov decision-analytic modelling, partitioned survival and multi-state modelling. An existing Markov decision-analytic model created by a manufacturer in a submission to NICE for a technology appraisal acted as motivation for the comparison. The model, and the trial data that was the main source underlying it, acted as a case study with a health economics context for this thesis.

The manufacturer's Markov decision-analytic model found the treatment to be cost-effective with a Cost per QALY gained of £13,189. The mean Life Years (QALYs) gained in the model was 1.07 (0.88) for the treatment RFC compared to FC. However, the Evidence Review Group (ERG) who reviewed the submission thought that an overall survival benefit of such a magnitude was inappropriate (PenTAG, 2009). This was mainly because the manufacturer assumed no treatment effect while in progression, which meant most of the benefit was found whilst progression-free. The ERG initiated two sensitivity analyses that completely removed the overall survival benefit. These resulted in Costs per QALY gained of £30,336 and £30,304, just above a willingness-to-pay threshold of £30,000 per QALY gained. After further investigation as part of the research in

this thesis, these analyses were equivalent to hazard ratios of death after progression of 4.451 and 1.758. However when a Cox regression was fitted to the observed data the hazard ratio was 1.364, and therefore a discrepancy was apparent. This motivated the exploration of alternative approaches to analysis – namely partitioned survival and multi-state modelling.

Unlike the Markov decision-analytic modelling and multi-state modelling that modelled transitions between all relevant health states, the partitioned survival approach was based on calculating area under the curves of survival outcomes in the trial. As such, progression → death was not modelled directly. It was instead derived from the area under the overall survival and progression-free survival curves, the outcomes in the trial. Parametric regressions were fitted to these curves to allow extrapolation. A Weibull regression fitted to progression-free survival observed over the trial period was extrapolated to 15 years. However this approach was not adequate for overall survival as there was no parametric fit that led to survival of zero upon extrapolation to 15 years. Therefore the predictions for overall survival were based on different fits during the observed and extrapolation periods. The partitioned survival approach resulted in a Cost per QALY gained of £16,308, and therefore deemed the treatment of interest cost-effective.

With the multi-state modelling approach, the Markov property was formally tested by using a state arrival-extended model. This involved including a covariate for the time in the previous state when modelling progression → death. Evidence of a violation of the Markov property was found and this led to modelling using a semi-Markov approach. It resulted in a Cost per QALY gained of £30,702. This exceeded the willingness-to-pay threshold of £30,000 per QALY gained and therefore the treatment was found not to be cost-effective.

Therefore, a discrepancy in results was evident between the three approaches. An analysis was carried out using identical assumptions with the Markov decision-analytic modelling and the multi-state modelling. This produced very similar results demonstrating that multi-state modelling can be used as an alternative to conventional Markov decision-analytic modelling. Furthermore, an accompanying

publication to this chapter provided a tutorial to encourage others to adopt the approach, with elegant implementation using the statistical software R rather than having to rely on spreadsheets (Williams et al., 2017a). The analysis with the identical assumptions confirmed that the disparity seen in the main analysis was due to the different assumptions used with the different approaches. Furthermore, the main analysis, where each approach used different assumptions, demonstrated that the approaches compared well over the observed period of the trial, with most of the discrepancy apparent in the extrapolated period. Consequently, the comparison highlighted that it is imperative to rigorously check any assumptions used to ensure they are realistic. It is these in particular that can influence clinical and cost-effectiveness results, much more than the approach used for the modelling itself.

Liaising with clinicians and/or using external data sources such as registry data are recommended ways of gathering evidence to help ensure assumptions are realistic. In the comparison illustrated in this chapter, the main discrepancy in mean benefits gained overall stemmed from the differences between the approaches in mean Life Years gained whilst in progression. It could be seen in Figure 7.33 that this was primarily due to differences in the magnitude, and timing, of the peak for the probability of being in the progression state. Therefore, it would be worthwhile for any modellers faced with a similar situation to gather information to reduce uncertainty surrounding this peak and the predictions over time generally.

With the Markov decision-analytic modelling, the probability for the progression-free → death transition was based on the maximum of the observed death rate and an age-related background mortality rate in the general population. However, the progression-free → death transition had the competing risk of the progression-free → progression transition. Therefore, the observed death rate used to inform the transition may have been inflated if the competing risk was not taken into account, analogous to naïve Kaplan-Meier estimates. Hence, it would also be worthwhile checking that any inflation of the observed death rate while progression-free did not bias the predicted probability of

progression-free → progression and the related prediction over time of being in the progression state.

The extrapolation for progression-free → death appeared to be unrealistically flat as mortality would be expected to increase as patients aged. While it was not the case in this illustration, it may be worth investigating whether the extrapolations are sensitive to the order in which transitions are considered when choosing distributions. In this research, the plateau over the extrapolation period was most likely due to the relatively few pre-progression deaths, with the vast majority of patients having a progression and then having a post-progression death sometime before the end of the time horizon.

In addition, it would be beneficial to verify that any assumption used for a treatment effect post-progression is clinically plausible. It would also be advantageous, for each transition, to use appropriate available information to help inform assumptions with regards to continuation of treatment effects in the extrapolation period. Finally, it is worth taking a common sense approach, in conjunction with clinical opinion and external data such as registry data, to decide on a time horizon and extrapolation to that point that is realistic in a clinical sense.

The visual assessment of fits from each of the approaches suggested the partitioned survival approach provided the most reasonable fit. At first, this might appear a surprising result as one might expect an approach that models each of the transitions between health states of interest to provide the most sensible predictions. However in this example, partitioned survival was the only approach in which the regressions were fitted directly to the observed (Kaplan-Meier) curves for the initial and absorbing states. In addition, the predictions for the intermediate state – the only other one that required prediction – were derived from the predictions for these two other (directly fitted) states. Therefore, partitioned survival would provide the best fit to the observed data. Furthermore, it was the only approach that had observed data (Kaplan-Meier curves) to check the fit to every transition/state.

It may be that a state-transition approach, such as Markov decision-analytic modelling or multi-state modelling, is only advantageous in more complicated models where more states/transitions are of interest. However, more complex models have the drawback of being more difficult to check against observed data. This is because more intermediate states will mean more states with flow in and out of them. A model with several of such states will be more difficult to check against observed data because this will not just involve Kaplan-Meier curves, but will have to take account of the fact that patients can enter such states but also subsequently leave them. That said, prior knowledge of, or relevant external data to inform, the expected shape of such curves should help. It is also worth noting that in the illustration in this research, when the situation arose that a survival outcome over the observed period did not reach zero upon extrapolation, it was reasonably straightforward to use separate fits for the observed and extrapolated periods with the partitioned survival approach. It would be more complicated to use a similar technique with a state-transition approach. This is because some of the transitions would be interlinked and therefore it would not just be survival outcomes in isolation that would need to be considered.

An important consideration should be the number of covariates used in the modelling. In this illustration, treatment was the only covariate used in the modelling for each of the transitions. When data for other possible predictors and confounders are available from the trial/study, it would be worthwhile considering them in the modelling as they should result in more accurate predictions of the states/transitions of interest. However there can no longer be the same reliance on visual assessment of fits, by comparing observed data with model predictions, as they become impractical. That said, a workaround might be comparing predictions from parametric regressions with those from a Cox regression as a surrogate for the observed data.

Multi-state modelling has huge potential as an alternative elegant way of estimating transition probabilities in health economic modelling. The multi-state modelling approach uses the individual patient data directly to model the transitions and negates deciding on transition probabilities *a priori*. It uses the

exact times of transition and as such does not require modelling over (arbitrary) discrete cycles, nor does it require the use of tunnel states. Additionally, this particular multi-state modelling approach incorporates parametric distributions for hazards, which as well as allowing extrapolation of survival, can permit hazards that vary over time if required. Given the modelling at the individual patient level, multi-state modelling also provides an alternative to microsimulation (Davis et al., 2014).

Using a statistical package to build multi-state models is less time consuming than building a model using a series of spreadsheets, the approach frequently used in Markov decision-analytic modelling. The creation of a model in R, the calculation of the associated transition probabilities, the areas under the curves to obtain the mean life years/ QALYs and the discounting each only require one line of R code. The syntax-based approach of statistical software also means that it is easily reproducible. It is somewhat easier to understand and perform all the required modelling and other calculations when the syntax is all contained in one document rather than in a series of spreadsheets. The functions available in my published tutorial paper (Williams et al., 2017a) contain modifiable arguments that can be changed to suit the requirements of the user. This provides a valuable resource to health economic modellers as it overcomes the lack of available software that was a barrier to adopting the approach.

Chapter 8 Main insights, limitations and potential areas of future research

8.1 Introduction

The overall aim of this thesis was to explore the potential of competing risks analysis and multi-state modelling in an epidemiological and health economics context, in areas where they were not widely used. The thesis described the background to each of the methods. This included emphasising the importance of recognising a competing risks scenario when it exists, methodological issues that might need to be considered and the extra insight that can be gained from using each of the methods over standard survival analysis. The research was centred around two different case studies – one in epidemiology and one in health economics. The epidemiological one focused on stroke and the specific outcomes recurrence and death. The health economics case study was motivated by the economic model included in a technical appraisal submitted to NICE. This evaluated the cost-effectiveness of a chemotherapy regime for the first-line treatment of chronic lymphocytic leukaemia.

The background chapters also included reviews of the extent of use of competing risks (section 2.10.2) and multi-state modelling (section 4.7) in stroke epidemiology, thereby achieving the first of the stated objectives on page 3 of this thesis. These reviews found this was an area where the full potential of the methods was not demonstrated and motivated the empirical chapters in stroke epidemiology. This thesis appears to include the first study to demonstrate the greater understanding that can be achieved by examining both the cause-specific hazards and the subdistribution hazards/cumulative incidences of each of the two competing risks recurrence and death without recurrence in stroke patients. Furthermore, it appears to be the first research to showcase the potential of dynamic predictions in a multi-state model of stroke, recurrence and death.

Multi-state modelling was the focus for the health economics context. In particular, this thesis aimed to illustrate that multi-state modelling can provide

an alternative to common approaches in health economic modelling. The review of the use of multi-state modelling in health economics found that a barrier to implementation was a lack of readily available software. This motivated a tutorial paper I have subsequently published with accompanying `R` code and customisable functions to allow analysts to adopt the approach with their own data (Williams et al., 2017a).

Importantly throughout the chapters that described the methods, including the reviews of the literature into the extent of their use specific to the two case studies, misconceptions and misunderstandings related to some of the methodological issues were highlighted. The purpose of this was to alleviate the confusion that could arise from some of the conflicting messages in the literature.

Section 8.2 of this final chapter summarises the main insights from the empirical chapters of this thesis, demonstrating where appropriate some of the issues and considerations emphasised in the background chapters. Specifically, sub-section 8.2.1 focuses on how an aspect of the second objective for the stroke study was met, namely demonstrating the extra insight gained from competing risks analysis over standard survival analysis. Next, sub-section 8.2.2 provides an overview of how the last aspect of that objective, that of demonstrating the insight from multi-state modelling in a stroke epidemiology context, was achieved. Then, sub-section 8.2.3 outlines how meeting the two objectives in the health economics context provided extra insight. In section 8.3 limitations of the analysis in this thesis are discussed alongside solutions to overcoming them. Finally, the thesis concludes with section 8.4 that discusses future potential directions for the research that would help encourage adoption of the multi-state modelling framework beyond the illustrated case studies.

8.2 Main insights

8.2.1 Insights gained from the competing risks analysis over and above standard survival analysis

Chapter 3 met the objective of demonstrating the extra insight gained from competing risks analysis of the stroke case study, with particular emphasis on

three different comparisons involving its use. These were:

(i) composite outcome versus decomposition of the outcome into the competing risks

(ii) cause-specific hazard and Fine and Gray subdistribution hazard modelling approaches

(iii) naïve Kaplan-Meier/Cox estimates versus the competing risks cumulative incidence.

The insights found from each of these will be summarised in turn.

8.2.1.1 Composite outcome versus decomposition of the outcome into the competing risks

Decomposing a composite outcome into its component competing risks can provide useful insight. This is especially the case when a covariate affects two competing risks but in opposite directions. The decomposition of a non-fatal outcome from death can be particularly useful because it could help distinguish those patients that are likely to die from those that may still benefit from a targeted intervention. In the analysis presented in section 3.2 of this thesis, extra insights were gained from the decomposition for the effects of the ability to lift both arms off the bed, haemorrhage on scan and gender. It was found that being able to lift both arms off the bed was associated with a higher hazard of recurrence and a reduced hazard of death without recurrence. However, this association with recurrence was undetectable from the analysis of the composite outcome. This was due to it being in a different direction from that for death without recurrence, together with the composite outcome being dominated with death without recurrence. Therefore the competing risks analysis was crucial in identifying the association between the ability to lift both arms and recurrence.

Another covariate that had opposing effects on each of the two competing risks was haemorrhage on scan. Similar to the ability to lift both arms, having a haemorrhage was associated with an increased hazard of recurrence and a marginally reduced hazard of death without recurrence. However, no evidence was found of an association with haemorrhage and the composite outcome, and

therefore the composite analysis masked both of the effects found for the competing risks. The opposing effects on the competing risks were effectively cancelling each other out in the analysis of the composite outcome. Therefore, this again demonstrated the extra insight gained from the competing risks analysis over that of the standard composite outcome survival analysis. In particular, it revealed that haemorrhage was in fact a risk factor for recurrence and therefore identified a subset of patients likely to benefit from strategies to prevent recurrence.

The analysis of the composite outcome also did not find any evidence of an effect of sex. However the competing risks analysis found that being female was associated with a higher hazard of recurrence. Therefore this illustrated another effect that was masked from the analysis of the composite outcome, again highlighting the value of the competing risks analyses.

8.2.1.2 Cause-specific hazard and Fine and Gray subdistribution hazard modelling

In section 2.5 of this thesis, it was emphasised that a greater understanding of a competing risks scenario can be gained when both the hazards and cumulative incidence functions for each of the competing risks are considered. In particular, Latouche et al. (2013) recommended displaying both cumulative hazard plots and cumulative incidence plots to assess the effects of covariates. With this in mind, section 3.3 of this thesis showed such plots for each covariate. This appeared to be the first research to consider both cumulative hazards and cumulative incidences for the outcomes recurrence and death without recurrence in stroke patients. These helped to demonstrate the methodological issue emphasised in section 2.5 that the effects of a covariate on the cause-specific hazard and cumulative incidence of an outcome may differ. For instance, the cumulative hazard plot for orientated speech in section 3.3 showed that the hazard of recurrence was lower for those with orientated speech than those without it. However the cumulative incidence plot demonstrated that orientated speech had no effect on the cumulative incidence of recurrence. While this might seem surprising at first, this could be explained by knowing that those with orientated speech also had a lower hazard of death without recurrence, but that importantly this reduced hazard was more pronounced than

that for recurrence. This meant that there was more of those with orientated speech having a recurrence than was perhaps expected, due to there being more of them left to be at risk of recurrence owing to the decreased hazard of death without recurrence. Consequently, the cumulative incidence of recurrence for this group was in line with that for those without orientated speech, instead of being lower than it. Many similar examples were also illustrated in section 3.5 of this thesis. They demonstrated that exploring both the hazards and cumulative incidences of each of the two competing risks can help in the interpretation of such outcomes, which would not be possible if considering only the hazard or cumulative incidence in isolation.

8.2.1.3 naïve Kaplan-Meier/Cox estimates versus the competing risks cumulative incidences

Section 2.3 of this thesis emphasised that the naïve Kaplan/Cox approach is inappropriate in the presence of competing risks. Estimates using this approach introduce bias in the form of inflated cumulative incidences, with the extent of the bias relating to the strength of the effect of the competing outcome. Section 3.4 of this thesis illustrated the bias that can be introduced by not appropriately taking into account the competing risks in the stroke epidemiology case study. For example, it was found that inflation of 18% resulted in the estimate of the 4-year cumulative incidence of recurrence when using the biased Kaplan-Meier method instead of the more appropriate competing risks cumulative incidence approach. Furthermore, when the effects of covariates were taken into account, the naïve Cox prediction introduced 11% inflation in the 3.5-year cumulative incidence of death without recurrence for the reference patient. Therefore, the illustration in section 3.4 demonstrated it is imperative to use an approach that appropriately takes into account competing risks to avoid introducing such inflation bias. Over-estimating the incidence of an event could lead to misguided clinical decisions.

This piece of empirical analysis has reiterated the bias that can be introduced by using a naïve Kaplan-Meier approach when competing risks are present. Before conducting any of the analysis in this thesis, a literature search was conducted to find reviews that raised awareness of the need to recognise competing risks and analyse them appropriately (section 2.10.1). Two such reviews were found.

Mell et al. (2012) found that only 40% of the cancer studies they reviewed analysed cancer events and non-cancer mortality separately. Koller et al. (2012) found that 70% of studies they reviewed in high impact clinical journals either used a naïve Kaplan-Meier approach, neglected to report competing risks or failed to analyse them.

More recent contributions have also found that the naïve Kaplan-Meier approach has continued to be used [van Walraven and McAlister (2016) and Schumacher et al. (2016)]. In the studies reviewed in each contribution, 46% (van Walraven and McAlister, 2016) and 49% (Schumacher et al., 2016) were found to be susceptible to such competing risks bias. It is apparent that there is some delay in recognising the message regarding the need to analyse competing risks appropriately. However, the suggestion by van Walraven and McAlister (2016) to explicitly include the number of competing outcomes in the CONSORT/STROBE guidelines is a step in the right direction. In addition, as van Walraven and McAlister (2016) point out, there has quite recently been a move towards the base packages of popular software packages including the facility to create cumulative incidence curves for competing risks, rather than just Kaplan-Meier analyses. SAS, Stata and R all now have readily available procedures to create these and therefore there is some cause for optimism that more appropriate analysis will be more widely adopted.

8.2.2 Insights gained from the multi-state modelling of stroke, recurrence and death

Section 4.5 of this thesis highlighted that one of the main advantages of multi-state modelling is the flexibility of predictions it can offer. The usual predictions at time zero of being in a health state by some future time point can be accommodated. However, dynamic predictions are also possible with multi-state modelling. These dynamic predictions update a patient's prognosis taking into account the time elapsed and any subsequent events experienced.

Chapter 5 illustrated both types of prediction from a model of stroke, recurrence and death. In particular, it demonstrated dynamic predictions that could be used by clinicians as a communication tool to update their patients on their risk of relevant outcomes given their current stage in their illness.

The illustration of the flexibility of predictions offered by multi-state modelling in section 5.2 met the second objective in a stroke epidemiology context on page 3 of this thesis, that is to say to demonstrate the added insight of multi-state modelling over standard survival analysis. In addition, section 5.3 demonstrated that predictions from multi-state modelling reflect that intermediate states have flow in and out of them, in contrast to the competing risks cumulative incidence predictions that are monotonic. In doing so, it highlighted the extra insight gained from multi-state modelling compared to competing risks analysis.

8.2.3 Insights gained from exploring multi-state modelling in a health economics context

8.2.3.1 Publishing a tutorial paper to help encourage adoption of multi-state modelling for cost-effectiveness analysis in R

Section 6.3 of this thesis met the objective of exploring the extent of use of multi-state modelling in the health economics literature. It identified that a barrier to its use was the lack of readily available software. Motivated by this, a large component of this PhD research involved creating a series of customisable `R` functions for use by others to perform multi-state modelling for the purposes of cost-effectiveness analysis. All the coding now accompanies a tutorial paper I published providing a step-by-step guide to using the approach (Williams et al., 2017a). The paper demonstrates how to use multi-state modelling to calculate mean (quality-adjusted) life years gained and incremental cost-effectiveness ratios. It also illustrates how to carry out deterministic and probabilistic sensitivity analyses with the approach. It is based on adaptions to the existing `R` package `mstate` to accommodate parametric multi-state modelling, thereby facilitating extrapolation of survival curves. All functions have arguments that can be changed to suit the data and context of the user e.g. the number of transitions, covariates included, the discount rate and the time horizon. Furthermore, in using a syntax-based approach, it benefits from the transparency of the code used and analysis that is all contained in one file.

8.2.3.2 Illustration of multi-state modelling as an alternative to partitioned survival and Markov decision-analytic modelling

Chapter 7 achieved the objective of demonstrating, with the help of the aforementioned tutorial paper, a comparison of multi-state modelling with two common approaches in health economic modelling. An existing Markov decision-analytic model developed by a manufacturer in a submission to NICE for a technology appraisal was used as a basis for the comparison. That economic model sought to evaluate the clinical and cost-effectiveness of rituximab in combination with fludarabine and cyclophosphamide compared to fludarabine and cyclophosphamide alone for the first-line treatment of chronic lymphocytic leukaemia. The model had the three health states progression-free, progression and death & the transitions progression-free → progression, progression-free → death and progression → death. The manufacturers took the usual approach in state-transition decision-analytic modelling of making *a prior* assumptions about the transition probabilities before the modelling commenced.

By way of comparison, Chapter 7 used the trial data on which that economic model was based to carry out analyses using the partitioned survival and multi-state modelling approaches. The partitioned survival approach involved fitting parametric regression to the overall survival and progression-free survival curves to allow extrapolation to the desired time horizon. The mean time spent in progression was then derived from the area between the two survival curves. The multi-state modelling used the individual patient level data directly to build parametric regression models for the hazard of each of the transitions in the model. These hazards for each transition were then appropriately combined to calculate the state occupancy probabilities.

For each of the three approaches, the (incremental) mean (quality-adjusted) time in each state was presented along with the incremental cost-effectiveness ratio. Furthermore, for the multi-state modelling that was the main focus of this thesis, sensitivity analyses were presented. A tornado diagram was shown to summarise the results of one-way sensitivity analyses and a cost-effectiveness plane and cost-effectiveness acceptability curve were presented to summarise the extent of the uncertainty from the probabilistic sensitivity analyses. The chapter helped to demonstrate that any output required from a conventional

decision-analytic model can just as easily be produced using the multi-state modelling framework.

Chapter 7 showed that the conventional state-transition decision-analytic modelling approach and multi-state modelling produced equivalent results when they made the same assumptions with regards to the transition probabilities/hazards. Furthermore, when the comparison of the approaches each using different assumptions were split into the results in the observed and extrapolated period, it was found that the approaches mainly differed over the period of extrapolation. Therefore, the research demonstrated that it is imperative to check whether the assumptions used are realistic. It is the choice of assumptions that can influence the clinical and cost-effectiveness results, much more than the approach to the modelling itself.

8.3 Limitations

The purpose of this thesis was to showcase multi-state modelling in an epidemiological and health economics context, in areas where their potential had not been fully realised. However the research did have some limitations. This section highlights some of the limitations and outlines, where possible, some solutions that could overcome them in future.

8.3.1 Limited follow-up with covariates versus extended follow-up with limited covariates

As is often the case with studies/trials of health outcomes, the case studies used as examples in this thesis had relatively short follow-up. However the studies did have a wealth of covariate information to help explain which factors were associated with the outcomes of interest. This was particularly the case for the stroke epidemiology study. This would have also been the case in the trial data used for the health economics example. However this was less evident because only the covariate treatment was used in the analysis, for comparison purposes with the existing economic model. On the other hand, external data that could help inform extrapolation of the outcomes to the time horizon, e.g. life tables or hospital episode data, tend to be limited in terms of covariate information. Therefore, some trade-off is necessary between study data with limited follow-

up but covariates, and routine administrative data with longer follow-up but a lack of covariates. However, the lack of covariates in the longer term data may be mitigated if the datasets are linked at individual patient level, and longer follow-up is only required to identify occurrence of outcomes. Furthermore, the use of longitudinal cohort studies may alleviate some of the limitations. In addition, such studies offer the opportunity to benefit from the emergence of joint modelling of longitudinal and competing risks/multi-state modelling survival data [e.g. Williamson et al. (2008) and Ferrer et al. (2016)].

8.3.2 Variable selection with competing risks and multi-state modelling

Prognostic models are used to predict outcomes for individual patients with a given set of risk factors. When developing such models, careful consideration should be given to the predictor variables (i.e. risk factors) selected for inclusion in the model. Parsimonious models are deemed to be the most practical because they are easier to use for predicting outcomes for new patients and more likely to generalise to different populations. Recognised strategies for curtailing the number of variables include choosing only those that are clinically relevant, removing obvious collinearity and ignoring those with data quality issues. Examples of the latter include measurement error or a high level of missing data. However, even after following such strategies to disregard variables of obvious limited value, the number of variables can still exceed Peduzzi et al. (1995)'s standard rule of thumb of 10 events per variable. Therefore, modelling-based variable selection procedures are often required. This section outlines some of the challenges and limitations of variable selection in the context of competing risks and multi-state modelling and describes very recent advances that could help overcome them.

8.3.2.1 Variable selection with Fine and Gray's proportional subdistribution hazard modelling

It was previously highlighted in section 2.7.1 of this thesis that Kuk and Varadhan (2013) have developed an information criterion-based test to help select variables for Fine and Gray's proportional subdistribution hazards modelling. The approach incorporated stepwise regression techniques. Such techniques are popular due to being easy to use. However, they have also been

heavily criticised for stability problems and their automation process, the latter meaning that analysts unaccustomed to prognostic modelling may not make appropriate considerations such as ensuring all variables are clinically relevant. Fu et al. (2017) have proposed an alternative approach using penalised regression with Fine and Gray's model. As well as providing an alternative to stepwise techniques, penalised regression is a shrinkage method and as such can overcome difficulties due to overfitting. In particular, many penalised regression techniques can be used for variable selection as they facilitate removal of variables, with lasso-based regression being perhaps the most well-known example of this.

Kuk and Varadhan (2013)'s aforementioned stepwise information-criteria based approach appears to be the first article to address variable selection using Fine and Gray's model. Consequently, it was a useful contribution in that it highlighted the issue and provided a starting point to build on. Furthermore, Fu et al. (2017)'s development to a penalised regression approach is a valuable addition to the literature on variable selection for competing risks model for the reasons described above.

8.3.2.2 Variable selection with cause-specific hazard modelling

It was previously mentioned in section 2.6 of this thesis that prediction is possible with the cause-specific hazard approach to competing risks modelling. However, variable selection is much more challenging with this approach. This is because of the phenomenon that the effect of a covariate on the cause-specific hazard will not necessarily translate into the same effect on the corresponding cumulative incidence. The cumulative incidence may also be affected by the cause-specific hazards of any competing events (section 2.5). Consequently, when selecting covariates to include in modelling, it is not recommended to focus on the cause-specific hazard for one event in isolation. While the two aforementioned methods with Fine and Gray's model can result in models for each event that differ in the covariates they contain, a different approach is required for cause-specific hazard modelling.

Instead, each of the events could be considered simultaneously and the effects of a given covariate on each event taken together as a group when deciding on

selection of that covariate. As Most et al. (2016) point out, in multiple response models such as competing risks, a distinction can be made between variable selection and parameter selection. This is because in multiple response models the effect of one covariate can be represented by several parameters e.g. a parameter for the effect on each event.

When aiming for parsimonious modelling using the cause-specific hazard approach with a view to using it for prediction, Most et al. (2016) emphasis on selecting variables instead of selecting individual parameters is a very useful recommendation. It means that all coefficients that correspond to the same covariate can enter or leave the model jointly (Tutz et al., 2015) i.e. even if a covariate only has an association with one event, that covariate will be kept in the models for the competing events. However, the work by Most et al. (2016) is from the perspective of the discrete-time, rather than the continuous-time, competing risks framework. There does not appear to be any contributions in the literature detailing an equivalent for the latter as yet, and unfortunately Most et al. (2016)'s piece does not include any code that could be adapted for continuous-time.

Some may question why prediction using cause-specific hazard modelling is considered, when variable selection is a more challenging process than it is with Fine and Gray's subdistribution hazard modelling. While there is the option of Fine and Gray's approach for modelling competing risks, this is not the case in multi-state modelling. In addition, Fine and Gray's model and cause-specific hazard modelling may not both simultaneously meet the proportional hazards (PH) assumption for the covariates. Therefore, in instances when the PH assumption is met with the cause-specific hazard modelling approach but not with Fine and Gray's model, analysts may favour the former for calculating predictions.

Therefore, in the (continuous-time) cause-specific hazard modelling approach to competing risks and multi-state modelling, a notable limitation is the lack of a recognised appropriate variable selection method in preparation for calculating predictions. This is also true in general for any state transition model, including the commonly used Markov decision-analytic model in health economics, because the presence of competing risks means the transition-specific covariate

effects may not necessarily translate into the same effect on the state occupancy probability that is ultimately of interest.

However, it may be possible to overcome the lack of a suitable variable selection method. Adapting the aforementioned work by Most et al. (2016) from a discrete-time to a continuous-time framework may solve the problem for competing risks. With regards to multi-state modelling, Reulen and Kneib (2016) proposes a boosting method using penalisation of the (partial) likelihood as a way of selecting a parsimonious model. However, this is a variable selection approach tailored to modelling for etiological reasons, rather than for subsequent prognostic modelling, because it results in models for transition-specific hazards that do not all contain the same variables. On the other hand, Turlach et al. (2005), Simon et al. (2013) and Tutz et al. (2015) have developed variable selection methods using lasso-based penalisation for multinomial logit models. While not formulated for a survival analysis framework, their contributions may provide some insight into a solution for competing risks/multi-state modelling because multinomial logit models share with them the feature of being multi-response. In particular, the approaches use grouped penalisation with groups that contain all the coefficients belonging to the same variable in the context previously mentioned in the piece by Most et al. (2016). Of particular note, the original lasso (Tibshirani, 1996) is adapted to accommodate grouping so that all coefficients that correspond to the same covariate enter or level the model jointly.

It is encouraging that variable selection methods have already been developed for multivariate modelling, and therefore it is hoped that an adaption to cause-specific hazard modelling will appear in the literature in the near feature. The resultant parsimonious modelling for each cause-specific hazard could then be used to calculate predictions as detailed previously in section 2.6.

8.3.2.3 Limitation for analysis in the wider health economics context

Some health economic modellers may be reluctant to adopt a multi-state modelling approach using syntax with statistical software. They may see `R`, and other software packages that allow users to write their own code, as less user-friendly than spreadsheet packages such as Excel. However this might be due to

fear of the unknown rather than a barrier that cannot be overcome. The next section of this thesis includes details of anticipated developments of the R code, including incorporating a more user-friendly interface.

8.4 Areas of future research

This section gives an overview of some potential directions this research could take in future.

8.4.1 Enhancing health economic multi-state modelling by combining trial data with longer-term routine administrative data

8.4.1.1 Reform of the Cancer Drugs Fund

The Cancer Drugs Fund was introduced in 2010/2011 to fund cancer treatments not recommended by NICE for routine use in the NHS. In a news piece in the British Medical Journal, Mayor (2016) explains that reasons for such non-approval include that the drug was new and not yet assessed, or that it was for a rare cancer not being appraised by NICE. Mayor (2016) also emphasises that another reason why drugs have not been recommended for routine use is the uncertainty regarding their clinical-effectiveness and cost-effectiveness. It was envisaged that the Cancer Drugs Fund would address this. However it has been acknowledged that the fund needs reform, because as the Mayor (2016) piece highlights, it did not collect data on outcomes of use of the drugs it funded even though it had spent £1.27 billion.

A reform of the Cancer Drugs Fund was implemented in July 2016 that entails that drugs be funded for up to 2 years in order for further evidence to be collected. After this a short NICE appraisal will take place that considers this additional evidence. If approved, the drug will be licenced for routine use in the NHS. If it is rejected, it will only be considered for use on a case-by-case basis.

That a reform of the Cancer Drugs Fund was required, highlights that more attention needs to be given to evaluating the clinical-effectiveness and cost-effectiveness of such drugs and the uncertainty in the results. Indeed a recent article by Aggarwal A. (2017) found that only 38% (18/47) of drugs funded by the

Cancer Drugs Fund in 2015 were found to be effective. This need for more evidence means that there will be a greater demand for modelling in health economic evaluations.

However, because the reform does not stipulate exactly how the evidence should be collected, there is some debate over whether this reform will address the uncertainty in effectiveness. Grieve et al. (2016) have concerns that, while the reform encourages early access to drugs, it is at a high cost and still does not address uncertainty surrounding the clinical benefits. The article argues that the reform does not encourage manufacturers to conduct randomised clinical trials (RCTs), and does not advocate the sole use of observational data due to the biased estimates of effectiveness that can result. The authors point out that, once the drugs are widely used, if real world data is used alone to assess treatments randomisation will be impossible weakening the evidence base. A solution is proposed by Grieve et al. (2016) involving assessing treatment effectiveness using timely and pragmatic low-cost RCTs in conjunction with follow-up from routine collected data. Furthermore, if there is an ongoing trial it may be more efficient to wait and make decisions based on assessing the long-term outcomes from that trial, because a new trial may not provide sufficient evidence to justify the expense. The authors also argue that when RCTs are deemed unethical or impractical, then the non-randomised studies that are conducted should minimise confounding. These studies should collect longitudinal data on all relevant prognostic characteristics for those with and without the drug. Recent developments in strategies to minimise the selection bias from confounding include flexible regression incorporating machine learning for model selection Kreif et al. (2015), propensity score matching (PSM) with regression adjustment Kreif et al. (2013) and an extension of PSM called Genetic Matching Radice et al. (2012). Incorporating such methods into state-transition approaches, such as multi-state modelling, is an area of further research worth exploring.

In a rapid reply to Grieve et al. (2016)'s article, Hatswell (2016) stresses that the current Cancer Drugs Fund reform does not address the most pertinent issue in terms of uncertainty, that of effectiveness in the long-term. He agrees with Grieve et al. (2016) that collection of evidence on long-term outcomes would be more effective in studies that complement and extend the original clinical trials.

In a rapid reply from Longson et al. (2016) of NICE to Grieve et al. (2016)'s article, the authors refute the implication that the reform will rely heavily on observational data. The authors emphasise that the reform will include a data collection plan that will outline the key areas of uncertainty and how these will be addressed. On an optimistic note, they highlight that there is scope to use RCTs alongside observational data. Furthermore, they mention that ongoing RCTs can be used to help evaluate long-term effects and reduce uncertainty in extrapolated estimates. This renewed focus on addressing uncertainty, and not just relying on trials or observational data alone but instead combining them to obtain evidence long-term, should mean that the use of modelling that can facilitate this will increase. In particular, it could provide a window of opportunity for increasing adoption of multi-state modelling.

8.4.1.2 GetReal project on methods for real word evidence collection and synthesis

Combining trial data with the real world evidence from observational data is also being investigated by the GetReal consortium (www.imi-getreal.eu), with a specific aim of improving estimates of real word effectiveness. In particular, Workpackage 4 of the GetReal project is concerned with identifying and sharing best practice in synthesis and prediction modelling to estimate the real world effectiveness of treatments, beyond the efficacy estimated in the controlled environment of an RCT. An article by this group concerned with mathematical modelling to predict drug effectiveness from trials has already identified multi-state modelling as a method that has potential, but that has scope to be more widely-applied (Panayidou et al., 2016). However, the multi-state models found in the review described in that paper were under the discrete-time framework, i.e. they were typically what this thesis has termed conventional (Markov) decision-analytic models. The authors acknowledge that building such models can be cumbersome and time-consuming. However, the continuous-time multi-state modelling approach that has been the focus of this thesis overcomes such problems. In particular, continuous-time Markov multi-state models can be build in seconds (Williams et al., 2017a) and are therefore much less time-consuming than their discrete-time counterparts. This is mainly due to the modelling not requiring any form of simulation, such as cohort simulation. It is hoped that this

will help encourage the adoption of continuous-time multi-state modelling for cost-effectiveness analysis.

8.4.1.3 Extrapolation

Increasingly in the literature, more importance is being placed on assessing the suitability of extrapolations in terms of being clinically plausible, and in particular using external data to help inform the extrapolation. In the UK, NICE (2013) recommend that any extrapolation is assessed for "both clinical and biological plausibility of the inferred outcome as well as its coherence with external data sources".

Bayesian evidence synthesis that incorporates external information is demonstrated in the work by Demiris and Sharples (2006). In another article, Nelson et al. (2008) show how internal and external data could be used to inform the extrapolation in conjunction with time-metric and age-metric survival analysis. That is to say, in particular, the time to event was measured using time since initiation of treatment in the initial period when there was a high risk of death. Then, when the hazard of death could be assumed to be more stable, time was measured using age to facilitate extrapolation. Hwang and Wang (1999) proposed a method that utilises general population life table estimates to simulate a reference population to help inform the extrapolation. Specifically, their method involved a linear extrapolation of a logit-transformation of the survival ratio between the reference and specific population of interest, assuming a constant excess hazard. This approach has been applied by many authors including Chu et al. (2008a), Chu et al. (2008b), Lee et al. (2012), Yang et al. (2012), Chen et al. (2013) and Ho et al. (2006).

A very recent special issue of the journal Medical Decision Making specifically focused on methods for extrapolation in cost-effectiveness analyses was published in 2017. The contributions outlined how analysts could incorporate external data into their models [e.g. Jackson et al. (2017), Guyot et al. (2017), Negrín et al. (2017) and Meacock et al. (2017)].

The contribution by Jackson et al. (2017) is a valuable addition to the literature that starts by stating that the conventional approach to extrapolation is

parametric regression. However, the authors include the caveat that the observed follow-up should represent a large enough proportion of the whole lifetime when using this approach. In addition, they emphasise that this assumes that the observed hazards and trends will continue long term, but that this becomes more implausible as the proportion of the unobserved lifetime increases. They also stress that the extent of uncertainty in the extrapolation needs quantified, which can be difficult if observed data is immature. Jackson et al. (2017) reiterate the NICE recommendation that extrapolation be evaluated as to its clinical and biological plausibility and also coherence with external data. The article then proceeds to review methods of extrapolation involving external data, discuss the required assumptions and the circumstances under which method can be appropriate. In particular, the paper provides guidance on what to do when the disease and external (or treatment and control) populations have the same mortality all the time, the same mortality after some time and different mortalities in the short and long term. For the latter, the authors give an overview of how the external data could be adjusted to represent the population of interest. This included making assumptions about the proportionality of the all-cause or cause-specific hazards or about the additive excess hazard. The authors then discuss possible forms the survival function could take.

Guyot et al. (2017) detail how Bayesian multi-parameter evidence synthesis incorporating splines can be used to combine the observed trial data with external data sources to inform the extrapolation. In doing so, the authors demonstrate a useful approach encouraging analysts to explicitly use all available information outwith the trial to obtain extrapolations that are sensible and clinically plausible.

The work by Negrín et al. (2017) adds value in that it raises awareness of the need to consider the uncertainty in the estimates of extrapolated outcomes. The authors caution against the use of relying on “best-fit” models based on information criteria because they fail to address two areas of model uncertainty. Specifically, the two areas are uncertainty surrounding the true distribution and that related to the stability over time of the model parameters. Instead, their article demonstrates the use of Bayesian model averaging which can take into account both. The authors use two different approaches to this: one based on

selection using Bayesian information criterion (BIC), and an alternative incorporating optimistic and pessimistic scenarios. Negrín et al. (2017) include the alternative because they acknowledge that information criteria can only assess the goodness of fit of observed data, and cannot help decide how reasonable the extrapolation looks.

The premise of the paper by Meacock et al. (2017) is the authors' position that the usual binary outcome of 30-day mortality used to evaluate policy evaluations has room for improvement. The authors explain that this dichotomy assumes that those that die do so instantly, and therefore any time spent alive is ignored. In addition, those that do remain alive beyond 30 days are given a life expectancy based on that of the general population. However, those that the policy or program has an influence on are likely to be different from the general population. Furthermore, it is often possible to observe patients for more than 30 days but this longer term follow-up is not always utilised. The authors point out that trials often use all the follow-up data available and use parametric survival analysis to extrapolate the desired outcome over the lifetime. The aim of Meacock et al. (2017)'s article was to illustrate how this approach could be of use as a method of extrapolation in the evaluation of health policies. They did this by comparing the crude 30-day mortality outcome with parametric survival regression, both of which were coupled with general population life expectancy estimates to assess differences in remaining life years in the short term i.e. over 1 year. Furthermore, they also utilised the observed survival data over 1 year to carry out parametric survival regression, therefore basing the estimates on the population of interest rather than the general population. They performed extrapolation to a lifetime horizon and external validation using 3-year observed survival data they had available. Because of the initial very high risk of survival they fitted two separate models. The first was for short-term survival and was based on the observed 1-year data. The second was for long-term extrapolation, but excluded the high risk period, instead of being based on the data from 30 days onwards. After demonstrating the improved estimation with parametric survival regression, they then proceeded to use a difference-in-differences design to evaluate the effect on life expectancy of a policy.

Meacock et al. (2017) demonstrate the added value of using the observed data to estimate remaining life expectancy over the 30-day mortality estimates,

especially when with the latter the life expectancy estimates were from the general population rather than the population of interest. This was to be expected because in their example the life expectancy estimates for their population of interest differed quite considerably from that of the general population. However, the external validity of the extrapolation demonstrated in the paper was somewhat limited because it was only based on 3-years follow-up of mortality. Given their population had a mean age of 72 it would have been preferable if at least 10-years follow-up of mortality was obtained for a more comprehensive external validation to the lifetime horizon. Estimates of mortality at more points over the lifecourse would ideally have been available to check that the extrapolation was sensible and clinically plausible, especially as their population were aged from 18 years onwards.

In the analysis presented in Chapter 7 of this thesis the extrapolation and choice of time horizon did not benefit substantially from being informed by external data. This was primarily because the approaches were contrasted with an existing decision-analytic model already created by a manufacturer, and used the same time horizon as that model for comparison purposes. This constraint on the time horizon is likely only to be specific to this illustration due to it being based on a comparison with an existing model. It is not a limitation to the future use of multi-state modelling. Indeed, the `R` functions already created as part of this research (Williams et al., 2017a) facilitate choosing a different fit in the observed period from that in the period that requires extrapolation. Their scope could be easily widened to incorporate external data.

Another point worthy of consideration related to extrapolation in health economic modelling, but true of the whole time-frame of interest, is the impact of discounting on the predictions of being in relevant states. Given that after discounting the corresponding prediction curves will be lower and possibly somewhat different in shape, raises the issue as to whether it should be the fit to the curves after discounting that are evaluated instead.

While it is encouraging that guidance on extrapolation is increasing in the literature, its primary focus has been on the extrapolation of individual outcomes in isolation. Research into adaptions to existing methods of extrapolation for state-transition modelling would be a welcome addition. The

next section outlines how extrapolation can be more complex when multiple outcomes are considered simultaneously.

8.4.1.4 Implications for the choice of modelling approach

The previous sub-sections on the Cancer Drug Funds reform, the GetReal project and the use of external data for extrapolation have highlighted that there is huge scope to improve the evaluation of effectiveness and cost-effectiveness; specifically, by making use of the mechanisms that are available to combine study data with routine administrative data for longer-term follow-up. Within medical research there is an increasing ethos of data sharing. Indeed, public-funded research grant bodies are stipulating that data generated from research is shared with the wider research community, where possible. Furthermore, initiatives are emerging that allow trial data to be assessed by researchers e.g. https://clinicalstudydatarequest.com/. In addition, linkages with data from studies and routine administrative data are become more readily available, especially as the latter matures and new linkages become increasingly possible. Data such as hospital episode statistics and disease-specific (e.g. cancer) and death registries are among the sources of information that are likely to be useful.

With health economic modelling, interest is often on modelling several health states relevant to a (disease) process simultaneously. This allows for each of the different health states to be assigned a utility value representing quality of life, facilitating the calculation of cost-effectiveness measures such as cost per quality of life year gained. Multi-state modelling is a natural choice for this, especially when all the different data sources are linked at individual patient level. It also has the flexibility to incorporate aggregated data as sources of information to inform assumptions about transitions. However, analysts should consider carefully where to source the information required before embarking on any modelling to ensure that sensible and clinically relevant results emerge. This is particularly true of any extrapolation required as this is often the aspect with the greatest uncertainty.

While external data on death is likely to be available from death registries, information for other outcomes/health states in an intended model may be more

difficult to access. Furthermore, because with state-transition modelling the transitions can be interlinked, it can be problematic to achieve sensible results for each of the health states simultaneously. A less complex model may provide a solution when there is difficulty obtaining data to inform/check the plausibility of the results for the transitions and states. This could involve combining states so that a state-transition model consists of less transitions. In a competing risks scenario, it could involve combining the competing risks into composite outcomes and conducting standard survival analysis or taking the partitioned survival approach instead. The use of less complex models may also overcome problems that may arise when there is sparse information on covariates for an outcome. Combining events into composite health states may support allowing more covariates to be assessed.

8.4.2 Consideration of multiple comparators

In the health economics case study illustrated in this thesis, there was one comparator for the treatment and the two treatments were compared head-to-head in a single trial. However a full economic evaluation can involve multiple comparator treatments, the data for which can come from several different sources, particularly if multiple clinical trials and/or manufacturers are involved. The multi-state modelling approach can still be used in this context if an estimate of effect, either direct or indirect, between the different comparators can be found. Further research in this area would be worthwhile.

8.4.3 Simulation exercise to dissect the differences in the approaches compared

While multi-state modelling and decision-analytic modelling – the two state-transition modelling approaches – were compared when the assumptions used differed and when they were equivalent, this was not extended to the partitioned survival approach. A worthwhile area of future research would be a simulation exercise to fully understand the mechanisms that differ between each of the three approaches. This could focus on varying specific aspects to help dissect what was driving the discrepancies between the approaches e.g. hazard ratios used including assumed distributions, effects of treatment across transitions/events, level of censoring, extrapolation and time horizon.

8.4.4 Validation of models

The stroke epidemiology case study used in this thesis is part of a wider Stroke Complications and Outcomes Prediction Engine (SCOPE) study. SCOPE is a collaboration that brings together individual patient level data from various studies and trials for the purpose of developing and validating prognostic models. Therefore this is a valuable resource in which to internally and externally validate the competing risks and multi-state models for stroke developed in Chapter 3 and Chapter 5 of this thesis. In particular, both the calibration and discrimination of the models could be evaluated. However, an important consideration for external validation in particular, of competing risks analyses and multi-state modelling, is that the proportion of events due to a specific competing risk/health state is comparable to that in the data used to develop the model. It would also be worthwhile to validate the modelling in the health economics context.

8.4.5 Predictions and modifiable lifestyle factors

Chapter 5 illustrated survival predictions both from the start of the study and dynamically. These predictions could provide a useful communication tool for clinicians to discuss prognosis with their patients. However mortality may be a difficult subject for patients to engage with, especially if they feel they do not have control over their future. In the models demonstrated in this thesis, no modifiable lifestyle factors were included as covariates. However there are many areas where modifiable lifestyle factors are predictors, such as diabetes and heart disease. The flexibility of predictions that multi-state modelling facilitates has huge potential in chronic disease where changes in lifestyle could result in improvements in outcomes for patients. The predictions from multi-state modelling could aid clinicians when trying to encourage patients to take control of aspects of their lifestyle that could improve their health. For example, tobacco and alcohol consumption, body mass index, diet, exercise, cholesterol and blood pressure. Clinicians could discuss with patients the predictions of being in each of the relevant health states applicable to their disease if they continue with their current lifestyle. They then could try and motivate their patients to change by discussing how the predictions could improve with a change in lifestyle. The dynamic predictions in particular could

help the discussions in routine follow-up clinics because they would reflect the time that has elapsed since diagnosis and any health states experienced by the patient.

8.4.6 Widening the functionality of the coding for cost-effectiveness analysis in R using multi-state modelling

The R functions and accompanying code in the tutorial paper developed as part of this research has provided health economic modellers with the tools to start adopting multi-state modelling with their own data. It acts as a strong foundation to build on, with there being much scope to develop it further. A few examples of intended developments will now be outlined.

- It is anticipated that the existing facility within `mstate` to produce standard errors for predictions will be extended to the parametric approach. In particular, the current bootstrapping procedure in `mstate` for the semi-Markov model would benefit from adaption.

- Another area that could be developed within this multi-state modelling framework is Value of Information analysis. This would be a worthwhile extension to the existing functionality that carries out probabilistic sensitivity analysis. It would allow evaluation of whether future research, and what aspects in particular, would be value for money and worthwhile doing. Specifically, aspects where more information would be of benefit to help reduce uncertainty could be investigated, while at the same time considering the cost of a proposed study to assess whether the outlay would be justified.

- One barrier that may make analysts reluctant to adopt multi-state modelling is that conventionally it requires data at individual patient level data. However, health economists do not always have access to this and anyway economic evaluations are often strengthened by synthesising evidence from various different sources to inform the assumptions with regards to transitions. Consequently, the functions developed as part of this research include those that can be used when individual patient level data are not available or when several different sources are used.

However there is scope to widen the contexts in which they could be used.

- In the case studies used in this research missing covariate information was not such an issue. However missing information is very common in medical research. Therefore, it is intended that the coding will incorporate the option to perform multiple imputation for those users who require it.

To widen the exposure and accessibility of the existing code and intended developments above, it is envisaged that I will bring them all together into a comprehensive `R` package available through CRAN (https://cran.r-project.org/) This would be the ideal platform for those interested in using the code to keep abreast of any developments in this ongoing work. The package would also include vignettes with step-by-step guides on various aspects of the multi-state modelling framework in a cost-effectiveness analysis context. It may also encourage users to suggest developments and improvements of their own to help widen the adoption of the approach among health economic modellers.

Alongside the `R` package it is also my intention to use the shiny application (http://shiny.rstudio.com/) within `R` studio (https://www.rstudio.com/) to create a more user-friendly web-based front-end to the package of functions and related code. Finally, to raise awareness of the approach and engage with those that could benefit from using it, I intend running courses on using `R` for cost-effectiveness analysis at conferences aimed at health economic modellers.

The research conducted as part of this thesis has demonstrated the usefulness of competing risks and multi-state modelling, in areas where they are not widely applied. It has also provided motivation for the future directions detailed above. It is hoped this will encourage further adoption of the multi-state modelling framework, in the many areas where it has not yet reached its full potential.

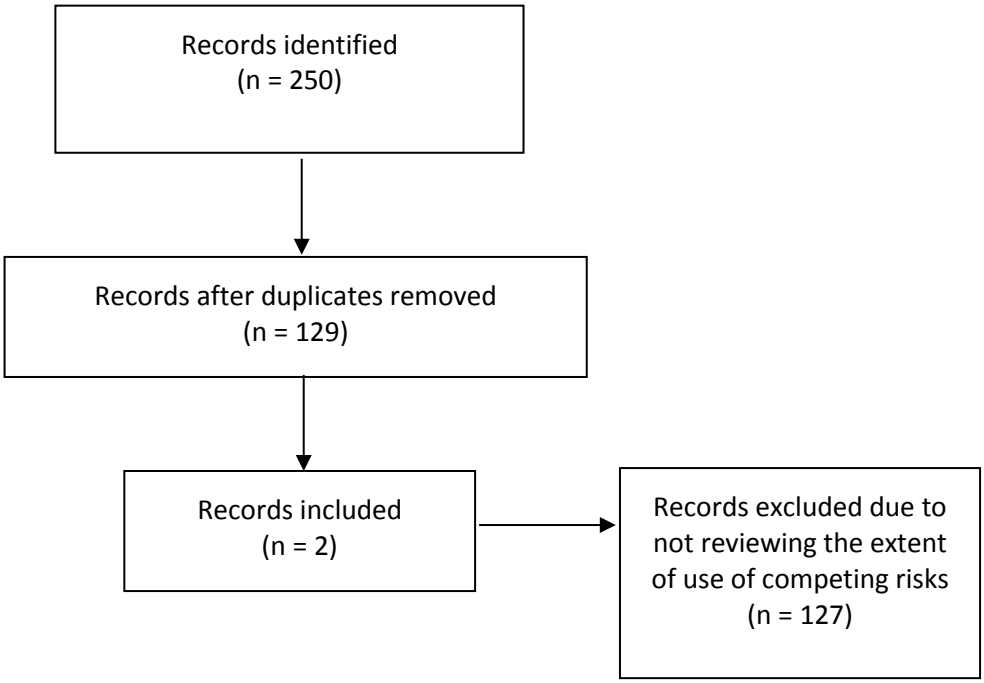
Appendices

Appendix I Literature search of reviews of extent of use of competing risks

A search was conducted using OvidSP (All Resources – including MEDLINE and Embase) of titles, abstracts and keywords to 31/08/2016 using the following search terms:

("competing risks" and review and (liter* or stud* or systematic)).ab,ti,kw.

The following diagram summaries the search leading to 2 contributions.



Appendix II Koller et al. (2012)'s search strategy

Pubmed search for studies on competing risks that used the following search terms:

(a) involving selected biostatistical journals

("competing cause" OR "competing causes" OR "competing risk" OR "competing risks" OR "competing outcome" OR "competing outcomes" OR "competing endpoints") AND (Biom J[Jour] OR Biometrics[Jour] OR Biostatistics[Jour] OR Lifetime Data Anal[Jour] OR Stat Methods Med Res[Jour] OR Stat Med[Jour])

Limits activated: English, Field: Title/Abstract

Searched time period: 1 January 2000 until 28 October 2010

(b) involving high-impact medical journals

("competing cause" OR "competing causes" OR "competing risk" OR "competing risks" OR "competing outcome" OR "competing outcomes" OR "competing endpoints") AND (N Engl J Med[Jour] OR JAMA[Jour] OR bmj[Jour] OR Ann Intern Med[Jour] OR Lancet[Jour] OR PLoS Med[Jour])

Limits activated: English, Field: Title/Abstract

Searched time period: 1 January 2000 until 28 October 2010

and (c) involving core clinical journals

("competing cause" OR "competing causes" OR "competing risk" OR "competing risks" OR "competing outcome" OR "competing outcomes" OR "competing endpoints")

Limits activated: English, Field: Title/Abstract, core clinical journals
Searched time period: 1 January 2000 until 28 October 2010

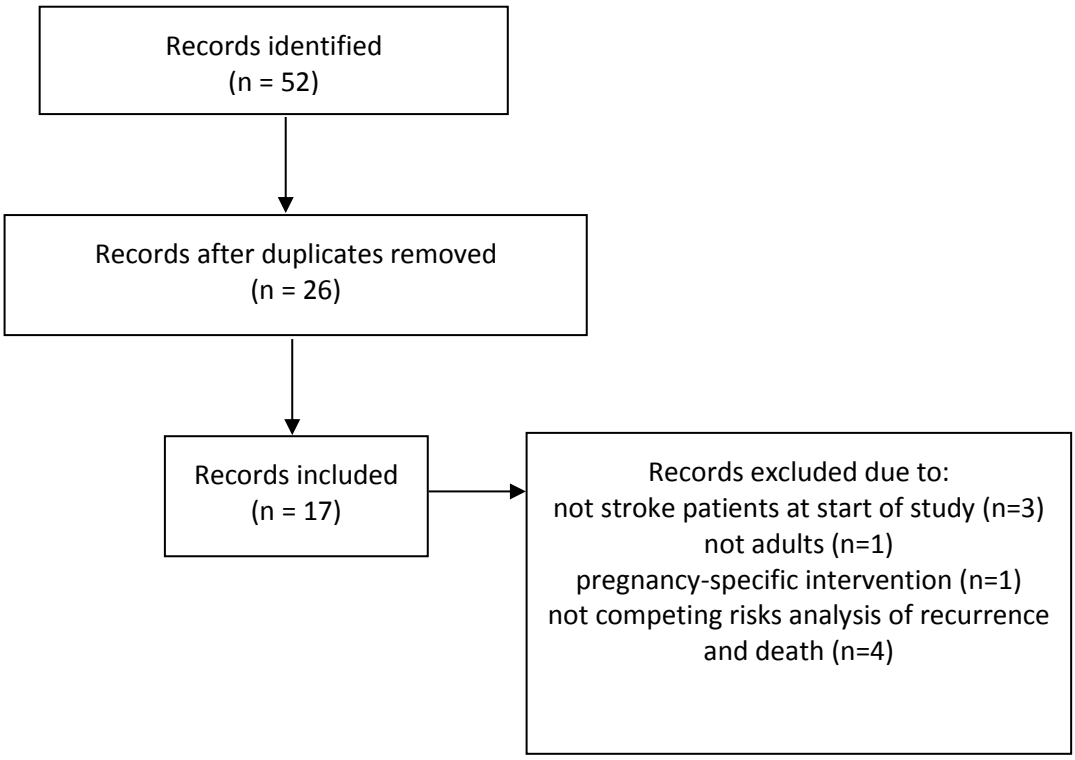
I replicated each of these searches and then also updated them by revising the searched time period to 1 January 2000 until 31 December 2015.

Appendix III Literature search of competing risks analyses of stroke recurrence and death

A search was conducted using OvidSP of titles, abstracts and keywords to 31/08/2016 using the following search terms:

competing cause OR competing causes OR competing risk OR competing risks OR competing outcome OR competing outcomes OR competing endpoints AND stroke AND recur*

The following diagram summaries the search leading to 17 contributions.



Appendix IV Evidence from the literature of competing risks analyses of stroke recurrence and death

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Wollenweber et al. (2014) Germany	ischemic infarct or ICH within 3 days and no prior dementia ≥ 18 years	Study time period not specified proposed follow-up of 5 years plan to include n=600	vague mention of competing risks analysis	n/a - protocol, no modelling presented	post-stroke dementia but stroke recurrence included as a secondary end point	n/a - protocol only
Sun et al. (2013)	index ischemic stroke, ICH or SAH	2000 - 2004 5 year follow-up	standard Cox for death and F&G model for recurrence (but not death without recurrence)	admitting year, age, gender, ethnicity, stroke subtype	all-cause mortality (41.7%) and stroke recurrence (15.7%)	recurrence more likely with haemorrhages and with aging ethnicity, gender and admitting year had no effect on recurrence after haemorrhagic (ICH or SAH) strokes

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Singapore	age-range not specified	n=12,559				earlier year, male and being Malay or other was associated with a high risk of recurrence after ischemic stroke
Arntz et al (2014)	first-ever TIA, ischemic stroke or intracerebral hamorrhage	Study time period not specified	intention to use KM primariliy, F&G model only as secondary	n/a - protocol, no modelling presented	all-cause mortality and risk of recurrent vascular events	n/a - protocol only
		proposed follow-up of 3 years				
Netherlands	aged 18-49	plan to recruit 1500 patients over 3-4 years				
Dhamoon et al. (2016a) + Conference abstract by Dhamoon et al. (2016b)	diabetics with first ischemic stroke	1 Apr 2002 - 31 Mar 2012	F&G model for recurrence with death as competing risk, but no corresponding model for death with recurrence as competing risk	ethnicity, age, sex, income, hypertension, atrial fibrillation, stroke or TIA, MI, CAD, PVD and Charlson score	death, all-cause readmission. Readmission for stroke/TIA, readmission for CAD, composite endpoint of all of the above combined	South Asian diabetic stroke patients aged 65+ have a higher recurrent stroke rate than their non- South Asian counterparts
		maximum follow-up to 31 Mar 2013 with median follow-up of 3.18 years				
Canada	≥ 18 years	n= 25,495	Instead, KM and Cox are used for death	Also medication use in those aged ≥ 65 years		

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Lewsey et al. (2010) + Conference abstract by Inglis et al. (2009)	strokes based on ICD9/10 codes	1986 - 2001	acknowledge should show results for each competing risk	age, sex, socioeconomic status and comorbidity	recurrent stroke within 5 years (10.8%)	Adjusted risk of recurrent stroke decreased by 27% between 1986 and 2001. Adjusted risk of death being first event decreased by 28% between 1986 and 2001.
		5 year follow-up	F&G model for recurrence with death as competing risk		death without recurrence within 5 years (57.8%)	recurrences more likely in males, the most deprived and those with depression, hypertension, diabetes and atrial fibrillation
Scotland		n=128, 511	F&G model for death with recurrence as competing risk			death without recurrence was associated with aging, being female, and having heart failure, cancer, renal failure, Parkinsonism, dementia and falls and fractures. Having hypertension was associated with a reduced risk of death without recurrence

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Andersen et al. (2011)	hospitalized stroke patients	2002-2010			recurrent stroke (10%), death (26%)	men at higher risk of dying *after* stroke (probable typo)
Conference abstract		median follow-up of 2.6 years	F&G model with recurrent stroke and death treated as competing risks	age, stroke severity score, gender, cardiovascular risk factors		stroke recurrence not found to be related to age
Denmark		n=29, 599				higher risk of stroke recurrence in males and with an initial mild stroke
He et al (2015)	First-ever hamorrhagic, ischemic or other strokes: based on ICD9/10 codes	1996 - 2011	F&G model for recurrence with death as competing risk	indigenous status, age, gender, stroke subtype, year of diagnosis, remoteness of residence, atrial fibrillation, hypercholesterolemia, modified Charlson Comorbidity Index and age*indigenous interaction	recurrent stroke (13.5%), long-term death, case fatality	comorbidities did not explain the disparity in stroke recurrence between the indigenous and non-indigenous populations, whereas comorbidities did partly explain the disparity in case fatality and long-term survival
		follow-up to 30 Jun 2013	F&G model for death with recurrence as competing risk			
Australia	age range not specified but ≤ 44, 45-64 and ≥ 65 used in analysis	n=2,105				

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Ovbiagele (2012)	index ischemic stroke	Sep 1996 - May 2003	Cox models with competing risks	low eGFR + others not specified	primary -time to first (recurrent) stroke, MI or vascular death	low eGFR (CKD) found to be associated with primary outcome in adjusted model
Conference abstract		2 year follow-up			secondary - time to first (recurrent) stroke	no evidence of significant association of CKD with first (recurrent) stroke alone
Country not specified		n=3,673				
Stamplecoski et al. (2012)	ischemic stroke, TIA or ICH	July 1 2003 -Mar 31 2008	competing risks Cox modelling with death and readmission as competing risks	age, sex, initial stroke severity, comorbidities, type of stroke, discharge medications	readmission (any reason) within 30d, 1yr and 2yr respectively: 9.6%, 31.4%, 42.3%	no association results given in abstract for recurrent stroke readmission specifically
Conference abstract		30day, 1year and 2year follow-up			death: 2.9%, 13.2% and 19.6%	

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Ovbiagele (2013)	non-disabling cerebral infarction	Sep 1996 - May 2003	competing risk Cox regression but used LRTs	age, sex , race, history of previous stroke, severity of stroke, history of congestive heart failure, history of CAD, history of diabetes, history of carotid endarterectomy , body mass index, alcohol use in the previous year, current smoking, lipid modifier drug use, baseline antithrombotic use, total cholesterol, and low density lipoproteincholesterol	primary - recurrence of stroke	U-shape for association with systolic blood pressure for both primary and secondary outcomes. Low-normal and high have greater hazard of recurrence than normal-high
		2 year follow-up	appear to show "naïve" Cox predictions		secondary - composite outcome of stroke, MI or vascular death	
	≥ 35 years	n=3,680				
Castilloux et al. (2015)	acute ischemic stroke (AIS)	2011-2012	competing risk model due to high death rate	age, gender, past medical conditions, patient's region	AIS recurrence , death	aging and past medical conditions associated with higher hazard of death

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Conference abstract		1 year follow-up	Cox model specified			past medical conditions associated with higher hazard of recurrence
Canada	age range not specified but 85% are 65+	n=6,609				Age not predictor of AIS recurrence. Gender and patient's region not predictive of death or AIS recurrence
Choi et al. (2016)	First-ever stroke patients with atrial fibrillation (AF)	2008 -2012	competing risk Cox model	free-fatty acid, fasting blood sugar, diabetes, previous coronary artery disease, high-density lipoprotein cholesterol, $CHADS_2$ score, CHA_2DS_2 - VAS_C score	(1) any recurrent stroke after 3 years (14.8%) i.e. ischemic or hemorrhagic	free-fatty acid is a predictor for recurrent stroke in AF patients, even after adjusting for covariates in established models. It is associated with a higher hazard
		median follow-up of 17.5 months			(2) ischemic stroke or systemic embolism (ISSE) and	
Korea	age-range not specified	n=279	Gray's test to compare cumulative incidence curves		(3) ischemic stroke	

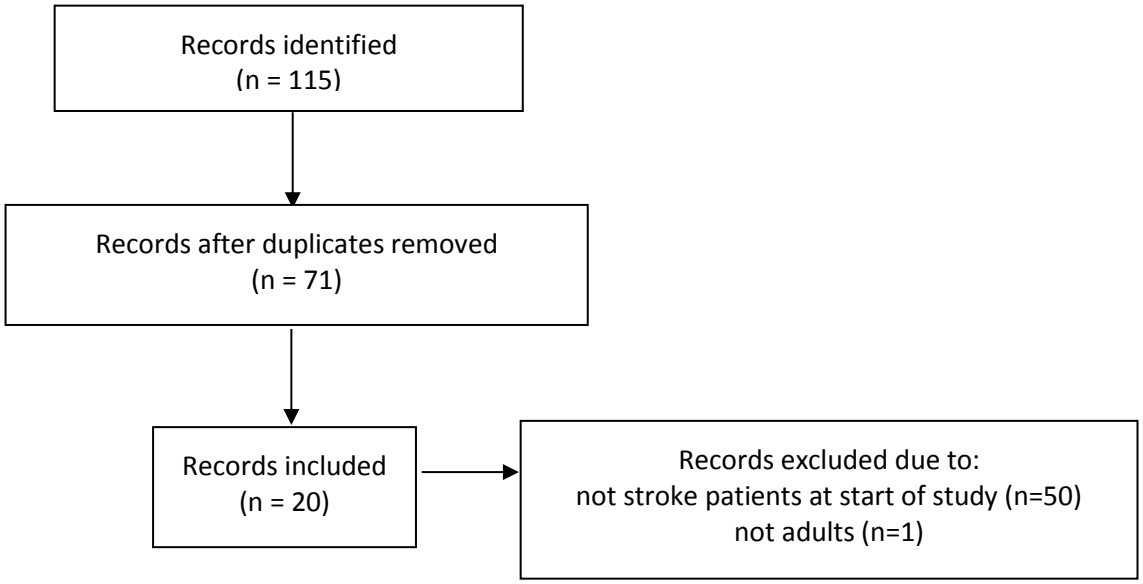
Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Approaches to analysis	covariates used	Primary outcome	Main findings
Rutten-Jacobs et al. (2013)	first-ever TIA, ischemic stroke or intracerebral hamorrhage	1 Jan 1980 - 1 Nov 2010	Gray's test to compare cumulative incidence curves and F&G model	age, sex and decade which index event occurred	any vascular event with separate analyses for fatal or non-fatal stroke and other arterial events	no effect of age or sex on recurrent stroke
		follow-up mean (range): 9.1 (0-31) years				
Netherlands	aged 18-50	n=724				type of stroke and stroke severity found to be associated with recurrent stroke
Wolinsky et al. (2009)	≥ 70 years	1993 - 2005	didn't have stroke recurrence specifically as an outcome, it was combined with first-ever strokes			
USA		n=5,511				
Zhan et al. (2015)	first incidence of stroke	Time period of study and follow-up not specified	F&G model as SHRs specified	smoking + others that are not specified	ischemic stroke recurrence (108/594)	trend association was found between smoking and stroke recurrence but did not reach significance
English abstract for article that is not written in English		n=594	recurrence as outcome, competing events are unspecified but frequency of occurrence is			

Abbreviations used in Table: CAD=coronary artery disease, CKD=chronic kidney disease, ICH= intracerebral hemorrhage, eGRF=estimated Glomerular Filtration Rate, F&G model=Fine and Gray's proportional subdistribution hazards model, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MI=myocardial infarction, PVD=peripheral vascular disease, SAH=subarachnoid hemorrhage, TIA=transient ischemic attack

Appendix V PRISMA diagram for literature search for use of recurrence-free survival as an outcome in stroke

The following diagram summaries the search leading to 20 contributions.



Appendix VI Evidence of contributions to the literature using recurrence-free survival after stroke

Contributions that may have benefited from decomposition of the outcomes

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Elneihoum et al. (1998)	WHO definition used for stroke	1989 - 1992	all-cause mortality and recurrence	KM for survival, Cox for all-cause mortality, Cox for non-fatal recurrences
		3 year follow-up	137 (6%) recurrences	show probability of recurrence figure by age, can only assume it must be naïve KM as paper does not mention any appropriate competing risks approach
Sweden		n=2,290	959 (43.4%) deaths	not clear whether death was treated as a competing risk (i.e. censored) in the modelling of the hazard of recurrence

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Ogasawara et al. (2002b)	unilateral carotid artery or middle cerebral artery occlusion	Jan 1993-Mar 1996	stroke recurrence or death	KM for recurrence-free survival
		2 year follow-up	11 recurrences	Cox for recurrence
Japan	mean (range) age of 57 (38-69)	n=70	8/23 + 3/47 recurrences	
Hillen et al. (2003a)	First-ever strokes	Jan 1995-Aug 2000	recurrence-free survival, recurrence	Use KM for both recurrence-free survival and stroke recurrence
		3 year follow-up	153 (16.6%) recurrences by 5 years	Parametric PH model for recurrence or death. Also parametric PH for recurrence and specify that deaths are censored
UK		n=1,626	recurrence or death, 65.3%	Use likelihood to decide between parametric distributions. Use LRTs to select covariates

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Hillen et al. (2003b)	patients surviving 3 months after first ever stroke	1995-1998	disability + recurrence-free survival over 5 years	Cox for recurrence-free survival
		3 year follow-up	212 recurrences or death	
UK		n=561	66 recurrences, 146 deaths	
Yokota et al. (2004)	first-ever acute strokes	1 Apr 1978 - 31 Mar 1997	recurrence-free survival but	
		min 3 year follow-up, follow-up to 31 Mar 2000 unless had an event	198 recurrences and 286 deaths within 3 years	KM for recurrence-free survival, Cox for recurrence-free survival
Japan		n=1,565		

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Lee et al. (2010)	cryptogenic stroke with patent foramen ovale	Jan 2000 - Apr 2007	recurrent ischemic stroke	KM for recurrence in methods
+ conference abstract by Lee et al. (2009)		median follow-up of 3.5 years, follow-up to Aug 2008	14 (7.7%) recurrences	Cox for recurrence
South Korea		n=181		KM for recurrence-free survival
Toschke et al. (2011)	Diagnosed with index stroke between 1997-2006 with no stroke for at least 2yrs prior	1997-2006	effects of antihypertensive use on: survival after 1yr, recurrence after 1yr	Cox for all-cause mortality, Cox also for recurrence
			survival, recurrence-free survival	Not clear whether death was treated as a competing risk in modelling of hazard of recurrence
UK		n=44,244	~30% deaths by 5years among 3690 patients surviving 90 days 35-40% recurrences by 6 years among 3121 patients surviving 1year	A naïve Cox probability of recurrence is presented

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Chan et al. (2012)	prior history of atherothrombotic/ hemorrhagic stroke		recurrence free survival	KM for recurrence-free survival
		30 months follow-up	10 recurrences, 12 MACEs, 8 deaths	KM for MACE-free survival
Hong Kong	mean(sd) age of 67(11)	n=127		Cox for stroke recurrence and MACE

Abbreviations used: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Contributions not likely to benefit from decomposition of the outcomes

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Nadeau et al. (1992)	vertebrobasilar stroke		30 day and 3 year survival	KM used for survival and recurrences
	all male	3 year follow-up	stroke recurrence	KM curves shown for overall survival and recurrence-free survival
	mean age of 63	n=57	13 recurrences	
Yokota et al. (1998)	ischemic stroke	1987-1995	stroke recurrence	KM for recurrence-free survival
		median follow-up of 2.7 years, range 2 days - 7.8 years	13 recurrences	Cox for recurrence
Japan	mean (range) age of 63 (28-81)	n=105	11 deaths	

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Ogasawara et al. (2002a)	symptomatic internal carotid artery or middle cerebral artery occlusion	Jan 1993-Mar 1996	5 year risk of recurrent stroke	KM and Cox for recurrence-free survival
		5 year follow-up	stroke recurrence or death	
Japan	mean (range) age of 57 (38-69)	n=70	13 recurrences and 4 deaths	
Marnane et al. (2010)	ischemic stroke		stroke recurrence, functional outcome (modified Rankin score)	KM and Cox for recurrence-free survival
Conference abstract		1 year follow-up	6 and 10 recurrences by 7 and 14 days respectively	
Ireland		n=365		

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Kuwashiro et al. (2012)	ischemic stroke	Jun 2007 - Apr 2008	recurrence-free survival	KM for recurrence-free survival
		1year follow-up	25 (9.6%) recurrences	mentions KM to compare recurrence vs non-recurrence groups in methods but not results
Japan	mean age 67	n=256	8 deaths	used logistic regression with recurrence by 1year as outcome
Kim et al. (2014)	acute ishaemic strokes	Mar 2003 - Apr 2011	recurrence-free survival	KM for recurrence-free survival
		90 days follow-up	recurrent ischemic stroke within 90 days	KM for recurrence
USA	median (IQR) age of 70 (58-80)	n=2,378	106 (5.9%) recurrences by 90 days	Cox for recurrence

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Author/ Country	Cohort/ age range	Time period/ length of follow-up/ sample size	Primary outcome + decomposition (n)	Comments on approaches to analysis
Fujimoto et al. (2015)	acute ishaemic strokes		stroke recurrence and death	KM for recurrence-free survival
Conference abstract		median follow-up of 3.2 years	recurrence 12%, 11.8%, 18.2% and 6.7% when split by embolic source	Cox for recurrence and death
+ another conference abstract by Mezuki et al (2013)		n=542	recurrence and death 14.4%, 15.1%, 21.6% and 6.7% when split by embolic source	

Abbreviations used: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Abbreviations used in Tables: IQR=interquartile range, KM=Kaplan-Meier estimate, LRTs= likelihood ratio tests, MACE=major adverse cardiovascular event, PH=proportional hazards, WHO=World Health Organisation

Appendix VII Literature search of broad use of multi-state modelling in medical literature

Pubmed search for multi-state modelling that used the following search terms:

(a) involving selected biostatistical journals

"multi-state model" OR "multistate model" OR "multi state model" OR "illness-death model" OR "illness death model" OR "multi-state models" OR "multistate models" OR "multi state models" OR "illness-death models" OR "illness death models" OR "multi-state modelling" OR "multistate modelling" OR "multi state modelling" OR "illness-death modelling" OR "illness death modelling" OR "multi-state modeling" OR "multistate modeling" OR "multi state modeling" OR "illness-death modeling" OR "illness death modeling" OR "disability model*" OR ("multi-state" AND "Markov") OR ("multistate" AND "Markov") OR ("multi state" AND "Markov") OR ("illness-death" AND "Markov") OR ("illness death" AND "Markov") OR ("disability AND "Markov") OR ("multi-state" AND "semi-Markov") OR ("multistate" AND "semi-Markov") OR ("multi state" AND "semi-Markov") OR ("illness-death" AND "semi-Markov") OR ("illness death" AND "semi-Markov") OR ("disability AND "semi-Markov") OR ("multi-state" AND "model") OR ("multistate" AND "model") OR ("multi state" AND "model") OR ("illness-death" AND "model") OR ("illness death" AND "model") OR ("multi-state" AND "models") OR ("multistate" AND "models") OR ("multi state" AND "models") OR ("illness-death" AND "models") OR ("illness death" AND "models")OR ("multi-state" AND "modelling") OR ("multistate" AND "modelling") OR ("multi state" AND "modelling") OR ("illness-death" AND "modelling") OR ("illness death" AND "modelling") OR ("multi-state" AND "modeling") OR ("multistate" AND "modeling") OR ("multi state" AND "modeling") OR ("illness-death" AND "modeling") OR ("illness death" AND "modeling") AND (Biom J[Jour] OR Biometrics[Jour] OR Biostatistics[Jour] OR Lifetime Data Anal[Jour] OR Stat Methods Med Res[Jour] OR Stat Med[Jour])

Limits activated: Title/Abstract

Searched time period: 1 January 2000 until 31 December 2015

(b) involving high-impact medical journals

"multi-state model" OR "multistate model" OR "multi state model" OR "illness-death model" OR "illness death model" OR "multi-state models" OR "multistate models" OR "multi state models" OR "illness-death models" OR "illness death models" OR "multi-state modelling" OR "multistate modelling" OR "multi state modelling" OR "illness-death modelling" OR "illness death modelling" OR "multi-state modeling" OR "multistate modeling" OR "multi state modeling" OR "illness-death modeling" OR "illness death modeling" OR "disability model*" OR ("multi-state" AND "Markov") OR ("multistate" AND "Markov") OR ("multi state" AND "Markov") OR ("illness-death" AND "Markov") OR ("illness death" AND "Markov") OR ("disability AND "Markov") OR ("multi-state" AND "semi-Markov") OR ("multistate" AND "semi-Markov") OR ("multi state" AND "semi-Markov") OR ("illness-death" AND "semi-Markov") OR ("illness death" AND "semi-Markov") OR ("disability AND "semi-

Markov") OR ("multi-state" AND "model") OR ("multistate" AND "model") OR ("multi state" AND "model") OR ("illness-death" AND "model") OR ("illness death" AND "model") OR ("multi-state" AND "models") OR ("multistate" AND "models") OR ("multi state" AND "models") OR ("illness-death" AND "models") OR ("illness death" AND "models")OR ("multi-state" AND "modelling") OR ("multistate" AND "modelling") OR ("multi state" AND "modelling") OR ("illness-death" AND "modelling") OR ("illness death" AND "modelling") OR ("multi-state" AND "modeling") OR ("multistate" AND "modeling") OR ("multi state" AND "modeling") OR ("illness-death" AND "modeling") OR ("illness death" AND "modeling") AND (N Engl J Med[Jour] OR JAMA[Jour] OR bmj[Jour] OR Ann Intern Med[Jour] OR Lancet[Jour] OR PLoS Med[Jour])

Limits activated: Title/Abstract

Searched time period: 1 January 2000 until 31 December 2015

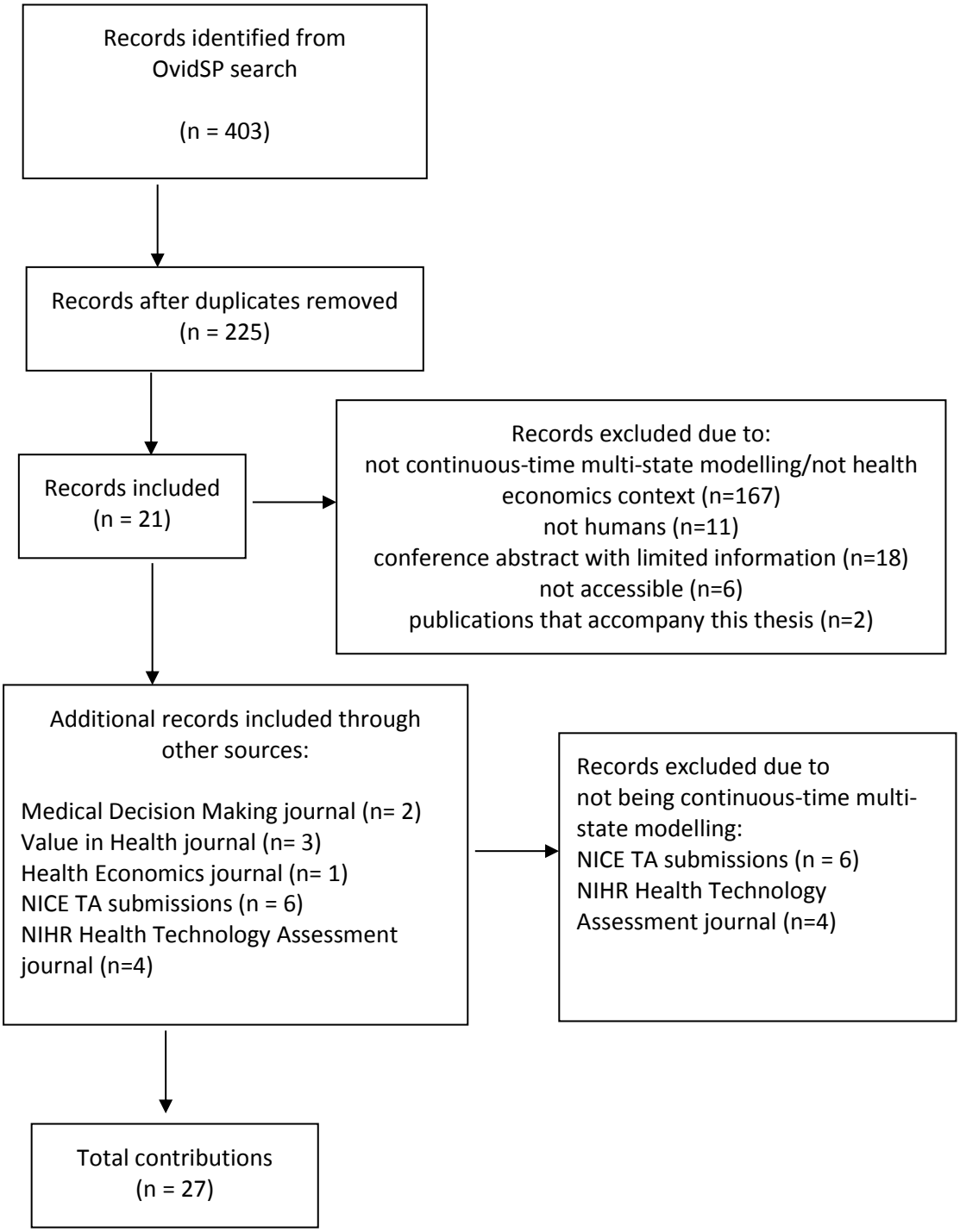
and (c) involving core clinical journals

"multi-state model" OR "multistate model" OR "multi state model" OR "illness-death model" OR "illness death model" OR "multi-state models" OR "multistate models" OR "multi state models" OR "illness-death models" OR "illness death models" OR "multi-state modelling" OR "multistate modelling" OR "multi state modelling" OR "illness-death modelling" OR "illness death modelling" OR "multi-state modeling" OR "multistate modeling" OR "multi state modeling" OR "illness-death modeling" OR "illness death modeling" OR "disability model*" OR ("multi-state" AND "Markov") OR ("multistate" AND "Markov") OR ("multi state" AND "Markov") OR ("illness-death" AND "Markov") OR ("illness death" AND "Markov") OR ("disability AND "Markov") OR ("multi-state" AND "semi-Markov") OR ("multistate" AND "semi-Markov") OR ("multi state" AND "semi-Markov") OR ("illness-death" AND "semi-Markov") OR ("illness death" AND "semi-Markov") OR ("disability AND "semi-Markov") OR ("multi-state" AND "model") OR ("multistate" AND "model") OR ("multi state" AND "model") OR ("illness-death" AND "model") OR ("illness death" AND "model") OR ("multi-state" AND "models") OR ("multistate" AND "models") OR ("multi state" AND "models") OR ("illness-death" AND "models") OR ("illness death" AND "models")OR ("multi-state" AND "modelling") OR ("multistate" AND "modelling") OR ("multi state" AND "modelling") OR ("illness-death" AND "modelling") OR ("illness death" AND "modelling") OR ("multi-state" AND "modeling") OR ("multistate" AND "modeling") OR ("multi state" AND "modeling") OR ("illness-death" AND "modeling") OR ("illness death" AND "modeling")

Limits activated: Title/Abstract, core clinical journals
Searched time period: 1 January 2000 until 31 December 2015

In addition, for each of the three searches above, the term “cost*” was added to help identify studies involving cost-effectiveness analysis. If found, these were removed. This was the case for 1, 1 and 4 articles in biostatistical, high-impact medical and core clinical journals respectively.

Appendix VIII PRISMA diagram for literature search of use of multi-state modelling in health economics literature



Appendix IX

Table A1 Cost per QALYs results in all 216 combinations

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
Gompertz	**Gompertz**	**Gompertz**	**0.331**	**10149**	**30702**
exponential	Gompertz	Gompertz	0.844	9620	11393
exponential	exponential	Gompertz	0.841	9613	11424
exponential	Log-logistic	Gompertz	0.829	9548	11515
exponential	Log normal	Gompertz	0.829	9570	11543
exponential	Gompertz	generalised gamma	0.797	9324	11699
exponential	generalised gamma	Gompertz	0.805	9534	11839
exponential	Weibull	Gompertz	0.807	9557	11844
exponential	exponential	generalised gamma	0.791	9372	11848
exponential	Log-logistic	generalised gamma	0.782	9261	11849
exponential	Log normal	generalised gamma	0.783	9293	11867
exponential	Gompertz	exponential	0.756	9160	12117
exponential	Weibull	generalised gamma	0.760	9275	12205
exponential	generalised gamma	generalised gamma	0.752	9215	12249
exponential	exponential	Weibull	0.753	9241	12272
exponential	Log-logistic	Weibull	0.750	9217	12294
exponential	Gompertz	Log-logistic	0.777	9585	12343
exponential	Weibull	Weibull	0.735	9215	12539
exponential	Gompertz	Weibull	0.728	9221	12667
exponential	generalised gamma	Log-logistic	0.729	9359	12841
exponential	Log normal	Weibull	0.714	9175	12850
exponential	Log normal	Log-logistic	0.713	9466	13279
exponential	generalised gamma	exponential	0.676	9019	13336
exponential	generalised gamma	Weibull	0.683	9138	13384
exponential	Log-logistic	Log-logistic	0.705	9480	13450
exponential	exponential	exponential	0.665	9048	13597
exponential	Log normal	exponential	0.664	9044	13611
exponential	Weibull	Log-logistic	0.697	9503	13643
exponential	Log-logistic	exponential	0.654	9095	13907
exponential	Weibull	Log normal	0.664	9239	13913
exponential	Log-logistic	Log normal	0.659	9215	13993
exponential	Weibull	exponential	0.651	9134	14040
exponential	Gompertz	Log normal	0.650	9170	14096
Log normal	Gompertz	Gompertz	0.691	9847	14245
exponential	exponential	Log-logistic	0.659	9503	14427

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
Log normal	Log normal	Gompertz	0.675	9780	14481
Log-logistic	Gompertz	Gompertz	0.687	9992	14535
Log normal	Gompertz	generalised gamma	0.645	9518	14760
exponential	generalised gamma	Log normal	0.609	9028	14828
exponential	Log normal	Log normal	0.617	9163	14848
Log normal	Weibull	Gompertz	0.666	9893	14860
Log normal	Log normal	generalised gamma	0.630	9459	15010
Log normal	exponential	Gompertz	0.654	9848	15050
Log-logistic	Log normal	Gompertz	0.652	9820	15061
Log normal	Log-logistic	Gompertz	0.649	9835	15162
Log-logistic	Gompertz	generalised gamma	0.639	9696	15171
Log normal	generalised gamma	Gompertz	0.631	9726	15405
Log normal	Weibull	generalised gamma	0.616	9595	15587
exponential	exponential	Log normal	0.595	9325	15685
Log normal	Gompertz	Weibull	0.601	9488	15796
Log normal	exponential	generalised gamma	0.604	9577	15843
Log-logistic	Log normal	generalised gamma	0.598	9552	15972
Log normal	Log-logistic	generalised gamma	0.596	9534	16007
Log-logistic	exponential	Gompertz	0.612	9863	16111
Log normal	generalised gamma	generalised gamma	0.579	9398	16228
Log normal	Weibull	Weibull	0.577	9527	16508
Log normal	Log normal	Weibull	0.571	9420	16510
Log normal	Log normal	Log-logistic	0.591	9783	16565
Log normal	exponential	Weibull	0.573	9510	16602
Log-logistic	Gompertz	Weibull	0.579	9741	16822
Log-logistic	generalised gamma	Gompertz	0.581	9775	16836
Log normal	Log-logistic	Log-logistic	0.577	9750	16910
Log normal	Log-logistic	Weibull	0.551	9425	17118
Log-logistic	Gompertz	exponential	0.557	9556	17144
Log normal	Weibull	Log-logistic	0.568	9781	17214
Log-logistic	Log normal	Weibull	0.546	9467	17344
Log-logistic	exponential	generalised gamma	0.547	9611	17556
Log normal	Log normal	exponential	0.536	9435	17587
Log-logistic	exponential	Weibull	0.539	9495	17604
Log normal	exponential	Log-logistic	0.550	9703	17633
Log normal	exponential	exponential	0.528	9352	17717
Log normal	generalised gamma	Weibull	0.532	9426	17725
Weibull	Log normal	Gompertz	0.569	10111	17781

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
Log normal	Gompertz	exponential	0.521	9342	17917
Log normal	Log normal	Log normal	0.529	9515	17993
Log-logistic	Gompertz	Log-logistic	0.557	10080	18084
Weibull	Weibull	Gompertz	0.551	10000	18157
Log normal	Gompertz	Log-logistic	0.543	9876	18192
Log-logistic	Weibull	Gompertz	0.546	9946	18219
generalised gamma	Gompertz	Gompertz	0.558	10186	18270
Log-logistic	Log-logistic	Gompertz	0.548	10033	18316
Log normal	Weibull	Log normal	0.525	9633	18353
Log-logistic	generalised gamma	generalised gamma	0.512	9433	18424
Log-logistic	Log normal	Log-logistic	0.524	9703	18514
Weibull	Gompertz	Gompertz	0.542	10084	18615
Log-logistic	generalised gamma	Log-logistic	0.528	9890	18741
Log normal	generalised gamma	Log-logistic	0.524	9909	18897
Log normal	Log-logistic	Log normal	0.502	9519	18948
Log normal	Log-logistic	exponential	0.496	9416	18997
Log normal	exponential	Log normal	0.498	9502	19095
Weibull	exponential	Gompertz	0.530	10143	19127
Log-logistic	exponential	exponential	0.491	9421	19200
Weibull	Log normal	generalised gamma	0.505	9713	19219
Log normal	Weibull	exponential	0.489	9392	19219
Log-logistic	Weibull	generalised gamma	0.505	9726	19244
generalised gamma	generalised gamma	Gompertz	0.523	10068	19250
Log-logistic	generalised gamma	Weibull	0.480	9333	19442
Log normal	generalised gamma	exponential	0.480	9350	19483
Log-logistic	Log normal	exponential	0.483	9410	19489
Weibull	generalised gamma	Gompertz	0.508	9983	19652
Log-logistic	Log-logistic	generalised gamma	0.495	9734	19658
generalised gamma	Gompertz	generalised gamma	0.501	9872	19693
generalised gamma	Log normal	Gompertz	0.509	10029	19721
Log-logistic	Log-logistic	Log-logistic	0.502	9899	19735
Log-logistic	generalised gamma	exponential	0.471	9353	19862
generalised gamma	Gompertz	Weibull	0.489	9753	19933
Log-logistic	Log-logistic	exponential	0.481	9682	20125

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
generalised gamma	Weibull	Gompertz	0.491	9908	20161
Weibull	Log normal	Log-logistic	0.508	10285	20233
generalised gamma	Gompertz	Log-logistic	0.508	10288	20244
Weibull	Log normal	Weibull	0.477	9667	20256
Log normal	generalised gamma	Log normal	0.470	9551	20303
generalised gamma	Log-logistic	Gompertz	0.492	9989	20320
Log normal	Gompertz	Log normal	0.472	9590	20330
Log-logistic	exponential	Log-logistic	0.487	9911	20352
Log-logistic	Weibull	Log-logistic	0.478	9747	20396
generalised gamma	exponential	Gompertz	0.489	9983	20413
Weibull	Gompertz	Log-logistic	0.502	10267	20470
Weibull	Weibull	generalised gamma	0.472	9693	20515
Log-logistic	Log-logistic	Weibull	0.470	9721	20664
Log-logistic	Gompertz	Log normal	0.476	9851	20686
Weibull	Gompertz	generalised gamma	0.468	9724	20788
Weibull	exponential	Weibull	0.468	9774	20879
Log-logistic	Weibull	Weibull	0.462	9653	20880
Weibull	Weibull	Weibull	0.460	9642	20964
Log-logistic	Weibull	exponential	0.456	9568	20969
Weibull	Gompertz	Weibull	0.456	9733	21334
Weibull	generalised gamma	generalised gamma	0.448	9578	21364
generalised gamma	Log normal	Log-logistic	0.477	10286	21555
Weibull	generalised gamma	Log-logistic	0.465	10060	21636
Weibull	Log-logistic	Gompertz	0.458	9986	21801
Log-logistic	Log normal	Log normal	0.440	9625	21871
Weibull	exponential	generalised gamma	0.445	9759	21926
Weibull	exponential	Log-logistic	0.466	10289	22065
generalised gamma	generalised gamma	generalised gamma	0.436	9648	22146
generalised gamma	Weibull	Log-logistic	0.449	9990	22248
Log-logistic	exponential	Log normal	0.439	9779	22289
generalised gamma	Log normal	generalised gamma	0.433	9649	22294
Weibull	Gompertz	exponential	0.437	9754	22336
generalised gamma	generalised gamma	exponential	0.433	9744	22479
Weibull	generalised gamma	Weibull	0.435	9802	22513
Weibull	Log normal	exponential	0.417	9440	22617

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
generalised gamma	Gompertz	exponential	0.422	9575	22674
generalised gamma	Log-logistic	Log-logistic	0.439	10002	22800
Log-logistic	Log-logistic	Log normal	0.423	9664	22872
generalised gamma	Weibull	generalised gamma	0.421	9654	22956
generalised gamma	exponential	Weibull	0.416	9612	23131
Weibull	exponential	exponential	0.418	9740	23281
generalised gamma	exponential	generalised gamma	0.416	9693	23307
Weibull	Weibull	Log-logistic	0.438	10203	23321
generalised gamma	Log normal	Weibull	0.416	9740	23389
Weibull	Log normal	Log normal	0.425	9988	23516
Log-logistic	Weibull	Log normal	0.408	9630	23622
generalised gamma	generalised gamma	Log-logistic	0.421	9941	23626
generalised gamma	Log-logistic	generalised gamma	0.407	9707	23869
Weibull	Log-logistic	generalised gamma	0.408	9746	23916
Log-logistic	generalised gamma	Log normal	0.403	9635	23926
Weibull	Weibull	exponential	0.393	9420	23949
Weibull	Log-logistic	Log-logistic	0.422	10125	23985
generalised gamma	Weibull	Weibull	0.394	9539	24237
Weibull	Gompertz	Log normal	0.412	9995	24289
generalised gamma	generalised gamma	Weibull	0.396	9633	24354
Weibull	Log-logistic	exponential	0.394	9599	24377
generalised gamma	exponential	Log-logistic	0.413	10106	24463
Weibull	exponential	Log normal	0.404	9960	24676
generalised gamma	generalised gamma	Log normal	0.394	9864	25031
generalised gamma	Log-logistic	Weibull	0.379	9592	25281
Weibull	Log-logistic	Weibull	0.380	9624	25342
generalised gamma	Log normal	exponential	0.376	9624	25567
generalised gamma	Weibull	exponential	0.370	9585	25896
generalised gamma	Log-logistic	exponential	0.370	9621	26027
generalised gamma	Weibull	Log normal	0.376	9833	26135

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
generalised gamma	Log normal	Log normal	0.376	9882	26250
Weibull	generalised gamma	exponential	0.360	9509	26402
generalised gamma	Gompertz	Log normal	0.375	9913	26420
Weibull	generalised gamma	Log normal	0.367	9702	26443
Weibull	Weibull	Log normal	0.376	9955	26509
generalised gamma	exponential	exponential	0.341	9481	27813
generalised gamma	Log-logistic	Log normal	0.340	9686	28504
generalised gamma	exponential	Log normal	0.349	9966	28583
Gompertz	Weibull	Gompertz	0.340	10218	30056
Weibull	Log-logistic	Log normal	0.318	9799	30769
Gompertz	Log-logistic	Gompertz	0.316	10202	32299
Gompertz	Log normal	Gompertz	0.314	10159	32353
Gompertz	exponential	Gompertz	0.287	10097	35230
Gompertz	Weibull	generalised gamma	0.258	9860	38175
Gompertz	Gompertz	generalised gamma	0.251	9801	39043
Gompertz	generalised gamma	Gompertz	0.252	10011	39781
Gompertz	Log-logistic	generalised gamma	0.230	9799	42651
Gompertz	Log normal	generalised gamma	0.219	9775	44704
Gompertz	Weibull	Log-logistic	0.226	10279	45570
Gompertz	generalised gamma	Log-logistic	0.220	10235	46418
Gompertz	Log-logistic	Log-logistic	0.217	10244	47134
Gompertz	Gompertz	Log-logistic	0.208	10191	48889
Gompertz	Weibull	Weibull	0.194	9686	49907
Gompertz	Gompertz	Weibull	0.188	9676	51582
Gompertz	exponential	Log-logistic	0.197	10179	51781
Gompertz	exponential	generalised gamma	0.179	9666	53957
Gompertz	Log normal	Log-logistic	0.180	10181	56610
Gompertz	Log-logistic	Weibull	0.165	9707	58659
Gompertz	Log normal	Weibull	0.164	9692	59199
Gompertz	Log-logistic	exponential	0.159	9865	62101
Gompertz	Weibull	exponential	0.156	9744	62327
Gompertz	Gompertz	exponential	0.147	9686	65697

Distribution used for:					
progression-free -> progression	progression-free -> death	progression-> death	incremental QALYs	incremental Costs (£)	Cost per QALY gained (£)
Gompertz	exponential	Weibull	0.146	9668	66291
Gompertz	generalised gamma	Weibull	0.144	9567	66552
Gompertz	generalised gamma	generalised gamma	0.143	9516	66668
Gompertz	Log normal	exponential	0.128	9780	76128
Gompertz	generalised gamma	exponential	0.120	9671	80492
Gompertz	exponential	exponential	0.114	9635	84734
Gompertz	Gompertz	Log normal	0.108	9938	92303
Gompertz	generalised gamma	Log normal	0.091	9873	108802
Gompertz	Weibull	Log normal	0.090	9882	109412
Gompertz	Log-logistic	Log normal	0.087	9892	113980
Gompertz	exponential	Log normal	0.079	9824	124952
Gompertz	Log normal	Log normal	0.028	9781	344585

References

Aalen, O. O., Borgan, O. & Gjessing, H. K. 2008. Survival and Event History Analysis: A Process Point of View. *Survival and Event History Analysis: A Process Point of View*, 1-539.

Aggarwal A., T. F., C. Chamberlain, C. Davis, R. Sullivan 2017. Do patient access schemes for high-cost cancer drugs deliver value to society?—lessons from the NHS Cancer Drugs Fund. *Ann Oncol*.

Akaike, H. 1974. New Look at Statistical-Model Identification. *Ieee Transactions on Automatic Control,* Ac19, 716-723.

Allignol, A., Schumacher, M., Wanner, C., Drechsler, C. & Beyersmann, J. 2011. Understanding competing risks: a simulation point of view. *Bmc Medical Research Methodology,* 11.

Amorim, L. D. A. F. & Cai, J. W. 2015. Modelling recurrent events: a tutorial for analysis in epidemiology. *International Journal of Epidemiology,* 44, 324-333.

Andersen, K. K., Andersen, Z. J. & Olsen, T. S. 2011. Stroke recurrence: Higher risk in men and patients with minor strokes but no association with age. *Cerebrovascular Diseases,* 31, 211.

Andersen, P. K., Abildstrom, S. Z. & Rosthoj, S. 2002. Competing risks as a multi-state model. *Statistical Methods in Medical Research,* 11, 203-215.

Andersen, P. K., Geskus, R. B., de Witte, T. & Putter, H. 2012. Competing risks in epidemiology: possibilities and pitfalls. *International Journal of Epidemiology,* 41, 861-870.

Andersen, P. K. & Keiding, N. 2002. Multi-state models for event history analysis. *Statistical Methods in Medical Research,* 11, 91-115.

Andersen, P. K. & Perme, M. P. 2008. Inference for outcome probabilities in multi-state models. *Lifetime Data Analysis,* 14, 405-431.

Andersson, T. M. L., Dickman, P. W., Eloranta, S., Lambe, M. & Lambert, P. C. 2013. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine,* 32, 5286-5300.

Arefian, H., Hagel, S., Heublein, S., Rissner, F., Scherag, A., Brunkhorst, F. M., Baldessarini, R. J. & Hartmann, M. 2016. Extra length of stay and costs because of health care-associated infections at a German university hospital. *American Journal of Infection Control,* 44, 160-166.

Arntz, R. M., van Alebeek, M. E., Synhaeve, N. E., Brouwers, P. J., van Dijk, G. W., Gons, R. A., den Heijer, T., de Kort, P. L. M., de Laat, K. F., van Norden, A. G., Vermeer, S. E., van der Vlugt, M. J., Kessels, R. P. C., van Dijk, E. J. & de Leeuw, F. E. 2014. Observational Dutch Young

Symptomatic StrokE studY (ODYSSEY): study rationale and protocol of a multicentre prospective cohort study. *Bmc Neurology,* 14.

Asaria, M., Walker, S., Palmer, S., Gale, C. P., Shah, A. D., Abrams, K. R., Crowther, M., Manca, A., Timmis, A., Hemingway, H. & Sculpher, M. 2016. Using electronic health records to predict costs and outcomes in stable coronary artery disease. *Heart,* 102, 755-762.

Athanasakis, K., Boubouchairopoulou, N., Retsa, M. P., Maiese, E. M., Elbasha, E. H. & Kyriopoulos, J. 2014. Cost-effectiveness analysis of ralt egravir in HIV-infected treatment naive patients in Greece. *Value in Health,* 17 (7), A675.

Austin, P. C., Lee, D. S. & Fine, J. P. 2016. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation,* 133, 601-609.

Bagust, A. & Beale, S. 2014. Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation: An Alternative Approach. *Medical Decision Making,* 34, 343-351.

Bakoyannis, G., Siannis, F. & Touloumi, G. 2010. Modelling competing risks data with missing cause of failure. *Statistics in Medicine,* 29, 3172-3185.

Bakoyannis, G. & Touloumi, G. 2012. Practical methods for competing risks data: A review. *Statistical Methods in Medical Research,* 21, 257-272.

Barnett, A. G., Beyersmann, J., Allignol, A., Rosenthal, V. D., Graves, N. & Wolkewitz, M. The Time-Dependent Bias and its Effect on Extra Length of Stay due to Nosocomial Infection. *Value in Health,* 14, 381-386.

Begun, A., Icks, A., Waldeyer, R., Landwehr, S., Koch, M. & Giani, G. 2013. Identification of a Multistate Continuous-Time Nonhomogeneous Markov Chain Model for Patients with Decreased Renal Function. *Medical Decision Making,* 33, 298-306.

Belot, A., Abrahamowicz, M., Remontet, L. & Giorgi, R. 2010. Flexible modeling of competing risks in survival analysis. *Statistics in Medicine,* 29, 2453-2468.

Bongers, M. L., de Ruysscher, D., Oberije, C., Lambin, P., Uyl-de Groot, C. A. & Coupe, V. M. H. 2016. Multistate Statistical Modeling: A Tool to Build a Lung Cancer Microsimulation Model That Includes Parameter Uncertainty and Patient Heterogeneity. *Medical Decision Making,* 36, 86-100.

Bradburn, M. J., Clark, T. G., Love, S. B. & Altman, D. G. 2003. Survival Analysis Part III: Multivariate data analysis - choosing a model and assessing its adequacy and fit. *British Journal of Cancer,* 89, 605-611.

Briggs, A. & Sculpher, M. 1998. An introduction to Markov modelling for economic evaluation. *Pharmacoeconomics,* 13, 397-409.

Briggs, A., Sculpher, M. & Claxton , K. 2006. *Decision Modelling for Health Economic Evaluation*, Oxford University Press.

Cao, Q., Buskens, E., Feenstra, T., Jaarsma, T., Hillege, H. & Postmus, D. 2016. Continuous-Time Semi-Markov Models in Health Economic Decision Making: An Illustrative Example in Heart Failure Disease Management. *Medical Decision Making,* 36, 59-71.

Cao, Q., Postmus, D., Hillege, H. L. & Buskens, E. 2013. Probability Elicitation to Inform Early Health Economic Evaluations of New Medical Technologies: A Case Study in Heart Failure Disease Management. *Value in Health,* 16, 529-535.

Castelli, C., Combescure, C., Foucher, Y. & Daures, J. P. 2007. Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. *Statistics in Medicine,* 26, 5557-5571.

Castilloux, A. M., Tore Jorgensen, K., Coste, F. & Moride, Y. 2015. Outcomes and patterns of health care usage among acute ischemic stroke patients in Canada. *Pharmacoepidemiology and Drug Safety,* 24, 484-485.

Chan, Y. H., Lau, K. K., Yiu, K. H., Siu, C. W., Chan, H. T., Li, S. W., Tam, S., Lam, T. H., Lau, C. P. & Tse, H. F. 2012. Prospective observational study of isoflavone and the risk of stroke recurrence: Potential clinical implications beyond vascular function. *Journal of Nutrition Health & Aging,* 16, 383-388.

Chappell, R. 2012. Competing Risk Analyses: How Are They Different and Why Should You Care? *Clinical Cancer Research,* 18, 2127-2129.

Chaudhary, M. A., Elbasha, E. H., Pereira, R. & Kumar, R. N. 2010. A continuous-time economic model to evaluate raltegravir use strategies in treatment-naive hiv-1 patients in portugal. *Value in Health,* 13 (7), A436.

Chen, W.-Y., Cheng, H.-C., Wang, J.-D. & Sheu, B.-S. 2013. Factors That Affect Life Expectancy of Patients With Gastric Adenocarcinoma. *Clinical Gastroenterology & Hepatology,* 11, 1595-1600.

Choi, J. Y., Jung, J. M., Kwon, D. Y., Park, M. H., Kim, J. H., Oh, K., Koh, S. B. & Seo, W. K. 2016. Free fatty acid as an outcome predictor of atrial fibrillation-associated stroke. *Annals of Neurology,* 79, 317-325.

Chu, P.-C. M. D., Wang, J.-D. M. D. S., Hwang, J.-S. P. & Chang, Y.-Y. M. S. 2008a. Estimation of Life Expectancy and the Expected Years of Life Lost in Patients with Major Cancers: Extrapolation of Survival Curves under High-Censored Rates. *Value in Health,* 11, 1102-1109.

Chu, P. C., Hwang, J. S., Wang, J. D. & Chang, Y. Y. 2008b. Estimation of the financial burden to the national health insurance for patients with major cancers in Taiwan. *Journal of the Formosan Medical Association,* 107, 54-63.

Clarke, A., Pulikottil-Jacob, R., Connock, M., Suri, G., Kandala, N. B., Maheswaran, H., Banner, N. R. & Sutcliffe, P. 2014. Cost-effectiveness of left ventricular assist devices (LVADs) for patients with advanced heart failure: Analysis of the British NHS bridge to transplant (BTT) program. *International Journal of Cardiology,* 171, 338-345.

Commenges, D. 1999. Multi-state models in epidemiology. *Lifetime Data Analysis,* 5, 315-327.

Commenges, D. 2002. Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research,* 11, 167-182.

Connock, M., Hyde, C. & Moore, D. 2011. Cautions Regarding the Fitting and Interpretation of Survival Curves Examples from NICE Single Technology Appraisals of Drugs for Cancer. *Pharmacoeconomics,* 29, 827-837.

Dabrowska, D. 1995. Estimation of transition probabilities and bootstrap in a semiparametric Markov renewal model. *Nonparametric Statistics* 5, 237-259.

Davis, S., Stevenson, M., Tappenden, P. & Wailoo, A. J. 2014. NICE DSU Technical Support Document 15: Cost-effectiveness modelling using patient-level simulation.

De Angelis, G., Allignol, A., Murthy, A., Wolkewitz, M., Beyersmann, J., Safran, E., Schrenzel, J., Pittet, D. & Harbarth, S. 2011. Multistate modelling to estimate the excess length of stay associated with meticillin-resistant Staphylococcus aureus colonisation and infection in surgical patients. *Journal of Hospital Infection,* 78, 86-91.

de Wreede, L., Fiocco, M. & Putter, H. 2010. The mstate package for estimation and prediction in non- and semi-parametric multi-state and competing risks models. *Comput Methods Programs Biomed,* 99, 261-274.

de Wreede, L., Fiocco, M. & Putter, H. 2011. mstate: An R Package for the Analysis of

Competing Risks and Multi-State Models. *Journal of Statistical Software,* 38.

Demiris, N., Lunn, D. & Sharples, L. D. 2015. Survival extrapolation using the poly-Weibull model. *Statistical Methods in Medical Research,* 24, 287-301.

Demiris, N. & Sharples, L. D. 2006. Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies. *Statistics in Medicine,* 25, 1960-1975.

Dhamoon, M. S., Zhou, L., Stamplecoski, M., Kapral, M. & Shah, B. 2016a. Stroke recurrence among South Asians with diabetes in Ontario, Canada. *International Journal of Stroke,* 15, 15.

Dhamoon, M. S., Zhou, L., Stamplecoski, M., Kapral, M. K. & Shah, B. R. 2016b. Stroke recurrence among south asians with diabetes in Ontario, Canada.

Stroke. Conference: American Heart Association/American Stroke Association, 47.

Dignam, J. J. & Kocherginsky, M. N. 2008. Choice and interpretation of statistical tests used when competing risks are present. *Journal of Clinical Oncology,* 26, 4027-4034.

Dignam, J. J., Zhang, Q. & Kocherginsky, M. 2012. The Use and Interpretation of Competing Risks Regression Models. *Clinical Cancer Research,* 18, 2301-2308.

Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J. & Stoddart, G. L. 2005. *Methods for the Economic Evaluation of Health Care Programmes*, Oxford University Press.

Egger, M. E., Myers, J. A., Arnold, F. W., Pass, L. A., Ramirez, J. A. & Brock, G. N. 2016. Cost effectiveness of adherence to IDSA/ATS guidelines in elderly patients hospitalized for Community-Aquired Pneumonia. *Bmc Medical Informatics and Decision Making,* 16.

Elneihoum, A. M., Goransson, M., Falke, P. & Janzon, L. 1998. Three-year survival and recurrence after stroke in Malmo, Sweden - An analysis of stroke registry data. *Stroke,* 29, 2114-2117.

Erdesz, D., Nagy, L., Brandtmuller, A., Kiss, Z., Chaudhary, M. A., Kumar, R. N. & Elbasha, E. H. 2010. An economic evaluation of raltegravir for the treatment of antiretroviral-naive hiv-1 infected patients in Hungary. *Value in Health,* 13 (7), A439.

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H. & Proust-Lima, C. 2016. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in Medicine,* 35, 3933-3948.

Fine, J. P. & Gray, R. J. 1999. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American Statistical Association,* 94, 496-509.

Fiocco, M., Putter, H. & van Houwelingen, H. C. 2008. Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models. *Statistics in Medicine,* 27, 4340-4358.

Fu, Parikh, C. R. & Zhou, B. 2017. Penalized variable selection in competing risks regression. *Lifetime Data Anal* 23, 353-376.

Fujimoto, S., Osaki, M., Kumamoto, M., Kanazawa, M., Tagawa, N., Ishitsuka, T. & Kitazono, T. 2015. Impact of Aortic Arch Atheroma for Stroke Recurrence in Patients With or Without Atrial Fibrillation. *Stroke,* 46.

Gail, M. 1975. A review and critique of some models used in

competing risk analysis. *Biometrics* 31, 209-222.

Gelber, R. D., Cole, B. F., Gelber, S. & Goldhirsch, A. 1995. Comparing Treatments Using Quality-Adjusted Survival - the Q-Twist Method. *American Statistician,* 49, 161-169.

Gelber, R. D., Goldhirsch, A. & Cavalli, F. 1991. Quality-of-Life-Adjusted Evaluation of Adjuvant Therapies for Operable Breast-Cancer. *Annals of Internal Medicine,* 114, 621-628.

Glasziou, P. P., Simes, R. J. & Gelber, R. D. 1990. Quality Adjusted Survival Analysis. *Statistics in Medicine,* 9, 1259-1276.

Gooley, T. A., Leisenring, W., Crowley, J. & Storer, B. E. 1999. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. *Statistics in Medicine,* 18, 695-706.

Gray, R. J. 1988. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Annals of Statistics,* 16, 1141-1154.

Grieve, R., Abrams, K., Claxton, K., Goldacre, B., James, N., Nicholl, J., Parmar, M., Parker, C., Sekhon, J. S., Smeeth, L., Spiegelhalter, D. & Sculpher, M. 2016. Cancer Drugs Fund requires further reform. *Bmj-British Medical Journal,* 354, i5090.

Grieve, R., Hawkins, N. & Pennington, M. 2013. Extrapolation of Survival Data in Cost-effectiveness Analyses: Improving the Current State of Play. *Medical Decision Making,* 33, 740-742.

Guyot, Ades A, Beasley M, Beranger L, Pignon J-P & N, W. 2017. Extrapolation of Survival Curves from Cancer Trials Using External Information. *Medical Decision Making,* 37, 353-366.

Hallek, M., Fischer, K., Fingerle-Rowson, G., Fink, A. M., Busch, R., Mayer, J., Hensel, M., Hopfinger, G., Hess, G., von Grunhagen, U., Bergmann, M., Catalano, J., Zinzani, P. L., Caligaris-Cappio, F., Seymour, J. F., Berrebi, A., Jager, U., Cazin, B., Trneny, M., Westermann, A., Wendtner, C. M., Eichhorst, B. F., Staib, P., Buhler, A., Winkler, D., Zenz, T., Bottcher, S., Ritgen, M., Mendila, M., Kneba, M., Dohner, H., Stilgenbauer, S., Investigators, I. G. & Leukaem, G. C. L. 2010. Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet,* 376, 1164-1174.

Hatswell, A. J. 2016. REFORM OF CANCER DRUGS FUND New Cancer Drugs Fund solves the wrong problem. *Bmj-British Medical Journal,* 355.

Hawkins, N., Sculpher, M. & Epstein, D. 2005. Cost-effectiveness analysis of treatments for chronic disease: Using R to incorporate time dependency of treatment response. *Medical Decision Making,* 25, 511-519.

He, V. Y., Condon, J. R., You, J., Zhao, Y. & Burrow, J. N. 2015. Adverse outcome after incident stroke hospitalization for Indigenous and non-Indigenous Australians in the Northern Territory. *International Journal of Stroke,* 10, 89-95.

Hettle, R., Wouters, H., Ayres, J., Gani, R., Kelly, S., Lion, M. & Decramer, M. 2012. Cost-utility analysis of tiotropium versus usual care in patients with COPD in the UK and Belgium. *Respiratory Medicine,* 106, 1722-1733.

Hillen, T., Coshall, C., Tilling, K., Rudd, A. G., McGovern, R. & Wolfe, C. D. A. 2003a. Cause of stroke recurrence is multifactorial - Patterns, risk factors, and outcomes of stroke recurrence in the South London stroke register. *Stroke,* 34, 1457-1463.

Hillen, T., Davies, S., Rudd, A. G., Kieselbach, T. & Wolfe, C. D. 2003b. Self ratings of health predict functional outcome and recurrence free survival after stroke. *Journal of Epidemiology and Community Health,* 57, 960-966.

Ho, J. J., Hwang, J. S. & Wang, J. D. 2006. Life-expectancy estimations and the determinants of survival after 15 years of follow-up for 81 249 workers with permanent occupational disabilities. *Scandinavian Journal of Work, Environment and Health,* 32, 91-98.

Hougaard, P. 1999. Multi-state models: A review. *Lifetime Data Analysis,* 5, 239-264.

Hudgens, M. G., Li, C. X. & Fine, J. P. 2014. Parametric Likelihood Inference for Interval Censored Competing Risks Data. *Biometrics,* 70, 1-9.

Hudgens, M. G., Satten, G. A. & Longini, I. M. 2001. Nonparametric maximum likelihood estimation for competing risks survival data subject to interval censoring and truncation. *Biometrics,* 57, 74-80.

Hwang, J. S. & Wang, J. D. 1999. Monte Carlo estimation of extrapolation of quality-adjusted survival for follow-up studies. *Statistics in Medicine,* 18, 1627-1640.

Inglis, S. C., Lewsey, J. D., Jhund, P., Gillies, M., Langhorne, P., Walters, M., Briggs, A., Chalmers, J., McMurray, J. J. V. & Macintyre, K. 2009. Temporal trends in hospitalisation for stroke recurrence following incident hospitalisation for stroke. *European Heart Journal,* 30, 6-7.

Jackson, C., J, S., S, R., N, L., L, B., A, M. & L, S. 2017. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. *Medical decision making : an international journal of the Society for Medical Decision Making,* 37, 377-390.

Jackson, C. A., Hutchison, A., Dennis, M. S., Wardlaw, J. M., Lewis, S. C. & Sudlow, C. L. M. 2009. Differences Between Ischemic Stroke Subtypes in Vascular Outcomes Support a Distinct Lacunar Ischemic Stroke Arteriopathy A Prospective, Hospital-Based Study. *Stroke,* 40, 3679-3684.

Jackson, C. H., Sharples, L. D. & Thompson, S. G. 2010. Survival Models in Health Economic Evaluations: Balancing Fit and Parsimony to Improve Prediction. *International Journal of Biostatistics,* 6.

Jensen, M. L., Jorgensen, M. E., Hansen, E. H., Aagaard, L. & Carstensen, B. 2014. A Multistate Model and an Algorithm for Measuring Long-Term Adherence to Medication: A Case of Diabetes Mellitus Type 2. *Value in Health,* 17, 266-274.

Kim, G. M., Park, K. Y., Avery, R., Helenius, J., Rost, N., Rosand, J., Rosen, B. & Ay, H. 2014. Extensive Leukoaraiosis Is Associated With High Early Risk of Recurrence After Ischemic Stroke. *Stroke,* 45, 479-485.

Kim, H. T. 2007. Cumulative incidence in competing risks data and competing risks regression analysis. *Clinical Cancer Research,* 13, 559-565.

Kim, L. G. & Thompson, S. G. 2011. Estimation of Life-Years Gained and Cost Effectiveness Based on Cause-Specific Mortality. *Health Economics,* 20, 842-852.

Klein, J. P. & Andersen, P. K. 2005. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics,* 61, 223-229.

Koller, M. T., Raatz, H., Steyerberg, E. W. & Wolbers, M. 2012. Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine,* 31, 1089-1097.

Kreif, N., Grieve, R., Díaz, I. & Harrison, D. 2015. Evaluation of the Effect of a Continuous Treatment A Machine Learning Approach with an Application to Treatment for Traumatic Brain Injury, Health Economics Volume 24, Issue 9. *Health Economics* [Online], 24. Available: http://onlinelibrary.wiley.com/doi/10.1002/hec.3189/abstract [Accessed 01].

Kreif, N., Grieve, R., Radice, R. & Sekhon, J. S. 2013. Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation. *Health Services and Outcomes Research Methodology,* 13, 174-202.

Kuk, D. & Varadhan, R. 2013. Model selection in competing risks regression. *Statistics in Medicine,* 32, 3077-3088.

Kuwashiro, T., Sugimori, H., Kamouchi, M., Ago, T., Kitazono, T. & Iida, M. 2012. Lower Levels of High-density Lipoprotein Cholesterol on Admission and A Recurrence of Ischemic Stroke: A 12-month Follow-up of the Fukuoka Stroke Registry. *Journal of Stroke & Cerebrovascular Diseases,* 21, 561-568.

Lambert, P. C., Dickman, P. W., Nelson, C. P. & Royston, P. 2010. Estimating the crude probability of death due to cancer and other causes using relative survival models. *Stat Med,* 29, 885-95.

Latimer, N. R. 2013. Survival Analysis for Economic Evaluations Alongside Clinical Trials-Extrapolation with Patient-Level Data: Inconsistencies, Limitations, and a Practical Guide. *Medical Decision Making,* 33, 743-754.

Latimer, N. R. 2014. Response to "Survival Analysis and Extrapolation Modeling of Time-to-Event Clinical Trial Data for Economic Evaluation: An Alternative Approach" by Bagust and Beale. *Medical Decision Making,* 34, 279-282.

Latouche, A., Allignol, A., Beyersmann, J., Labopin, M. & Fine, J. P. 2013. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *Journal of Clinical Epidemiology,* 66, 648-653.

Lau, B., Cole, S. R. & Gange, S. J. 2009. Competing Risk Regression Models for Epidemiologic Data. *American Journal of Epidemiology,* 170, 244-256.

Lau, B., Cole, S. R., Moore, R. D. & Gange, S. J. 2008. Evaluating competing adverse and beneficial outcomes using a mixture model. *Statistics in Medicine,* 27, 4313-4327.

Le Pen, C., Rozenbaum, W., Downs, A., Maurel, F., Lilliu, H. & Brun, C. Effect of HAART on health status and hospital costs of severe HIV-infected patients: a modeling approach. *HIV Clinical Trials,* 2, 136-45.

Lee, J. Y., Park, J. P., Min, S. Y., Kim, S. S., Kim, S. H., Yi, M. Z., Kang, D. H., Song, J. M. & Song, J. K. 2009. Echocardiographic Characteristics of Patent Foramen Ovale Associated With Recurrence of Cryptogenic Stroke. *Journal of the American College of Cardiology,* 53, A245-A245.

Lee, J. Y., Song, J. K., Song, J. M., Kang, D. H., Yun, S. C., Kang, D. W., Kwon, S. U. & Kim, J. S. 2010. Association Between Anatomic Features of Atrial Septal Abnormalities Obtained by Omni-Plane Transesophageal Echocardiography and Stroke Recurrence in Cryptogenic Stroke Patients with Patent Foramen Ovate. *American Journal of Cardiology,* 106, 129-134.

Lee, L. J. H., Chang, Y. Y., Liou, S. H. & Wang, J. D. 2012. Estimation of benefit of prevention of occupational cancer for comparative risk assessment: methods and examples. *Occupational and Environmental Medicine,* 69, 582-586.

Lewsey, J., Jhund, P. S., Gillies, M., Chalmers, J. W. T., Redpath, A., Briggs, A., Walters, M., Langhorne, P., Capewell, S., McMurray, J. J. V. & MacIntyre, K. 2010. Temporal trends in hospitalisation for stroke recurrence following incident hospitalisation for stroke in Scotland. *BMC Medicine,* 8 (no pagination).

Li, C. X. 2016. The Fine-Gray model under interval censored competing risks data. *Journal of Multivariate Analysis,* 143, 327-344.

Longson, C. M., Landells, L. J., George, E., Boysen, M. & Dillon, A. 2016. REFORM OF CANCER DRUGS FUND Cancer Drugs Fund: the bigger picture. *Bmj-British Medical Journal,* 355.

Lunn, M. & Mcneil, N. 1995. Applying Cox Regression to Competing Risks. *Biometrics,* 51, 524-532.

Macedo-Vinas, M., De Angelis, G., Fankhauser, C., Safran, E., Schrenzel, J., Pittet, D. & Harbarth, S. 2011. Excess length of stay due to methicillin-resistant Staphylococcus aureus (MRSA) infection at a large Swiss hospital estimated by multi-state modelling. *BMC Proceedings. Conference: International Conference on Prevention and Infection Control, ICPIC,* 5.

Macedo-Vinas, M., De Angelis, G., Rohner, P., Safran, E., Stewardson, A., Fankhauser, C., Schrenzel, J., Pittet, D. & Harbarth, S. 2013. Burden of meticillin-resistant Staphylococcus aureus infections at a Swiss University hospital: excess length of stay and costs. *Journal of Hospital Infection,* 84, 132-137.

Marnane, M., Ni Chroinin, D., Callaly , E., Merwick, A., Hannon, N., Sheehan, O., Williams, E., Horgan, G., Duggan, J., Kyne, L., Moroney, J., Williams, D., McCormack, P. M. E., Crispino-O'Connell, G. & Kelly, P. J. Very early stroke recurrence causing long-term disability is independently associated with carotid stenosis in the north dublin population stroke study. 19th European Stroke Conference 2010 Barcelona Spain.

Mayor, S. 2016. New "managed access" process for Cancer Drugs Fund to go ahead, NHS England confirms. *Bmj-British Medical Journal,* 352, i1208.

Meacock, R., Sutton, M., Kristensen, S. R. & Harrison, M. 2017. Using Survival Analysis to Improve Estimates of Life Year Gains in Policy Evaluations. *Medical Decision Making,* 37.

Meira-Machado, L., de Una-Alvarez, J., Cadarso-Suarez, C. & Andersen, P. K. 2009. Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research,* 18, 195-222.

Mell, L. K. & Jeong, J. H. 2010. Pitfalls of Using Composite Primary End Points in the Presence of Competing Risks. *Journal of Clinical Oncology,* 28, 4297-4299.

Mell, L. K., Lau, S. K., Rose, B. S. & Jeong, J. H. 2012. Reporting of cause-specific treatment effects in cancer clinical trials with competing risks: a systematic review. *Contemp Clin Trials,* 33, 920-4.

Mitchell, B. G. & Gardner, A. 2012. Prolongation of length of stay and Clostridium difficile infection: a review of the methods used to examine length of stay due to healthcare associated infections. *Antimicrobial Resistance and Infection Control,* 1 (no pagination).

Most, S., Possnecker, W. & Tutz, G. 2016. Variable selection for discrete competing risks models. *Quality & Quantity,* 50, 1589-1610.

Nadeau, S., Jordan, J. & Mishra, S. 1992. Clinical Presentation as a Guide to Early Prognosis in Vertebrobasilar Stroke. *Stroke,* 23, 165-170.

Ndir, A., Diop, A., Faye, P. M., Cisse, M. F., Ndoye, B. & Astagneau, P. 2016a. Epidemiology and burden of bloodstream infections caused by extended-spectrum beta-lactamase producing enterobacteriaceae in a Pediatric Hospital in Senegal. *PLoS ONE,* 11 (2) (no pagination).

Ndir, A., Diop, A., Ka, R., Faye, P. M., Dia-Badiane, N. M., Ndoye, B. & Astagneau, P. 2016b. Infections caused by extended-spectrum beta-lactamases producing Enterobacteriaceae: clinical and economic impact in patients hospitalized in 2 teaching hospitals in Dakar, Senegal. *Antimicrobial Resistance and Infection Control,* 5.

Negrín, M. A., Nam, J. & Briggs, A. 2017. Bayesian Solutions for Handling Uncertainty in Survival Extrapolation *Medical decision making : an international journal of the Society for Medical Decision Making,* 37**,** 367 -376.

Nelson, C. L., Sun, J. L., Tsiatis, A. A. & Mark, D. B. 2008. Empirical estimation of life expectancy from large clinical trials: use of left-truncated, right-censored survival analysis methodology. *Statistics in medicine,* 27**,** 5525-5555.

Nelson, R. E., Nelson, S. D., Khader, K., Perencevich, E. L., Schweizer, M. L., Rubin, M. A., Graves, N., Harbarth, S., Stevens, V. W. & Samore, M. H. 2015. The magnitude of time-dependent bias in the estimation of excess length of stay attributable to healthcare-associated infections. *Infection Control and Hospital Epidemiology,* 36**,** 1089-1094.

NICE 2009. Rituximab for the first-line treatment of chronic lymphocytic leukaemia. NICE technology appraisal guidance 174.

NICE. 2013. *Guide to methods of technology appraisal* [Online]. National Institute for Health and Care Excellence. Available: https://www.nice.org.uk/article/pmg9 [Accessed Jun2017 2017].

Nicolaie, M. A., van Houwelingen, H. C. & Putter, H. 2010. Vertical modeling: A pattern mixture approach for competing risks modeling. *Statistics in Medicine,* 29**,** 1190-1205.

Ogasawara, K., Ogawa, A., Terasaki, K., Shimizu, H., Tominaga, T. & Yoshimoto, T. 2002a. Use of cerebrovascular reactivity in patients with symptomatic major cerebral artery occlusion to predict 5-year outcome: Comparison of xenon-133 and iodine-123-IMP single-photon emission computed tomography. *Journal of Cerebral Blood Flow and Metabolism,* 22**,** 1142-1148.

Ogasawara, K., Ogawa, A. & Yoshimoto, T. 2002b. Cerebrovascular reactivity to acetazolamide and outcome in patients with symptomatic internal carotid or middle cerebral artery occlusion - A xenon-133 single-photon emission computed tomography study. *Stroke,* 33**,** 1857-1862.

Ovbiagele, B. 2012. Should recent stroke patients be routinely screened for kidney disease? *Stroke. Conference,* 43.

Ovbiagele, B. 2013. Low-normal systolic blood pressure and secondary stroke risk. *Journal of Stroke and Cerebrovascular Diseases,* 22**,** 633-638.

Panayidou, K., Gsteiger, S., Egger, M., Kilcher, G., Carreras, M., Efthimiou, O., Debray, T. P. A., Trelle, S., Hummel, N. & GetReal methods review, g.

2016. GetReal in mathematical modelling: a review of studies predicting drug effectiveness in the real world. *Research synthesis methods,* 7, 264-77.

Peduzzi, P., Concato, J., Feinstein, A. R. & Holford, T. R. 1995. Importance of events per independent variable in proportional hazards regression analysis .2. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology,* 48, 1503-1510.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R. & Feinstein, A. R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology,* 49, 1373-1379.

Penn, A. M., Maclure, M., Lu, L., Bibok, M. B., Morrison, J. L., Votova, K., Penn, M., Balshaw, R. & Lesperance, M. L. 2016. Large centralized TIA assessment unit associated with reduction of recurrent stroke by up to 70%. *Stroke. Conference: American Heart Association/American Stroke Association,* 47.

PenTAG 2009. The clinical and cost-effectiveness of Rituximab for the 1st line treatment of Chronic Lymphocytic Leukaemia: an evidence review of the submission from Roche.

Pfirrmann, M., Hochhaus, A., Lauseker, M., Saussele, S., Hehlmann, R. & Hasford, J. 2011. Recommendations to meet statistical challenges arising from endpoints beyond overall survival in clinical trials on chronic myeloid leukemia. *Leukemia,* 25, 1433-1438.

Pintilie, M. 2007a. Analysing and interpreting competing risk data. *Statistics in Medicine,* 26, 1360-1367.

Pintilie, M. 2007b. Comments on 'Analysing and interpreting competing risk data' (original article and author's reply) - Author's reply. *Statistics in Medicine,* 26, 3523-3523.

Pintilie, M. 2007c. Comments on 'Analysing and interpreting competing risk data' by M. Pintilie, Statistics in Medicine 2006. DOI: 10.1002/sim.2655 - Author's reply. *Statistics in Medicine,* 26, 3679-3680.

Prentice, R., Kalbfleisch, J., Peterson, A., Flournoy, N., Farewell, V. & Breslow, N. 1978. The analysis of failure times in the presence of competing risks. *Biometrics,* 34, 541-554.

Price, M. J., Welton, N. J. & Ades, A. E. 2011. Parameterization of treatment effects for meta-analysis in multi-state Markov models. *Statistics in Medicine,* 30, 140-151.

Putter, H., Fiocco, M. & Geskus, R. B. 2007. Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine,* 26, 2389-2430.

Putter, H. & Spitoni, C. 2016. Non-parametric estimation of transition probabilities in non-Markov multi-state models: The landmark Aalen–Johansen estimator *Statistical Methods in Medical Research*.

Putter, H. & van Houwelingen, H. C. 2015. Frailties in multi-state models: Are they identifiable? Do we need them? *Statistical Methods in Medical Research,* 24, 675-692.

Radice, R., Ramsahai, R., Grieve, R., Kreif, N., Sadique, Z. & Sekhon, J. S. 2012. Evaluating treatment effectiveness in patient subgroups: a comparison of propensity score methods with an automated matching approach. *International Journal of Biostatistics,* 8.

Rael, M., Ishak, K. J. & Benedict, A. TO Join or Not to Join? Analysis Of Progression-Free And Overall Survival Using Multi-State Modeling. *Value in Health,* 19, A107.

Reulen, H. & Kneib, T. 2016. Boosting multi-state models. *Lifetime Data Analysis,* 22, 241-262.

Roche. 2008. *Rituximab for the 1st line treatment of chronic lymphocytic leukaemia (TA174)* [Online]. London, UK: National Institute for Health and Clinical Excellence. Available: https://www.nice.org.uk/guidance/ta174/resources/roche-submission2 [Accessed June 2017 2017].

Rodríguez-Girondo, M. & Uña-Álvarez, J. d. 2016. Methods for testing the Markov condition in the illness?death model a comparative study, Statistics in Medicine Early View. *Statistics in Medicine* [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/sim.6940/abstract [Accessed 30/03/2016].

Rutten-Jacobs, L. C. A., Maaijwee, N. A. M., Arntz, R. M., Schoonderwaldt, H. C., Dorresteijn, L. D., van der Vlugt, M. J., van Dijk, E. J. & de Leeuw, F. E. 2013. Long-Term Risk of Recurrent Vascular Events after Young Stroke: The FUTURE Study. *Annals of Neurology,* 74, 592-601.

Satagopan, J. M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D. & Auerbach, A. D. 2004. A note on competing risks in survival data analysis. *British Journal of Cancer,* 91, 1229-1235.

Scheike, T. H. & Zhang, M. J. 2008. Flexible competing risks regression modeling and goodness-of-fit. *Lifetime Data Analysis,* 14, 464-483.

Schumacher, M., Ohneberg, K. & Beyersmann, J. 2016. Competing risk bias was common in a prominent medical journal. *Journal of Clinical Epidemiology,* 80, 135-136.

Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics,* 6, 461-464.

Scotland, G., McKeigue, P., Philip, S., Leese, G. P., Olson, J. A., Looker, H. C., Colhoun, H. M. & Javanbakht, M. 2016. Modelling the cost-effectiveness of adopting risk-stratified approaches to extended screening intervals in the national diabetic retinopathy screening programme in Scotland. *Diabetic Medicine,* 33, 886-895.

Shi, W., Greco, G., Gelijns, A., Blackstone, E., Moskowitz, A., O'Gara, Jeffries, N. & Bagiella, E. 2014. Estimating the cost of health care associated infections controlling for both patient variability and time-dependent bias. *Value in Health,* 17 (3), A186.

Simon, N., Friedman, J. & Hastie, T. 2013. A blockwise descent algorithm for group-penalized multiresponse and multinomial regression. *arXiv Preprint*.

Sonnenberg, F. A. & Beck, J. R. 1993. Markov-Models in Medical Decision-Making - a Practical Guide. *Medical Decision Making,* 13, 322-338.

Spitoni, C., Verduijn, M. & Putter, H. 2012. Estimation and Asymptotic Theory for Transition Probabilities in Markov Renewal Multi-State Models. *International Journal of Biostatistics,* 8.

Stamplecoski, M., Fang, J., Kapral, M. K. & Silver, F. L. 2012. Predictors of readmission to hospital following acute stroke. *Stroke. Conference,* 43.

Stevens, A. J. & Longson, C. 2013. At the Center of Health Care Policy Making: The Use of Health Technology Assessment at NICE. *Medical Decision Making,* 33, 320-324.

Stevens, V. W., Khader, K., Nelson, R. E., Jones, M., Rubin, M. A., Brown, K. A., Evans, M. E., Greene, T., Slade, E. & Samore, M. H. 2015. Excess length of stay attributable to clostridium difficile infection (CDI) in the acute care setting: A multistate model. *Infection Control and Hospital Epidemiology,* 36, 1024-1030.

Stewardson, A., Fankhauser, C., De Angelis, G., Rohner, P., Safran, E., Schrenzel, J., Pittet, D. & Harbarth, S. 2013. Burden of bloodstream infection caused by extended-spectrum beta-lactamase-producing enterobacteriaceae determined using multistate modeling at a swiss university hospital and a nationwide predictive model. *Infection Control and Hospital Epidemiology,* 34, 133-143.

Stewardson, A., Fankhauser, C., De Angelis, G., Rohner, P., Schrenzel, J., Pittet, D., Safran, E. & Harbarth, S. 2012. Cohort study examining the burden of bloodstream infection caused by extended-spectrum beta-lactamaseproducing Enterobacteriaceae at a Swiss university hospital using multistate modelling. *Clinical Microbiology and Infection,* 18, 529.

Stewardson, A. J., Graves, N., Allignol, A. & Harbarth, S. 2015. The cost bloodstream infections caused by antimicrobial susceptible and non-susceptible Enterobacteriaceae and Staphylococcus aureus in European hospitals. *Antimicrobial Resistance and Infection Control. Conference: 3rd International Conference on Prevention and Infection Control, ICPIC,* 4.

Sun, X. & Faunce, T. 2008. Decision-analytical modelling in health-care economic evaluations. *European Journal of Health Economics,* 9, 313-323.

Sun, Y., Lee, S. H., Heng, B. H. & Chin, V. S. 2013. 5-year survival and rehospitalization due to stroke recurrence among patients with hemorrhagic or ischemic strokes in Singapore. *BMC Neurology,* 13 (no pagination).

Sun, Y. Q., Hyun, S. G. & Gilbert, P. 2008. Testing and Estimation of Time-Varying Cause-Specific Hazard Ratios with Covariate Adjustment. *Biometrics,* 64, 1070-1079.

Tai, B. C., Wee, J. & Machin, D. 2011. Analysis and design of randomised clinical trials involving competing risks endpoints. *Trials,* 12.

Tappenden, P., Chilcott, J., Ward, S., Eggington, S., Hind, D. & Hummel, S. 2006. Methodological issues in the economic analysis of cancer treatments. *European Journal of Cancer,* 42, 2867-2875.

Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological,* 58, 267-288.

Tichy, E. 2014. Are Cycles Needed in Markov Models? - the Continuous Model as a Simpler Approach. *Value in Health,* 17, A558-A558.

Toschke, A. M., Gulliford, M. C., Wolfe, C. D. A., Rudd, A. G. & Heuschmann, P. U. 2011. Antihypertensive treatment after first stroke in primary care: results from the General Practitioner Research Database. *Journal of Hypertension,* 29, 154-160.

Tsiatis, A. A. 1975. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci U S A,* 72, 20-22.

Turlach, B. A., Venables, W. N. & Wright, S. J. 2005. Simultaneous variable selection. *Technometrics,* 47, 349-363.

Tutz, G., Possnecker, W. & Uhlmann, L. 2015. Variable selection in general multinomial logit models. *Computational Statistics & Data Analysis,* 82, 207-222.

van den Hout, A., Ogurtsova, E., Gampe, J. & Matthews, F. E. 2014. Investigating healthy life expectancy using a multi-state model in the presence of missing data and misclassification. *Demographic Research,* 30, 1219-1243.

van Rosmalen, J., Toy, M. & O'Mahony, J. F. 2013. A Mathematical Approach for Evaluating Markov Models in Continuous Time without Discrete-Event Simulation. *Medical Decision Making,* 33, 767-779.

van Walraven, C. & McAlister, F. A. 2016. Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. *Journal of Clinical Epidemiology,* 69, 170-173.

Varadhan, R., Weiss, C. O., Segal, J. B., Wu, A. W., Scharfstein, D. & Boyd, C. 2010. Evaluating Health Outcomes in the Presence of Competing Risks A

Review of Statistical Methods and Clinical Applications. *Medical Care,* 48, S96-S105.

Welton, N. J. & Ades, A. E. 2005. Estimation of Markov chain transition probabilities and rates from fully and partially observed data: Uncertainty propagation, evidence synthesis, and model calibration. *Medical Decision Making,* 25, 633-645.

Wetmore, J. B., Mahnken, J. D. & Phadnis, M. A. 2016. Association of multiple ischemic strokes with mortality in incident hemodialysis patients: An application of multistate model to determine transition probabilities in a retrospective observational cohort. *BMC Nephrology,* 17 (1) (no pagination).

Williams, C., Lewsey, J. D., Briggs, A. H. & Mackay, D. F. 2017a. Cost-effectiveness Analysis in R Using a Multi-state Modeling Survival Analysis Framework: A Tutorial. *Med Decis Making,* 37, 340-352.

Williams, C., Lewsey, J. D., Mackay, D. F. & Briggs, A. H. 2017b. Estimation of Survival Probabilities for Use in Cost-effectiveness Analyses: A Comparison of a Multi-state Modeling Survival Analysis Approach with Partitioned Survival and Markov Decision-Analytic Modeling. *Med Decis Making,* 37, 427-439.

Williamson, P. R., Kolamunnage-Dona, R., Philipson, R. & Marson, A. G. 2008. Joint modelling of longitudinal and competing risks data. *Statistics in Medicine,* 27, 6426-6438.

Wolbers, M., Koller, M. T., Stel, V. S., Schaer, B., Jager, K. J., Leffondre, K. & Heinze, G. 2014. Competing risks analyses: objectives and approaches. *European Heart Journal,* 35, 2936-2941.

Wollenweber, F. A., Zietemann, V., Rominger, A., Opherk, C., Bayer-Karpinska, A., Gschwendtner, A., Andrews, L. C., Burger, K., Duering, M. & Dichgans, M. 2014. The Determinants of Dementia After Stroke (DEDEMAS) Study: protocol and pilot data. *International Journal of Stroke,* 9, 387-392.

Woods, B., Sideris, E., Palmer, S., Latimer, N. & Soares, M. 2017. NICE DSU Technical Support Document 19. Partitioned Survival Analysis for Decision Modelling in Health Care: A Critical Review.

Yang, D.-C. M. D., Lee, L. J.-H. M. D., Phd, Hsu, C.-C. M. D., Drph, Chang, Y.-Y. M. S. C., Wang, M.-C. M. D., Lin, W.-H. M. D., Chang, C.-M. M. D., Wang, J.-D. M. D. & Scd 2012. Estimation of Expected Life-Years Saved From Successful Prevention of End-Stage Renal Disease in Elderly Patients With Diabetes: A nationwide study from Taiwan. *Diabetes Care,* 35, 2279-2285.

Yokota, C., Hasegawa, Y., Minematsu, K. & Yamaguchi, T. 1998. Effect of acetazolamide reactivity and long-term outcome in patients with major cerebral artery occlusive diseases. *Stroke,* 29, 640-644.

Yokota, C., Minematsu, K., Hasegawa, Y. & Yamaguchi, T. 2004. Long-term prognosis, by stroke subtypes, after a first-ever stroke: A hospital-based study over a 20-year period. *Cerebrovascular Diseases,* 18**,** 111-116.

Zhou, B. Q., Latouche, A., Rocha, V. & Fine, J. 2011. Competing Risks Regression for Stratified Data. *Biometrics,* 67**,** 661-670.